



Informatica® Mass Ingestion  
May 2022

# Mass Ingestion

Informatica Mass Ingestion Mass Ingestion  
May 2022

© Copyright Informatica LLC 2019, 2022

Publication Date: 2022-09-01

# Table of Contents

<b>Chapter 1: Introducing Mass Ingestion.....</b>	<b>12</b>
My Services page. . . . .	12
Explore page. . . . .	13
<b>Chapter 2: Getting Started with Mass Ingestion.....</b>	<b>16</b>
Installing Secure Agents. . . . .	17
Secure Agent installation on Windows. . . . .	17
Secure Agent installation on Linux. . . . .	21
Secure Agent services. . . . .	24
Database Ingestion service. . . . .	24
Mass Ingestion (Files). . . . .	28
CMI Streaming Agent. . . . .	30
Creating projects and project folders. . . . .	36
Editing your user profile. . . . .	37
<b>Chapter 3: Connectors and Connections.....</b>	<b>38</b>
Mass Ingestion connectors. . . . .	38
Mass Ingestion Applications connectors. . . . .	38
Mass Ingestion Databases connectors. . . . .	40
Mass Ingestion Files connectors. . . . .	42
Mass Ingestion Streaming connectors. . . . .	43
Mass Ingestion connection properties. . . . .	44
Configuring a connection. . . . .	44
Adobe Analytics Mass Ingestion connection properties. . . . .	44
Advanced FTP V2 connection properties. . . . .	45
Advanced FTPS V2 connection properties. . . . .	47
Advanced SFTP V2 connection properties. . . . .	49
Amazon Redshift V2 connection properties. . . . .	50
Amazon Kinesis connection properties. . . . .	52
Amazon S3 V2 connection properties. . . . .	56
AMQP connection properties. . . . .	61
Db2 for i Database Ingestion connection properties. . . . .	62
Db2 for LUW Database Ingestion connection properties. . . . .	63
Db2 for zOS Database Ingestion connection properties. . . . .	64
Databricks Delta connection properties. . . . .	65
Flat file connection properties. . . . .	68
Google Analytics Mass Ingestion connection properties. . . . .	69
Google BigQuery V2 connection properties. . . . .	69
Google Cloud Storage V2 connection properties. . . . .	71
Google PubSub - Mass Ingestion Streaming connection properties. . . . .	75

Hadoop Files V2 connection properties. . . . .	75
JDBC V2 connection properties. . . . .	77
JMS connection properties. . . . .	78
Kafka connection properties. . . . .	79
Marketo V3 connection properties. . . . .	85
Microsoft Azure Blob Storage V3 connection properties. . . . .	85
Microsoft Azure Data Lake Storage Gen2 connection properties . . . . .	86
Microsoft Azure Event Hub connection properties. . . . .	88
Microsoft Azure Synapse Analytics Database Ingestion connection properties. . . . .	88
Microsoft Azure Synapse SQL connection properties. . . . .	90
Microsoft Dynamics 365 Mass Ingestion connection properties. . . . .	91
Microsoft SQL Server connection properties. . . . .	94
MongoDB Mass Ingestion connection properties. . . . .	97
MQTT connection properties. . . . .	98
MySQL connection properties. . . . .	99
Netezza connection properties. . . . .	100
NetSuite Mass Ingestion connection properties. . . . .	100
OPC UA connection properties. . . . .	101
Oracle Database Ingestion connection properties. . . . .	103
Oracle Fusion Cloud Mass Ingestion connection properties. . . . .	108
PostgreSQL connection properties. . . . .	109
Salesforce Mass Ingestion connection properties. . . . .	111
SAP HANA Database Ingestion connection properties. . . . .	113
SAP ODP Extractor connection properties. . . . .	115
ServiceNow Mass Ingestion connection properties. . . . .	119
Snowflake Data Cloud connection properties. . . . .	121
REST V2 connection properties. . . . .	125
Teradata connection properties. . . . .	127
Workday Mass Ingestion connection properties. . . . .	129
Zendesk Mass Ingestion connection properties. . . . .	130

## **Chapter 4: Mass Ingestion Applications..... 132**

Use cases. . . . .	132
System requirements. . . . .	133
Mass Ingestion Applications architecture. . . . .	133
Supported sources . . . . .	134
Guidelines for Marketo sources. . . . .	135
Guidelines for Microsoft Dynamics 365 sources. . . . .	135
Guidelines for NetSuite sources. . . . .	136
Guidelines for Oracle Fusion Cloud sources. . . . .	136
Guidelines for Salesforce sources. . . . .	137
Guidelines for SAP ECC and SAP S4/HANA sources. . . . .	137
Guidelines for Workday sources. . . . .	138

Guidelines for ServiceNow sources. . . . .	139
Guidelines for Zendesk sources. . . . .	140
Supported targets. . . . .	141
Guidelines for Amazon Redshift targets. . . . .	142
Guidelines for Amazon S3, Google Cloud Storage, and Microsoft Azure Data Lake Storage Gen2 targets. . . . .	143
Guidelines for Databricks Delta targets. . . . .	150
Guidelines for Google BigQuery targets. . . . .	151
Guidelines for Microsoft Azure Synapse Analytics targets. . . . .	152
Guidelines for Snowflake targets. . . . .	152
Avro data types. . . . .	153
Handling source schema changes. . . . .	154
Configuring application ingestion tasks. . . . .	154
Before you begin. . . . .	155
Defining basic task information. . . . .	155
Configuring the source. . . . .	156
Configuring the target. . . . .	160
Configuring schedule and runtime options. . . . .	172
Deploying an application ingestion task. . . . .	174
Running an application ingestion job. . . . .	175
Stopping an application ingestion job . . . . .	175
Aborting an application ingestion job. . . . .	176
Resuming an application ingestion job. . . . .	176
Restart and recovery for incremental load jobs. . . . .	176
Redeploying an application ingestion job. . . . .	177
Undeploying an application ingestion job. . . . .	177
<b>Chapter 5: Mass Ingestion Databases.....</b>	<b>178</b>
Use cases. . . . .	178
Supported source and target types. . . . .	179
Mass Ingestion Databases architecture. . . . .	180
Mass Ingestion Databases system requirements. . . . .	182
General limitations and guidelines. . . . .	182
Mass Ingestion Databases sources - preparation and usage considerations. . . . .	183
Db2 for i sources. . . . .	183
Db2 for LUW sources. . . . .	184
Db2 for z/OS sources. . . . .	185
Microsoft SQL Server and Azure SQL Database sources. . . . .	188
MongoDB sources. . . . .	190
MySQL sources. . . . .	191
Netezza sources. . . . .	192
Oracle sources. . . . .	193
PostgreSQL sources. . . . .	205

SAP HANA sources. . . . .	208
Teradata sources. . . . .	210
Mass Ingestion Databases targets - usage considerations. . . . .	210
Amazon Redshift targets. . . . .	211
Amazon S3, Flat File, Google Cloud Storage, and Microsoft Azure Data Lake Storage targets. . . . .	211
Databricks Delta targets. . . . .	212
Google BigQuery targets. . . . .	213
Kafka targets and Kafka-enabled Azure Event Hubs targets. . . . .	214
Microsoft Azure Synapse Analytics targets. . . . .	215
Oracle targets. . . . .	216
Snowflake targets. . . . .	217
Default directory structure for CDC files on Amazon S3, Google Cloud Storage, and Azure Data Lake Storage Gen2 targets. . . . .	217
Custom directory structure for output files on Amazon S3, Google Cloud Storage, Flat File, and ADLS Gen2 targets. . . . .	220
Supported Avro data types. . . . .	228
Schema drift handling. . . . .	229
Ability to apply deletes as soft deletes on the target. . . . .	230
Configuring a database ingestion task. . . . .	230
Before you begin. . . . .	230
Defining basic task information. . . . .	231
Configuring the source. . . . .	232
Configuring the target. . . . .	243
Configuring schedule and runtime options. . . . .	267
Deploying a database ingestion task. . . . .	270
Running database ingestion jobs. . . . .	270
Running a deployed database ingestion job. . . . .	271
Running initial load jobs based on a schedule. . . . .	271
Managing database ingestion jobs. . . . .	271
Stopping a database ingestion job. . . . .	271
Aborting a database ingestion job. . . . .	272
Resuming a database ingestion job. . . . .	272
Overriding schema drift options when resuming a database ingestion job. . . . .	273
Redeploying a database ingestion job. . . . .	274
Undeploying a database ingestion job. . . . .	274
Resynchronizing source and target objects. . . . .	275
Restart and recovery for incremental load jobs. . . . .	275
Default Data Type Mappings. . . . .	276
Db2 for i Source and Amazon Redshift Target. . . . .	276
Db2 for i Source and Databricks Delta Target. . . . .	277
Db2 for i Source and Google BigQuery Target. . . . .	278
Db2 for i Source and Microsoft Azure Synapse Analytics Target. . . . .	290
Db2 for i Source and Oracle Target. . . . .	291

Db2 for i Source and Snowflake Target. . . . .	293
Db2 for Linux, UNIX, and Windows Source and Amazon Redshift Target. . . . .	294
DB2 for Linux, UNIX, and Windows Source and Databricks Delta Target. . . . .	295
DB2 for Linux, UNIX, and Windows Source and Google BigQuery Target. . . . .	296
DB2 for Linux, UNIX, and Windows Source and Microsoft Azure Synapse Analytics Target. . . . .	299
DB2 for Linux, UNIX, and Windows Source and Snowflake Target. . . . .	300
Db2 for z/OS Source and Amazon Redshift Target. . . . .	301
Db2 for zOS Source and Databricks Delta Target. . . . .	302
Db2 for zOS Source and Google BigQuery Target. . . . .	303
Db2 for z/OS Source and Microsoft Azure Synapse Analytics Target. . . . .	307
Db2 for z/OS Source and Snowflake Target. . . . .	308
Microsoft SQL Server or Azure SQL Database Source and Amazon Redshift Target. . . . .	309
Microsoft SQL Server Source and Databricks Delta Target. . . . .	311
Microsoft SQL Server Source and Google BigQuery Target. . . . .	312
Microsoft SQL Server or Azure SQL Database Source and Microsoft Azure Synapse Analytics Target. . . . .	317
Microsoft SQL Server Source and Oracle Target. . . . .	318
Microsoft SQL Server or Azure SQL Database Source and Snowflake Target. . . . .	320
MySQL Source and Amazon Redshift Target. . . . .	321
MySQL Source and Databricks Delta Target. . . . .	323
MySQL Source and Google BigQuery Target. . . . .	325
MySQL Source and Microsoft Azure Synapse Analytics Target. . . . .	332
MySQL Source and Snowflake Target. . . . .	333
Netezza Source and Amazon Redshift Target. . . . .	335
Netezza Source and Databricks Delta Target. . . . .	336
Netezza Source and Google BigQuery Target. . . . .	337
Netezza Source and Microsoft Azure Synapse Analytics Target. . . . .	340
Netezza Source and Snowflake Target. . . . .	341
Oracle Source and Amazon Redshift Target. . . . .	342
Oracle Source and Databricks Delta Target. . . . .	344
Oracle Source and Google BigQuery Target. . . . .	345
Oracle Source and Microsoft Azure Synapse Analytics Target. . . . .	352
Oracle Source and Oracle Target. . . . .	353
Oracle Source and Snowflake Target. . . . .	354
PostgreSQL Source and Amazon Redshift Target. . . . .	355
PostgreSQL Source and Databricks Delta Target. . . . .	357
PostgreSQL Source and Google BigQuery Target. . . . .	358
PostgreSQL Source and Microsoft Azure Synapse Analytics Target. . . . .	363
PostgreSQL Source and Snowflake Target. . . . .	365
SAP HANA Source and Amazon Redshift Target. . . . .	367
SAP HANA Source and Databricks Delta Target. . . . .	369
SAP HANA Source and Google BigQuery Target. . . . .	370
SAP HANA Source and Microsoft Azure Synapse Analytics Target. . . . .	372

SAP HANA Source and Snowflake Target. . . . .	374
Teradata Source and Amazon Redshift Target. . . . .	375
Teradata Source and Databricks Delta Target. . . . .	377
Teradata Source and Google BigQuery Target. . . . .	378
Teradata Source and Microsoft Azure Synapse Analytics Target. . . . .	387
Teradata Source and Snowflake Target. . . . .	389
<b>Chapter 6: Mass Ingestion Files. . . . .</b>	<b>392</b>
Use cases. . . . .	392
Mass Ingestion Files source types. . . . .	392
Mass Ingestion Files target types. . . . .	393
Mass Ingestion Files actions. . . . .	394
Mass Ingestion Files runtime options. . . . .	395
Mass Ingestion Files security. . . . .	395
Configuring a file ingestion task . . . . .	396
Before you begin. . . . .	396
Defining basic task information. . . . .	396
Configuring the source. . . . .	397
Configuring the target. . . . .	424
Configuring file processing actions. . . . .	444
Configuring runtime options. . . . .	445
Running a file ingestion task. . . . .	445
Aborting a file ingestion job. . . . .	446
Key ring command reference. . . . .	446
createKeyRing. . . . .	446
createKeyPair. . . . .	446
listKeys. . . . .	447
importKeys. . . . .	448
exportKeyPairs. . . . .	448
exportPublicKeys. . . . .	449
deleteKeys. . . . .	449
changePassphrase. . . . .	450
<b>Chapter 7: Mass Ingestion Streaming . . . . .</b>	<b>451</b>
Use cases. . . . .	451
Mass Ingestion Streaming sources. . . . .	452
Amazon Kinesis Streams sources. . . . .	452
AMQP sources. . . . .	453
Azure Event Hubs Kafka sources. . . . .	453
Flat File sources. . . . .	454
Google PubSub sources. . . . .	454
JMS sources. . . . .	454
Kafka sources. . . . .	455



MQTT sources. . . . .	455
OPC UA sources. . . . .	456
REST V2 sources. . . . .	456
Mass Ingestion Streaming targets. . . . .	456
Amazon Kinesis Data Firehose target. . . . .	457
Amazon Kinesis Streams target. . . . .	457
Amazon S3 target. . . . .	458
Databricks Delta target. . . . .	458
Flat file target. . . . .	458
Google Cloud Storage V2 target. . . . .	459
Google PubSub target. . . . .	459
JDBC V2 target. . . . .	460
Kafka target. . . . .	460
Microsoft Azure Data Lake Storage Gen2 target. . . . .	461
Microsoft Azure Event Hubs target. . . . .	461
Transformations in Mass Ingestion Streaming. . . . .	461
Supported data formats. . . . .	462
Combiner transformation. . . . .	462
Filter transformation. . . . .	463
Python transformation. . . . .	463
Splitter transformation. . . . .	464
Format Converter transformation. . . . .	464
Configuring a streaming ingestion task. . . . .	464
Before you begin. . . . .	465
Defining basic task information. . . . .	465
Configuring a source . . . . .	465
Configuring a target . . . . .	474
Configuring a transformation . . . . .	482
Configuring runtime options. . . . .	486
Deploying a streaming ingestion task. . . . .	487
Undeploying a streaming ingestion job. . . . .	487
Stopping and resuming streaming ingestion jobs. . . . .	488
Mass Ingestion Streaming REST API. . . . .	488
Dataflows. . . . .	488
jobs. . . . .	490
MIJobs. . . . .	492
status. . . . .	494
statistics. . . . .	496
history. . . . .	498
<b>Chapter 8: Monitoring Mass Ingestion Jobs.....</b>	<b>500</b>
Monitoring your ingestion jobs. . . . .	500
Monitoring all ingestion jobs. . . . .	501

Job properties. . . . .	504
Viewing details for an ingestion job. . . . .	505
Application ingestion job details. . . . .	505
Database ingestion job details. . . . .	511
File ingestion job details. . . . .	516
Streaming ingestion job details. . . . .	518
<b>Chapter 9: Asset Management. . . . .</b>	<b>523</b>
Editing ingestion tasks. . . . .	523
Copying projects, folders, and tasks. . . . .	524
Moving folders and tasks. . . . .	524
Renaming projects and folders. . . . .	524
Renaming database ingestion tasks. . . . .	525
Renaming file ingestion tasks. . . . .	525
Renaming streaming ingestion tasks. . . . .	526
Deleting projects, folders, and tasks. . . . .	526
Tags. . . . .	527
Creating tags. . . . .	527
Assigning tags to an ingestion task. . . . .	528
Editing and deleting tags. . . . .	529
Asset dependencies. . . . .	529
Configuring user permissions on an ingestion task. . . . .	530
Asset migration. . . . .	531
Dependent objects. . . . .	531
Schedules. . . . .	532
Asset export. . . . .	532
Asset import. . . . .	535
Post-import tasks. . . . .	537
Source control. . . . .	537
Source control actions. . . . .	538
Source control and the Git repository. . . . .	538
Configuring repository access. . . . .	539
Source control pulls. . . . .	539
Checking out and checking in objects. . . . .	541
Deleting an object. . . . .	542
Reverting to an older version. . . . .	542
Undoing a checkout. . . . .	542
Unlinking an object. . . . .	543
Working with multiple objects. . . . .	543
Viewing source control columns on the Explore page. . . . .	544
Source control best practices. . . . .	544

<b>Chapter 10: Troubleshooting.....</b>	<b>546</b>
Troubleshooting an application ingestion task. . . . .	546
Troubleshooting a database ingestion task. . . . .	547
Troubleshooting a streaming ingestion task. . . . .	550
<b>Index.....</b>	<b>552</b>

## CHAPTER 1

# Introducing Mass Ingestion

Use the Informatica Intelligent Cloud Services Mass Ingestion service to ingest data at scale from database, application, file, and streaming data sources and transfer the data with low latency to selected cloud targets and messaging systems.

Mass Ingestion provides the following ingestion solutions:

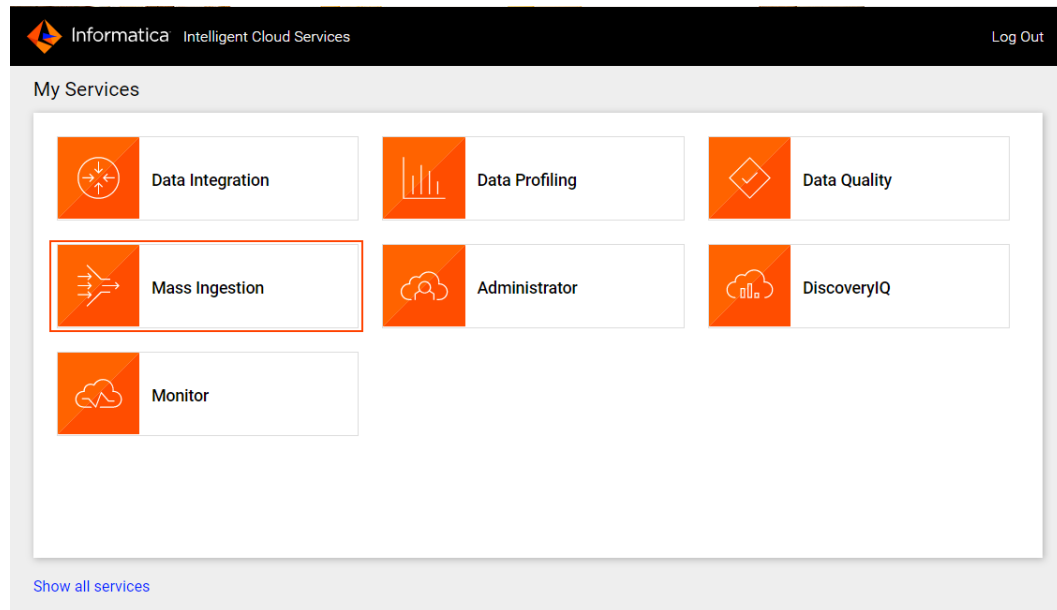
- **Mass Ingestion Applications.** Propagates data from objects in Software-as-a-Service (SaaS) and on-premise applications to cloud-based data lakes, data warehouses, and event streaming platforms. An application ingestion task can transfer a point-in-time snapshot of data in a source schema to a target in a batch operation. An application ingestion task can also incrementally propagate change data and schema changes in near real time from a source to a target on a continuous basis. If you select the combined initial and incremental load type, the application ingestion task performs an initial load and then automatically switches to incremental load processing of data changes.
- **Mass Ingestion Databases.** Propagates data from source objects in a database management system (DBMS) to multiple types of targets. A database ingestion task can transfer a point-in-time snapshot of all source data in a schema to a target in a batch operation. A database ingestion task can also incrementally propagate DML data and schema changes in near real time from a database source to a target on a continuous basis. If you select the combined initial and incremental load type, the database ingestion task performs an initial load and then automatically switches to incremental load processing of data changes.
- **Mass Ingestion Files.** Transfers a large number of files of different types between on-premises and cloud repositories. You can use Mass Ingestion Files to track and monitor file transfers.
- **Mass Ingestion Streaming.** Transfers real-time streaming and machine data from selected sources to selected messaging systems and batch targets.

## My Services page

When you log in to Informatica Intelligent Cloud Services, the **My Services** page displays the services that your organization is licensed to use and any common services that are available under the same license,

such as Administrator. If your organization has trial licenses for additional services, the page also displays those services.

The following image shows an example **My Services** page:



To use Mass Ingestion, click the **Mass Ingestion** box.

## Explore page

Use the **Explore** page to work with your Informatica Intelligent Cloud Services projects and assets.

### Finding projects and assets on the Explore page

Use any of the following methods to find your projects and assets on the **Explore** page:

- Explore by projects and folders. View all projects or select a particular project.
- Explore by asset types. View all assets or view assets of a particular type.
- Explore by tags. View assets associated with a particular tag.
- Search for projects or assets. To search all projects, folders, and assets in the organization, view the **Explore** page by **All Projects**, and then enter a name or description in the Find box. Or, to narrow your search, view the **Explore** page by **Asset Types** and select an asset type from the **All Assets** list. Then, in the Find box, enter a name or description in full or part.
- Sort the search results. Sort the **Explore** page by name, asset type, last update date, create date, or description. When you sort by type, the **Explore** page groups assets by asset type. It does not list the asset types in alphabetical order.

- Filter the objects on the page. To filter objects, click the **Filter** icon. To apply a filter, click **Add Field**, select the property to filter by, and then enter the property value. The filters available depend on how you view the page. You can specify multiple filters.

For example, to find all the assets in your organization with the tag SW Region, view the **Explore** page by **All Assets** and then click the **Filter** icon. Add the **Tags** filter and enter "SW Region."

**Tip:** Filtering is available on other pages in addition to the **Explore** page. For example, on the **Import Assets** page, you can filter by status to find the assets that imported successfully.

You can see projects, folders, and assets for all of the services that you use. If you select an asset to open it or perform an action, and the asset is created in a different service than the one you have open, the service opens in a new browser tab.

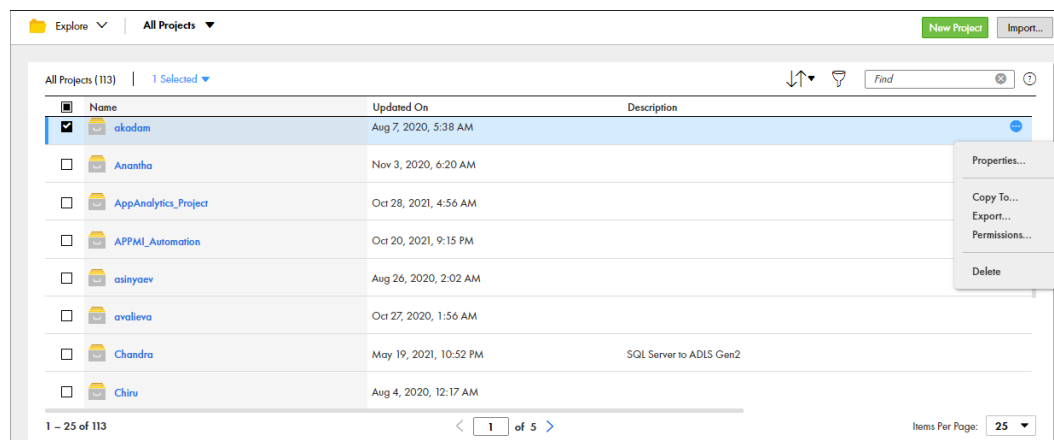
The following characters cannot be used on the **Explore** page:

# ? ' | { } " ^ & [ ] / \

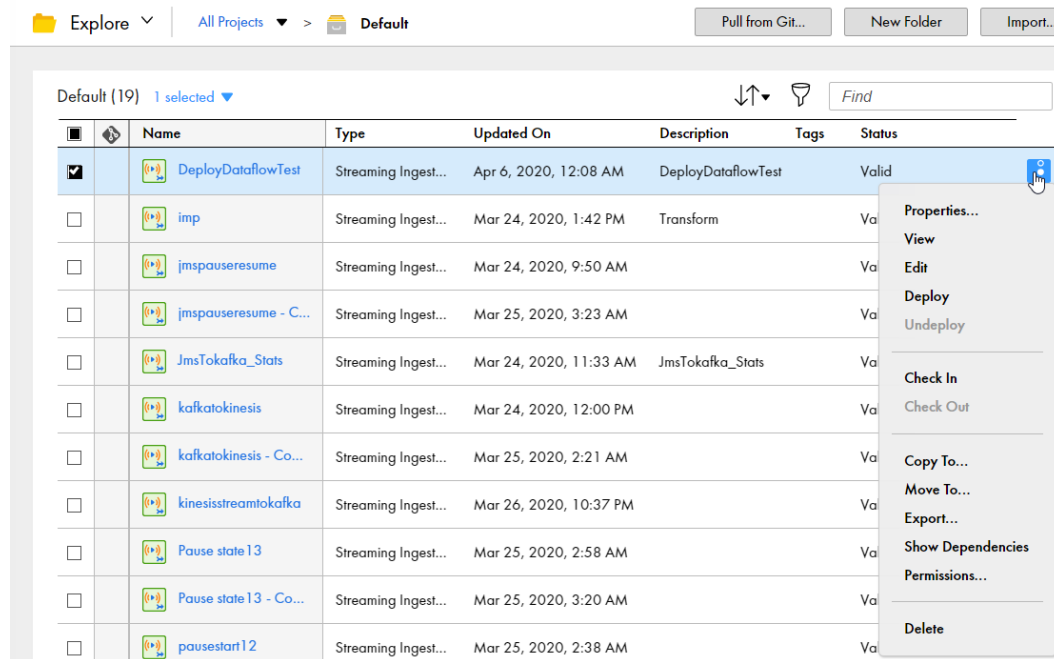
## Working with projects and assets on the Explore page

You can perform actions on projects and assets on the **Explore** page. Click the Actions menu in the row that contains the project or asset. The Actions menu lists the actions you can perform based on your user role privileges and the permissions for the selected objects. For example, your user role might have privileges to view and run tasks but not to delete tasks.

The following image shows the actions that you can perform on a project:



The following image shows actions that you can perform on a streaming ingestion task asset:



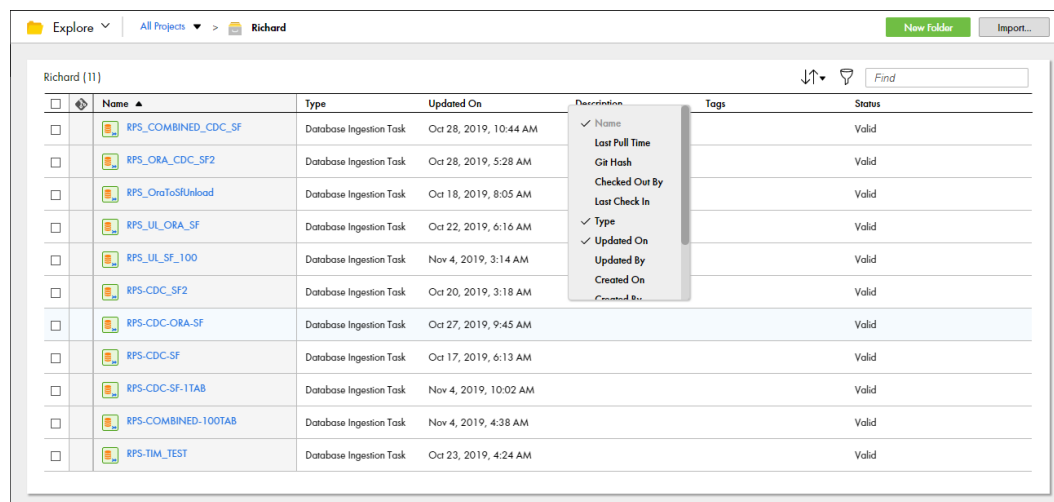
You can perform an action on multiple objects at one time. To select the objects individually, select the check box to the left of each object. To select all of the listed objects, select the check box at the top, next to the **Name** column heading. Then select an action from the **n selected** or **All selected** menu.

## Customizing the Explore page

You can display, hide, or rearrange columns on the **Explore** page. To display or hide columns, right-click any column heading and then select or deselect the column names in the menu.

**Note:** If source control is enabled in your environment, the menu contains source control options such as Last Pull time, Git Hash, Checked Out By, and Last Check In.

The following image shows the column customization menu:



To rearrange columns, click a column heading and drag it to a different location.

## CHAPTER 2

# Getting Started with Mass Ingestion

Before you configure an ingestion task, verify that all prerequisite tasks have been completed.

### Step 1. Check system requirements

Check the following items:

- For Mass Ingestion Databases minimum system requirements, see *Mass Ingestion Databases > Mass Ingestion Databases system requirements*.
- To determine the web browsers that are compatible with Mass Ingestion and the operating systems that are supported for the Secure Agent, check the Product Availability Matrix (PAM) for Informatica Intelligent Cloud Services (IICS) on the Informatica Network at <https://network.informatica.com/community/informatica-network/product-availability-matrices/>
- To determine the source and target types and versions that are supported for each ingestion type, check the KB article [What are the supported sources and targets for IICS Cloud Mass Ingestion service?](#).

### Step 2. Set up an organization.

If you are the administrator, set up an organization from the **Organization** page in Administrator. An organization is a secure area within the Informatica Intelligent Cloud Services repository that stores your licenses, user accounts, ingestion tasks, and information about jobs and security.

Then, configure users, user groups, and user role permissions for the organization.

If your organization has the Organization Hierarchy license, you can also create one or more sub-organizations within your organization. You can create sub-organizations to represent different business environments within your company. For example, you might create separate sub-organizations to represent your development, testing, and production environments.

For more information, see "Organizations" in the Administrator help.

### Step 3. Download and install a Secure Agent

On the **Runtime Environments** page in Administrator, download a Secure Agent and install it. A Secure Agent is a lightweight program that runs tasks and enables secure communication across the firewall between your organization and Informatica Intelligent Cloud Services. Mass Ingestion does not support the Hosted Agent or serverless runtime environments for any ingestion type.

When you download and install a Secure Agent, a *Secure Agent group*, also called a runtime environment, is created. A Secure Agent group can contain one Secure Agent or multiple agents if your licensing and ingestion type allows it.

**Note:** For Mass Ingestion Files, Mass Ingestion Databases, and Mass Ingestion Streaming, you can have multiple agents in a Secure Agent group, provided that you have the Secure Agent Cluster license. Then an



available agent is picked from the active agent list for the Secure Agent group. For Mass Ingestion Applications, the ingestion jobs must be deployed to a single Secure Agent.

Enable services and connectors for the Secure Agent group. This action downloads components and packages based on your selections and creates an associated runtime environment. For more information, see "Secure Agent Groups" in the Administrator help.

#### Step 4. Configure the runtime environment

On the **Runtime Environments** page in Administrator, select your runtime environment.

A runtime environment is the execution platform for running tasks. You must have at least one runtime environment in your organization for users to be able to run tasks. If you create another runtime environment, you must add an unassigned Secure Agent to it.

Under **System Configuration Details**, configure properties for the CMI Streaming Agent, Database Ingestion, or Mass Ingestion (file ingestion) service. For more information, see ["Secure Agent services" on page 24](#).

#### Step 5. Configure connections

On the **Connections** page, configure connection properties for the source and target connectors that you want to use in ingestion tasks.

#### Step 6. Create your project

From the **Explore** page in Mass Ingestion service, create projects and project folders to organize your ingestion tasks. A project can contain multiple subfolders. See ["Creating projects and project folders" on page 36](#).

## Installing Secure Agents

You can install Secure Agents on Windows or Linux.

### Secure Agent installation on Windows

On Windows, the Secure Agent runs as a Windows service. When you install the Secure Agent, you also install the Informatica Cloud Secure Agent Manager.

By default, the Secure Agent starts when you start Windows. You can stop and restart the Secure Agent using the Secure Agent Manager or Windows Services. If you install the Secure Agent on a different volume than you use to run the installation program, you must start and stop the Secure Agent from Windows Services.

You can also use the Secure Agent Manager to check the Secure Agent status and configure proxy information.

You can launch the Secure Agent Manager from the Start menu or desktop icon. When you close the Secure Agent Manager, it minimizes to the Windows taskbar notification area for quick access.

When you install a Secure Agent, you perform the following tasks:

1. Verify that the machine where you install the Secure Agent meets the minimum requirements.
2. Download the Secure Agent installer files.
3. Install and register the Secure Agent.

## Secure Agent requirements on Windows

You can install the Secure Agent on any machine that has internet connectivity and can access Informatica Intelligent Cloud Services.

Verify the following requirements before you install the Secure Agent on Windows:

- Verify that the machine on which you install the Secure Agent uses a supported operating system. For the list of supported operating systems for the Secure Agent, see the Product Availability Matrix (PAM) for Informatica Intelligent Cloud Services on the [Product Availability Matrices page](#) on Informatica Network.
- Verify that the machine where you install the Secure Agent has at least 5 GB of free disk space.
- Verify that the account you use to install the Secure Agent has access to all remote directories that contain flat source or target files.
- Verify that no other Secure Agent is installed on the machine. If another Secure Agent is installed on the machine, you must uninstall it first.

For more information about Secure Agent requirements, contact Informatica Global Customer Support.

## Configure the firewall

If your organization uses a protective firewall, include the Informatica Intelligent Cloud Services domain name or IP address ranges in the list of approved domain names or IP addresses. You should also enable the port that the Secure Agent uses. This ensures that the Secure Agent can perform all necessary tasks through the firewall.

The Secure Agent uses port 443 (HTTPS) to connect to the internet. Configure your firewall to allow traffic to pass over port 443.

The whitelists of domains and IP addresses can vary according to your data center, which is also called a POD (Point of Deployment). You can identify your POD through the URL that appears when you open any service in Informatica Intelligent Cloud Services. The first few characters of the URL string identify the POD. For example, if the URL starts with `usw3.dm-us.informaticacloud.com`, your POD is USW3.

You can find the whitelists of Informatica Intelligent Cloud Services domains and IP addresses for different PODs in [this Knowledge Base article](#) on Informatica Network or by clicking the link at the top of the **Runtime Environments** page in Administrator.

## Secure Agent permissions on Windows

A Secure Agent requires certain permissions to transfer data between sources and targets.

When you install a Secure Agent on Windows, the Secure Agent must be part of the local Administrators group.

## Configure Windows settings

Before you use the Secure Agent on Windows, configure proxy settings and a Windows Secure Agent service login.

You can configure proxy settings in Secure Agent Manager. Configure a login for the Windows Secure Agent service on Windows.

**Note:** If you use the Secure Agent for Informatica Cloud Data Wizard, you do not need to configure proxy settings or a Windows service login for the Secure Agent.

## Downloading and installing the Secure Agent on Windows

To install the Secure Agent on a Windows machine, you must download and run the Secure Agent installation program and then register the agent.

Secure Agent registration requires an install token. To get the install token, copy the token when you download the agent or use the **Generate Install Token** option in Administrator. The token expires after 24 hours.

Before you download and install the Secure Agent, verify that no other Secure Agent is installed on the machine. If there is, you must uninstall it.

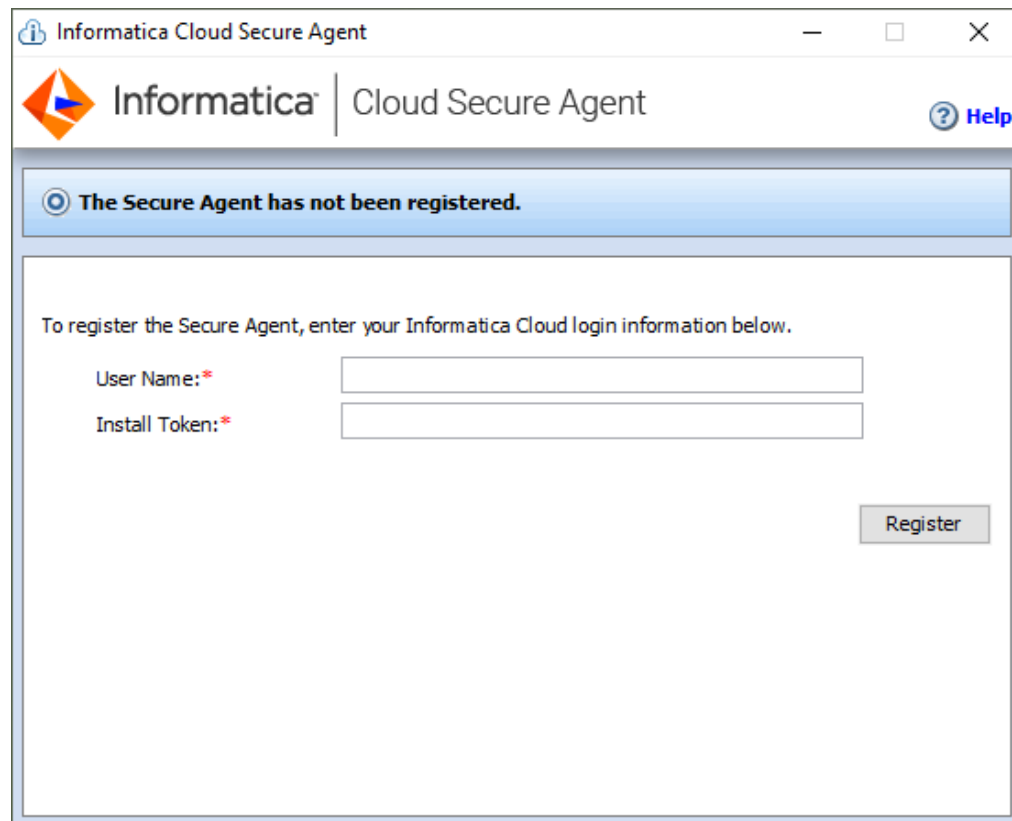
**Tip:** To verify the checksum of the Secure Agent installation program, use the agent REST API version 2 resource. For more information about the agent resource, see *REST API Reference*.

1. Open Administrator and select **Runtime Environments**.
2. On the **Runtime Environments** page, click **Download Secure Agent**.
3. Select the Windows 64-bit operating system platform, copy the install token, and then click **Download**.

The installation program is downloaded to your machine. The name of the installation program is `agent64_install_ng_ext.exe`.

4. Run the installation program:
  - a. Specify the Secure Agent installation directory, and click **Next**.
  - b. Click **Install** to install the agent.

The Secure Agent Manager opens and prompts you to register the agent as shown in the following image:



The screenshot shows a window titled "Informatica Cloud Secure Agent". The window has a blue header bar with the Informatica logo and the text "Cloud Secure Agent". Below the header, there is a blue banner with a circular icon and the text "The Secure Agent has not been registered." Below this banner, there is a white area with the text "To register the Secure Agent, enter your Informatica Cloud login information below." There are two input fields: "User Name: \*" and "Install Token: \*". A "Register" button is located at the bottom right of the input area.

5. If you did not copy the install token when you downloaded the agent, click **Generate Install Token** on the **Runtime Environments** page in Administrator, and copy the token.

6. In the Secure Agent Manager, enter the following information, and then click **Register**:

Option	Description
User Name	User name that you use to access Informatica Intelligent Cloud Services.
Install Token	Token that you copied.

The Secure Agent Manager displays the status of the Secure Agent. It takes a minute for all of the services to start.

7. If your organization uses an outgoing proxy server to connect to the internet, enter the proxy server information.
8. Close the Secure Agent Manager.

The Secure Agent Manager minimizes to the taskbar and continues to run as a service until stopped.

## Configure the proxy settings on Windows

If your organization uses an outgoing proxy server to connect to the internet, the Secure Agent connects to Informatica Intelligent Cloud Services through the proxy server. The Secure Agent installer configures the proxy server settings for the Secure Agent based on settings configured in the browser. You can change the proxy server settings through the Secure Agent Manager.

Contact your network administrator for the correct proxy settings.

1. In the Secure Agent Manager, click **Proxy**.
2. To enter proxy server settings, click **Use a Proxy Server**.
3. Enter the following information:

Field	Description
Proxy Host	Required. Host name of the outgoing proxy server that the Secure Agent uses.
Proxy Port	Required. Port number of the outgoing proxy server.
User Name	User name to connect to the outgoing proxy server.
Password	Password to connect to the outgoing proxy server.

4. Click **OK**.

The Secure Agent Manager restarts the Secure Agent to apply the settings.

## Configure a login for a Windows Secure Agent Service

On Windows, configure a network login for the Secure Agent service. The Secure Agent can access the network with the privileges and permissions associated with the login.

Configure a login for the machine on which the Secure Agent is installed to allow the Secure Agent to access directories to configure and run tasks. When you configure connections, configure tasks, and run tasks that use Flat File or FTP/SFTP connection types, the Secure Agent might require read and write permissions on the related directories.

For example, to browse to a directory to configure a Flat File or FTP/SFTP connection, the Secure Agent login might require permission to access the directory. Without a Secure Agent login with the appropriate permissions, Informatica Intelligent Cloud Services cannot display the directory in the **Browse for Directory** dialog box.

1. Go to the **Services** window from the Windows Administrative tools.
2. In the **Services** window, right-click the Informatica Cloud Secure Agent service and choose **Properties**.
3. In the **Properties** dialog box, click the **Log On** tab.
4. To configure a login, select **This Account**.
5. Enter an account and password.  
Use an account with the required privileges and permissions for the network security defined for the domain. By default, the account format is <domain name>\<user name>.
6. Click **OK**.
7. In the **Services** window, restart the Secure Agent service for the changes to take effect.

## Secure Agent installation on Linux

On Linux, the Secure Agent runs as a process. You can use a shell command line to install, register, start, stop, and uninstall the Secure Agent.

You can also use the shell command line to check the Secure Agent status.

When you install a Secure Agent, you perform the following tasks:

1. Verify that the machine where you install the Secure Agent meets the minimum requirements.
2. Download the Secure Agent installer files.
3. Install and register the Secure Agent.

## Secure Agent requirements on Linux

You can install the Secure Agent on any machine that has internet connectivity and can access Informatica Intelligent Cloud Services. Before you install the Secure Agent on Linux, verify the system requirements.

Verify the following requirements before you install the Secure Agent on Linux:

- Verify that the machine on which you install the Secure Agent uses a supported operating system. For the list of supported operating systems for the Secure Agent, see the Product Availability Matrix (PAM) for Informatica Intelligent Cloud Services on the [Product Availability Matrices page](#) on Informatica Network.
- Verify that the machine where you install the Secure Agent has at least 5 GB of free disk space.
- The account that you use to install the Secure Agent must have access to all remote directories that contain flat source or target files.
- If you use PowerCenter, install the Secure Agent using a different user account than the account you used to install PowerCenter.

Informatica Intelligent Cloud Services and PowerCenter use some common environment variables. If the environment variables are not set correctly for Informatica Intelligent Cloud Services, your jobs might fail at run time.

For more information about Secure Agent requirements, contact Informatica Global Customer Support.

## Configure the firewall

If your organization uses a protective firewall, include the Informatica Intelligent Cloud Services domain name or IP address ranges in the list of approved domain names or IP addresses. You should also enable the port that the Secure Agent uses. This ensures that the Secure Agent can perform all necessary tasks through the firewall.

The Secure Agent uses port 443 (HTTPS) to connect to the internet. Configure your firewall to allow traffic to pass over port 443.

The whitelists of domains and IP addresses can vary according to your data center, which is also called a POD (Point of Deployment). You can identify your POD through the URL that appears when you open any service in Informatica Intelligent Cloud Services. The first few characters of the URL string identify the POD. For example, if the URL starts with `usw3.dm-us.informaticacloud.com`, your POD is USW3.

You can find the whitelists of Informatica Intelligent Cloud Services domains and IP addresses for different PODs in [this Knowledge Base article](#) on Informatica Network or by clicking the link at the top of the **Runtime Environments** page in Administrator.

## Secure Agent permissions on Linux

A Secure Agent requires certain permissions to transfer data between sources and targets.

When you install a Secure Agent on Linux, the Secure Agent must have read/write/execute permissions for the installation directory.

## Downloading and installing the Secure Agent on Linux

To install the Secure Agent on a Linux machine, you must download and run the Secure Agent installation program and then register the agent.

Secure Agent registration requires an install token. To get the install token, copy the token when you download the agent or use the **Generate Install Token** option in Administrator. The token expires after 24 hours.

When you register the agent, it is added to its own Secure Agent group by default. You can add the agent to a different Secure Agent group.

Before you download and install the Secure Agent, verify that no other Secure Agent is installed on the machine using the same Linux user account. If there is, you must uninstall it.

**Tip:** To verify the checksum of the Secure Agent installation program, use the agent REST API version 2 resource. For more information about the agent resource, see *REST API Reference*.

1. Open Administrator and select **Runtime Environments**.
2. On the **Runtime Environments** page, click **Download Secure Agent**.
3. Select the Linux 64-bit operating system platform, copy the install token, and then click **Download**.  
The installation program is downloaded to your machine. The name of the installation program is `agent64_install_ng_ext.bin`.
4. Save the installation program to a directory on the machine where you want to run the Secure Agent.  
**Note:** If the file path contains spaces, the installation might fail.
5. From a shell command line, navigate to the directory where you downloaded the installation program and enter the following command:  

```
./agent64_install_ng_ext.bin -i console
```
6. When the installer completes, navigate to the following directory:

<Secure Agent installation directory>/apps/agentcore

7. To start the Secure Agent, enter the following command:

```
./infaagent startup
```

The Secure Agent Manager starts. You must register the agent using the user name that you use to access Informatica Intelligent Cloud Services. You must also supply the install token.

8. If you did not copy the install token when you downloaded the agent, click **Generate Install Token** on the **Runtime Environments** page in Administrator, and copy the token.
9. To register the agent, in the <Secure Agent installation directory>/apps/agentcore directory, enter one of the following commands using your Informatica Intelligent Cloud Services user name and the token that you copied:

- To add the agent to its own Secure Agent group, use the following command:

```
./consoleAgentManager.sh configureToken <user name> <install token>
```

- To add the agent to an existing Secure Agent group, use the following command:

```
./consoleAgentManager.sh configureTokenWithRuntime <user name> <install token>  
<Secure Agent group name>
```

**Note:** If the command includes a Secure Agent group name that doesn't exist, the Secure Agent is not assigned to a group. Be sure to use a valid Secure Agent group name.

The following table lists the command options:

Option	Description
User Name	Required. Informatica Intelligent Cloud Services user name of the user installing the Secure Agent.
Install Token	Required. The install token that you copied.
Secure Agent group name	Optional. Include when you want to add the agent to an existing Secure Agent group instead. If this option isn't included in the command, the agent will be in its own Secure Agent group.

You can check the registration status of a Secure Agent using the following command:

```
./consoleAgentManager.sh isConfigured
```

## Configure the proxy settings on Linux

If your organization uses an outgoing proxy server to connect to the internet, the Secure Agent connects to Informatica Intelligent Cloud Services through the proxy server.

The Secure Agent installer configures the proxy server settings for the Secure Agent based on settings configured in the browser. You can update the proxy server settings defined for the Secure Agent from the command line.

To configure the proxy server settings for the Secure Agent on a Linux machine, use a shell command that updates the `proxy.ini` file. Contact the network administrator to determine the proxy settings.

1. Navigate to the following directory:

```
<Secure Agent installation directory>/apps/agentcore
```

2. To update the `proxy.ini` file, enter the following command:

```
./consoleAgentManager.sh configureProxy <proxy host> <proxy port> <proxy user name>  
<proxy password>
```

3. Restart the Secure Agent.

# Secure Agent services

Secure Agent services are pluggable microservices that the Secure Agent uses for data processing. Each Secure Agent service runs independently of the other services that run on the agent.

The independent services architecture provides the following benefits:

- The Secure Agent does not restart when you add a connector or package.
- Services are not impacted when another service restarts.
- Downtime during upgrades is minimized. The upgrade process installs a new version of the Secure Agent, updates connector packages, and applies configuration changes for the services. To minimize downtime, the old agent remains available and continues to run ingestion jobs during the upgrade. The new version of the Secure Agent runs jobs that start after the upgrade process completes.

The services that run on a Secure Agent vary based on your licenses and the Informatica Intelligent Cloud Services that your organization uses. For Mass Ingestion, the following Secure Agent services are available:

- Database Ingestion - for running application ingestion jobs and database ingestion jobs
- CMI Streaming Agent - for running streaming ingestion jobs
- Mass Ingestion - for running file ingestion jobs

Each service has a unique set of configuration properties. You might need to configure a service or change the service properties to optimize performance or if you are instructed to do so by Informatica Global Customer Support.

## Database Ingestion service

Mass Ingestion Applications and Mass Ingestion Databases use the Database Ingestion agent service to run ingestion jobs.

After you download the Secure Agent to your runtime environment and enable the Database Ingestion service, the Database Ingestion packages are pushed to the on-premises system where the Secure Agent runs. You can then optionally configure properties for the Database Ingestion service that runs on the Secure Agent.

### Database Ingestion service properties

To change or optimize the behavior of the Database Ingestion service that your Secure Agent group uses, configure Database Ingestion properties for your runtime environment.

To configure the properties, open your runtime environment and click **Edit**. Under **System Configuration Details**, select the **Database Ingestion** service and the **DBMI\_AGENT\_CONFIG** type.



The following table describes the Database Ingestion agent service properties:

Property	Description
maxTaskUnits	<p>The maximum number application ingestion tasks and database ingestion tasks that can run concurrently on an on-premises machine where the Secure Agent is running. Each ingestion task requires at least 1 task unit to run.</p> <p>Informatica recommends that you configure no more than 2.5 task units per core. For example, if you have an 8-core machine, you can set the maxTaskUnits property to 20. If you have a 16-core machine, you can set it to 40.</p> <p>During initial load processing, this property determines the number of tables that can be unloaded simultaneously. Remaining tables are queued and start unload processing when resources become available.</p> <p>During incremental load processing, this property determines the number of application ingestion and database ingestion jobs that can run simultaneously.</p> <p>A single job can process many tables. The total number of tables that can be processed is limited only by available memory. On the average, 25 MB of RAM is required per table for an initial load task based on a 1 KB row size.</p>
serviceLogRetentionPeriod	<p>The number of days to retain each internal Database Ingestion service log file after the last update is written to the file. When this retention period elapses, the log file is deleted. The default value is 7 days.</p> <p>Service logs are retained on the Secure Agent host where they are created: <code>&lt;infaagent&gt;/apps/Database_Ingestion/logs</code>.</p> <p><b>Note:</b> This property is applicable to both Mass Ingestion Applications and Mass Ingestion Databases.</p>
taskLogRetentionPeriod	<p>The number of days to retain each job log file after the last update is written to the file. When this retention period elapses, the log file is deleted. The default value is 7 days.</p>
ociPath	<p>For Oracle sources and targets, the path to the Oracle Call Interface (OCI) file <code>oci.dll</code> or <code>libclntsh.so</code>. For a DBMI agent that is running, this value is appended to the path that is specified in the PATH environment variable on Windows or in the LD_LIBRARY_PATH environment variable on Linux.</p> <p><b>Note:</b> This property is applicable only to Mass Ingestion Databases.</p>
serviceUrl	<p>The URL that the Database Ingestion service uses to connect to the Informatica Intelligent Cloud Services cloud.</p> <p><b>Note:</b> This property is applicable to both Mass Ingestion Applications and Mass Ingestion Databases.</p>

Property	Description
logLevel	<p>The level of detail to include in the logs that the Database Ingestion service produces. Options are:</p> <ul style="list-style-type: none"> <li>- TRACE</li> <li>- DEBUG</li> <li>- INFO</li> <li>- WARN</li> <li>- ERROR</li> </ul> <p>The default value is TRACE.</p> <p><b>Note:</b> This property is applicable to both Mass Ingestion Applications and Mass Ingestion Databases.</p>
taskExecutionHeapSize	<p>The maximum heap size, in gigabytes, for the Task Execution service. This value, in conjunction with maxTaskUnits property, affects the number of concurrent application ingestion and database ingestion tasks that can run on a Secure Agent. Try increasing the heap size to run more tasks concurrently. Enter this value followed by "g" for gigabytes, for example, '9g'. The default value is '8g'.</p> <p><b>Note:</b> This property is applicable to both Mass Ingestion Applications and Mass Ingestion Databases.</p>
useProxy	<p>Set this property to true to enable the DBMI Agent to go through a proxy when connecting to or writing data to targets. The DBMI Agent then uses the proxy settings from the Secure Agent proxy configuration. By default, proxy settings are not used.</p> <p><b>Note:</b> This property is applicable to both Mass Ingestion Applications and Mass Ingestion Databases.</p>
intermediateStorageDirectory	<p>For incremental load and combined initial and incremental load jobs, the local root directory under which intermediate files that contain data are stored when the <b>Enable Persistent Storage</b> option is selected in the associated task definitions.</p> <p><b>Note:</b> This property is applicable only to Mass Ingestion Databases.</p>
storageBackupDirectory	<p>For incremental load and combined initial and incremental load jobs, the path to the directory that stores backup files when the <b>Enable Persistent Storage</b> option is selected in the associated task definitions.</p> <p><b>Note:</b> This property is applicable only to Mass Ingestion Databases.</p>
storageProperties	<p>For incremental load and combined initial and incremental load jobs, a comma-separated list of key=value pairs that is used when the <b>Enable Persistent Storage</b> option is selected in the associated task definitions. Specify this property only at the direction of Informatica Global Customer Support.</p> <p><b>Note:</b> This property is applicable only to Mass Ingestion Databases.</p>

Property	Description
task_container.jvm.allowExceptionForInvalidEncodedData	<p>If you receive transliteration errors that report invalid encoding to UTF-8, and you do not want to repair or correct the source data, set this property to false so that database ingestion jobs do not fail when trying to unload the data from the source. With this setting, the Database Ingestion service passes an equivalent Java property to the DataDirect JDBC driver to prevent the exception from occurring. After you set this property, you must restart the Database Ingestion service.</p> <p><b>Note:</b> This property is applicable only to Mass Ingestion Databases.</p>
testProperty	<p>Do not set this property. It is intended for internal use by Informatica Global Customer Support and technical staff. This property appears only if you select DBMI_AGENT_ENV in the <b>Type</b> field.</p>

## Database Ingestion Agent environment variables

To change or optimize the behavior of the Database Ingestion Agent, define the following environment variables:

Environment Variable	Description
DBMI_REPLACE_UNSUPPORTED_CHARS	<p>For Microsoft Azure Synapse Analytics targets, controls whether an application ingestion job or database ingestion job replaces characters in character data that the target cannot process correctly. To enable character replacement, set this environment variable to true.</p> <p><code>DBMI_REPLACE_UNSUPPORTED_CHARS=true</code></p> <p>Mass Ingestion Applications or Mass Ingestion Databases then uses the character that is specified in the DBMI_UNSUPPORTED_CHARS_REPLACEMENT environment variable to replace unsupported characters.</p>
DBMI_UNSUPPORTED_CHARS_REPLACEMENT	<p>If the DBMI_REPLACE_UNSUPPORTED_CHARS environment variable is set to true, specifies the character that replaces the characters in source data that a Microsoft Azure Synapse Analytics target cannot process correctly.</p> <p>Default value: ? (question mark)</p> <p><b>Note:</b> Define this environment variable only for Mass Ingestion Databases.</p>
DBMI_WRITER_CONN_POOL_SIZE	<p>Indicates the number of connections that an application ingestion job or database ingestion job uses to propagate the change data to the target. The default value is 8. Valid values are 4 through 8.</p>

Environment Variable	Description
DBMI_WRITER_RETRIES_MAX_COUNT	If a network issue occurs while a database ingestion job is loading source data to an Amazon S3 or Microsoft Azure Data Lake Storage Gen2 target, indicates the maximum number of times that the database ingestion job retries a request to continue the initial load or incremental load. If all of the retries fail, the job fails. The default value is 5.
DBMI_WRITER_RETRIES_INTERVAL_IN_MILLIS	Specifies the time interval, in milliseconds, that a database ingestion job waits before retrying the request to continue the initial load or incremental load to an Amazon S3 or Microsoft Azure Data Lake Storage Gen2 target if a network issue occurs. The default value is 1000.

**Note:** After you define or change an environment variable, restart the Database Ingestion Agent for the changes to take effect.

## Mass Ingestion (Files)

To change or optimize the behavior of Mass Ingestion Files that your Secure Agent group uses, configure Mass Ingestion properties for your runtime environment in Administrator.

You can configure the following properties:

Type	Name	Description
AGENT_RUNTIME_SETTINGS	file-listener-snapshot-dir	A directory where the snapshots of a new file listener components are added. You can add the following directory paths: <ul style="list-style-type: none"> <li>- A path relative to the <code>MassIngestionRuntime</code> directory. For example, <code>../data/monitor</code>.</li> <li>- The absolute path. For example, <code>&lt;Secure agent installation directory&gt;/apps/MassIngestionRuntime/data/monitor</code> where <i>Secure agent installation directory</i> is the name of the directory where the secure agent is installed.</li> </ul> <b>Note:</b> Use the snapshot directory shared with all agents when multiple Secure Agents are present in a group.
AGENT_RUNTIME_SETTINGS	mi-task-workspace-dir	A directory in the agent that file ingestion tasks use as an intermediate staging area when transferring files to a target. The directory is a custom location in the agent. The path can be a shared location, mounted location, or a location apart from the default location in the agent.

Type	Name	Description
AGENT_RUNTIME_SETTINGS	mi-task-quarantine-dir	<p>A directory where the file ingestion task stores the infected files detected when you run a virus scan. The directory is the custom location in the agent. The path can be a shared location, mounted location, or a location apart from the default location in the agent.</p> <p>For example, userdata\quarantine</p> <p><b>Note:</b> To automatically clean up the quarantine directory, set the agent property for the quarantine location to a system temporary files location such as /tmp/informatica/fmi/quarantine.</p>
AGENT_RUNTIME_SETTINGS	file-listener-max-pool-size	<p>The maximum number of threads to execute the file listener.</p> <p>Default is 20.</p>
AGENT_RUNTIME_SETTINGS	file-listener-core-pool-size	<p>The total number of threads.</p> <p>Default is 20.</p>
AGENT_RUNTIME_SETTINGS	fmi-task-max-pool-size	<p>The maximum number of threads to execute the file ingestion task.</p> <p>Default is 50.</p>
AGENT_RUNTIME_SETTINGS	fmi-task-core-pool-size	<p>The initial or minimum number of threads.</p> <p>Default is 20.</p>
AGENT_RUNTIME_SETTINGS	ftp-receive-socket-buffer-size	<p>The buffer size for FTP inbound packets.</p> <p>Default is 16 bytes.</p>
AGENT_RUNTIME_SETTINGS	ftp-send-socket-buffer-size	<p>The buffer size for FTP outbound packets.</p> <p>Default is 16 bytes.</p>
AGENT_RUNTIME_SETTINGS	http-client-timeout	<p>The timeout duration in seconds for Agent requests to Informatica Intelligent Cloud Services.</p> <p>Default is 30 seconds.</p>
PGP_SETTINGS	public-keyring-path	<p>The directory to store the public key ring. You can add the following directory paths:</p> <ul style="list-style-type: none"> <li>- A path relative to the directory where mass ingestion is installed. For example, <code>../data/pubring.pkr</code> where <i>pubring.pkr</i> is the name of the file where you store the public key ring.</li> <li>- The absolute path. For example, <code>&lt;Secure agent installation directory&gt;/apps/MassIngestionRuntime/data/pubring.pkr</code> where <i>pubring.pkr</i> is the name of the file where you store the public key ring and <i>Secure agent installation directory</i> is the name of the directory where the agent is installed.</li> </ul>

Type	Name	Description
PGP_SETTINGS	secret-keyring-path	The directory to store the secret key ring. You can add the following directory paths: <ul style="list-style-type: none"> <li>- A path relative to the directory where mass ingestion is installed. For example, <code>../data/secring.pkr</code> where <i>secring.pkr</i> is the name of the file where you store the secret key ring.</li> <li>- The absolute path. For example, <code>&lt;Secure agent installation directory&gt;/apps/MassIngestionRuntime/data/secring.pkr</code> where <i>secring.pkr</i> is the name of the file where you store the secret key ring and <i>Secure Agent installation directory</i> is the name of the directory where the agent is installed.</li> </ul>
JVM_SETTINGS	app-heap-size	The minimum and maximum heap sizes of the Mass Ingestion Files application. Default is <code>-Xms256m -Xmx2048m</code> .
JVM_SETTINGS	lcm-heap-size	The minimum and maximum heap sizes of life-cycle management scripts. Default is <code>-Xms32m -Xmx128m</code> .

You can configure the following properties in the **Custom Configuration Details** area when you edit a Secure Agent:

Type	Name	Description
AGENT_RUNTIME_SETTINGS	ComplexFileDisableWriteChecksum	Set the value to <b>True</b> to ignore the <code>crc</code> file. The job runs successfully with Hadoop Files V2 as source and Snowflake Cloud Data Warehouse V2 as the target.

## CMi Streaming Agent

Use the CMI Streaming Agent to define and deploy streaming ingestion tasks. You configure streaming ingestion tasks in the Mass Ingestion service.

A CMI Streaming Agent runs on an on-premise system and works in conjunction with the Mass Ingestion Streaming service. In an on-premise system, the CMI Streaming Agent runs the jobs deployed by Mass Ingestion Streaming. The agent provides status and statistics updates of each job.

On Linux, the CMI Streaming Agent does not start if the agent installation directory name contains a space. The agent returns a connection timeout status. After a few restart attempts, the agent goes into the error state.

**Note:** Prior to Spring 2020 April release of Informatica Intelligent Cloud Services Mass Ingestion service, CMI Streaming Agent was called Streaming Ingestion Agent.

## CMI Streaming Agent properties

To change or optimize the behavior of the CMI Streaming Agent, configure agent properties for your run-time environment. Configure CMI Streaming Agent properties in the **System Configuration Details** area when you edit a Secure Agent.

You can configure Engine, Agent, and Script properties of a CMI Streaming Agent. The following image shows some of the CMI Streaming Agent properties:

### ▼ System Configuration Details

Service: CMI Streaming Agent ▼

Type: All Types ▼

Type	Name	Value
Engine	MaxLogFileSize	'5MB'
Engine	LogLevel	'DEBUG'
Agent	DataflowPullInterval	60
Agent	JVM	'-Xms256M -Xmx256M'
Agent	LogLevel	'DEBUG'
Agent	MaxLogFileSize	'10MB'
Agent	MaxNumberOfBackups	5
Scripts	LogLevel	'DEBUG'
Scripts	MaxFileSize	'5MB'
Scripts	MaxBackupIndex	5

You can configure the following CMI Streaming Agent properties:

Type	Property Name	Description
Engine	MaxLogFileSize	The maximum size of the log file that the engine can create. Default is 5 MB.
Engine	LogLevel	The log level for the engine.
Agent	DataflowPullInterval	The time interval after which the agent checks for updates in the task. Default is 60 seconds.
Agent	JVM	List of JVM properties for the agent. For example: [-Xms256M -Xmx256M]

Type	Property Name	Description
Agent	LogLevel	The log level for the agent.
Agent	MaxLogFileSize	Maximum size of the log files that an agent can create. Default is 10 MB.
Agent	MaxNumberOfBackups	Maximum number of backup log files for the agent. Default is 5.
Scripts	LogLevel	The log level of the scripts.
Scripts	MaxFileSize	The maximum file size after which the log rolls over and creates a new file. Default is 10 MB.
Scripts	MaxBackupIndex	Maximum number of backup files maintained after rolling over. Default is 5.

## Streaming Agent offline mode

You can run and monitor a streaming ingestion job when the CMI Streaming Agent is offline or not connected to the internet.

The Streaming Agent supports both online and offline modes of communication. In the offline mode, the streaming ingestion job continues to run even if the Streaming Agent does not communicate with the Informatica Intelligent Cloud Services for an extended period of time. The Streaming Agent continues to monitor the health and statistics of the ingestion tasks locally. When the Streaming Agent turns online and connects to the cloud services, it updates any configuration changes for the agent and tasks, as well as updates the health and statistics to the services.

To switch between the offline and online modes, you can use the command line utility provided by the Mass Ingestion Streaming service. Run the following command to start the command line utility:

```
<Informatica Secure Agent>/apps/Streaming_Ingestion_Agent/<version>/runcli.bat
```

The command line utility uses the command prompt `infa/stream>` and provides three groups of commands.

You can change the communication modes only through the command line utility. The Streaming Agent preserves the communication mode when the agent restarts.



The following table lists the commands of this command line utility:

Command	Description	Example
app-config	Shows the current configuration of the Streaming Agent application.	<pre> infa/stream :&gt;app-config deploy.pull.interval : 60 health.poll.interval : 30 minifi.ingester.file.location : ./conf siagent.communication.mode : Online siagent.monitoring.persist.dir : ../data siagent.statistics.post.batchsize : 720 siagent.statistics.post.concurrency : 60 siagent.status.persist.dir : ../data statistics.poll.interval : 30 </pre>
app-setconfig	<p>Use this command to configure the following properties:</p> <ul style="list-style-type: none"> <li>- siagent.communication.mode. Use to configure offline or online communication mode.</li> <li>- siagent.statistics.post.batchsize. Use to define the number of snapshots in a batch.</li> <li>- siagent.statistics.post.concurrency. Use to define the number of worker threads to post statistics.</li> </ul> <p>The --key and --value tokens are optional.</p>	<pre> infa/stream :&gt;app-setconfig --key siagent.statistics.post.batchsize -- value 20  or  infa/stream :&gt;app-setconfig siagent.statistics.post.batchsize 20 </pre>
app-status	<p>Shows the current status of the Streaming Agent.</p> <p>The health status code and health error message indicates the status of the agent (service) shown on the <b>Administrator</b>.</p> <p>uptime indicates the number of seconds since the Streaming Agent application is available.</p>	<pre> infa/stream :&gt;app-status health error message : No errors health status code : RUNNING(0) uptime : 67828 </pre>
app-statistics	<p>Shows metadata and status of overall statistics collection in the Streaming Agent.</p> <ul style="list-style-type: none"> <li>- collection interval. Interval of statistics collection, in seconds.</li> <li>- post interval. Frequency of statistics posted or attempted post.</li> <li>- max batch size. Maximum number of snapshots posted in a single http post.</li> <li>- last batch size. Number of snapshots in the last http post.</li> <li>- last time collected. Timestamp when any statistics were last collected.</li> <li>- last time posted. Timestamp when any statistics were last posted.</li> </ul>	<pre> infa/stream :&gt;app-statistics collection interval : 30 last batch size : 2 last time collected : 7/3/20 10:19:03 AM IST last time posted : 7/3/20 10:18:53 AM IST max batch size : 20 pending snapshots : 3 post interval : 30 </pre>

Command	Description	Example
clear	Clears the screen.	-
exit, quit	Quits the application.	-
help	Shows a summary of all the commands available.	<pre> infa/stream :&gt;help AVAILABLE COMMANDS Agent Application Commands app-config: Show agent application configuration app-setmode: Set the communication mode [Online/Offline] app-status: Show agent application status Built-In Commands clear: Clear the shell screen. exit, quit: Exit the shell. help: Display help about available commands. Streaming Ingestion Task Commands task-health: Show streaming ingestion task health task-list: Show streaming ingestion task list task-metadata: Show streaming ingestion task metadata </pre>
task-list	Shows the list of streaming ingestion jobs currently deployed on the Streaming Agent.	<pre> infa/stream :&gt;task-list 6e61e76f-2618-4292-ab3d-dd181f47ee91 ad5053c7-5ac2-493f-8cbb-a24900b61f71 </pre>
task-health	Shows the health status of all streaming ingestion jobs in the Streaming Agent. Use the options --name or --id to specify a job. If none are specified, all jobs are listed.	<pre> infa/stream :&gt;task-health --name aby_df4 processors : [{"id":"14a7a095-7fac-4fc3-ac5c-705369132516","status":"ERROR"}, {"id":"821e6730-3aed-4d3f-b875-45f424b6b963","status":"RUNNING"}] status : ERROR timestamp : Sat May 09 06:04:08 IST 2020 infa/stream :&gt;task-health 6e61e76f-2618-4292-ab3d-dd181f47ee91 processors : [{"id":"2a0b8715-aa7a-46c5-9d6a-6a356f5a0102","status":"ERROR"}, {"id":"1172f3a8-35dd-41ef-be4b-bc0cf37e3794","status":"RUNNING"}] status : ERROR timestamp : Sat May 09 06:04:08 IST 2020 ad5053c7-5ac2-493f-8cbb-a24900b61f71 processors : [{"id":"14a7a095-7fac-4fc3-ac5c-705369132516","status":"ERROR"}, {"id":"821e6730-3aed-4d3f-b875-45f424b6b963","status":"RUNNING"}] status : ERROR timestamp : Sat May 09 06:04:08 IST 2020 </pre>

Command	Description	Example
task-metadata	Shows the metadata of all streaming ingestion jobs in the Streaming Agent. Use the options <code>--name</code> or <code>--id</code> to specify a job. If none are specified, all jobs are listed.	<pre> infa/stream :&gt;task-metadata --name aby_df4 id : ad5053c7-5ac2-493f-8cbb- a24900b61f71 name : aby_df4 runId : 9071 version : 1 infa/stream :&gt;task-metadata 6e61e76f-2618-4292-ab3d-dd181f47ee91 id : 6e61e76f-2618-4292-ab3d- dd181f47ee91 name : aby_df2 runId : 9069 version : 8 ad5053c7-5ac2-493f-8cbb-a24900b61f71 id : ad5053c7-5ac2-493f-8cbb- a24900b61f71 name : aby_df4 runId : 9071 version : 1 </pre>
task-statistics	Shows the statistics details of all streaming ingestion jobs in the Streaming Agent. Use the options <code>--name</code> or <code>--id</code> to specify a job. If none are specified, all jobs are listed.	<pre> infa/stream :&gt;task-statistics --name aby_df1 dataflow name : aby_df1 last time collected : 1590861803731 last time posted : 1590861806091 infa/stream :&gt;task-statistics 7b7d3c09-df43-482f-b6c8-8dd80187e6d7 dataflow name : aby_df2 last time collected : 1590861770731 last time posted : 1590861741132 decfad0a-20df-4226-84f9-1ff1ab6ef96a dataflow name : aby_df1 last time collected : 1590861768730 last time posted : 1590861771054 </pre>

## Online mode to Offline mode

By default, the Streaming Agent is in online mode.

To change the Streaming Agent to offline mode:

1. Launch the command line utility using the following command:

```

In Windows:
<Informatica_Secure_Agent>/apps/Streaming_Ingestion_Agent/<version>/runcli.bat

or

In UNIX:
runcli.sh

```

2. Set the Streaming Agent to offline mode:

```
app-setconfig --key siagent.communication.mode --value Offline
or
app-setconfig siagent.communication.mode Offline
```

The Streaming Agent stops sending health updates and statistics of any streaming ingestion job.

## Offline mode to Online mode

To change the Streaming Agent to online mode:

1. Launch the command line utility using the following command:

```
In Windows:
<Informatica_Secure_Agent>/apps/Streaming_Ingestion_Agent/<version>/runcli.bat

or

In UNIX:
runcli.sh
```

2. Set the Streaming Agent to online mode:

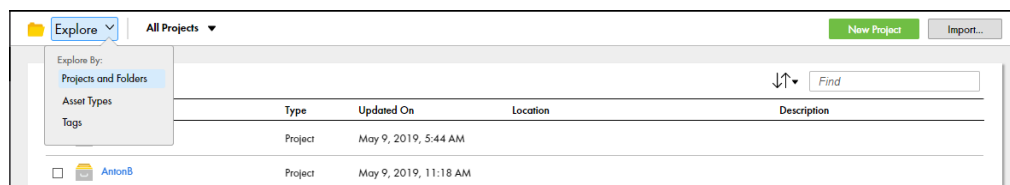
```
app-setconfig --key siagent.communication.mode --value Online
or
app-setconfig siagent.communication.mode Online
```

The Streaming Agent starts sending health updates of all the streaming ingestion jobs and the updates appear in the **Monitoring** page. It starts sending statistics of all the streaming ingestion jobs including the statistics backlog collected while it was offline to the service. It also synchronizes the updates to the streaming ingestion jobs or adds the new streaming ingestion job deployed while it was offline to the service.

# Creating projects and project folders

Create a project to contain your ingestion task assets. You can create one or more folders under the project to logically organize your assets. However, you cannot create subfolders under a folder.

1. In the Mass Ingestion service, open the **Explore** page.
2. If the **Explore** page shows objects other than projects, select **Projects and Folders** in the **Explore** menu.



3. To create a project, click **New Project**.
4. In the **New Project Properties** dialog box, enter a project name up to 255 characters in length. You can also enter an optional description of the project.

A project name cannot contain the following characters: # ? ` | { } " ^ & [ ] / \ .

5. If you want to add a folder under the project, select the project and click **New Folder**.
6. In the **New Folder Properties** dialog box, enter a folder name up to 255 characters in length. You can also enter an optional description of the project.

A folder name cannot contain the following characters: # ? ` | { } " ^ & [ ] / \ .

When you define an ingestion task, you must specify the project or project folder location to contain the task definition.

## Editing your user profile

Your user profile contains the details of your Informatica Intelligent Cloud Services user account.

You can update the following information in your profile:

- First and last name
- Job title
- Email address
- Phone number
- Time zone (used in the job execution time stamps on the **All Jobs**, **Running Jobs**, **My Jobs**, **Import/Export Logs**, and **My Import/Export Logs** pages)
- Password
- Security question and answer

**Note:** If you use SAML to sign on to Informatica Intelligent Cloud Services and your organization administrator has enabled SAML group and role mapping on the **SAML Setup** page in Administrator, you can only update the time zone. The other attributes are updated directly from your enterprise directory each time you log into Informatica Intelligent Cloud Services.

1. Click the **User** icon in the top right corner of the Informatica Intelligent Cloud Services window and then select **Profile**.
2. On the **Profile** page, add or edit personal information such as your name, job title, phone number, and time zone.
3. To update your email address, click **Update Email**.  
Informatica Intelligent Cloud Services sends a verification email to your new email address. The email contains a link that is valid for 24 hours. When you click the link in the email, the new address is verified, and it appears in your profile. If the link expires, you can resend the verification email.
4. Optionally, change your password or security question.
5. Click **Save**.

## CHAPTER 3

# Connectors and Connections

Connections provide access to data in cloud and on-premises applications, platforms, databases, and flat files. Before you can define a connection, ensure that the connector for the source or target type is installed in Informatica Intelligent Cloud Services.

If multiple connectors are available for a source or target type, get the one that your ingestion type supports. Some connectors are pre-installed. If you need a connector that is not pre-installed, you can download it from the **Add-On Connectors** page in Administrator.

## Mass Ingestion connectors

You must have the correct connectors to create connections for the sources and targets you use in your ingestion tasks.

Before you can define connections, your organization administrator must ensure that the source and target connectors that the organization uses are installed. Also, you must enable connectors for your runtime environment.

For more information about connectors and connections, see "Licenses," "Runtime Environment," and "Connections" in the Administrator help.

## Mass Ingestion Applications connectors

Before you define a connection for application ingestion tasks, verify that the connectors for your source and target types are available in Informatica Intelligent Cloud Services.

The following table lists the connectors that Mass Ingestion Applications requires to connect to a source or target:

Source or target type	Connector	Use for
Adobe Analytics	Adobe Analytics Mass Ingestion	Sources in initial load operations
Amazon Redshift	Amazon Redshift V2	Targets in initial load, incremental load, and combined initial and incremental load operations
Amazon S3	Amazon S3 V2	Targets in initial load, incremental load, and combined initial and incremental load operations
Apache Kafka	Kafka	Targets in incremental load operations

Source or target type	Connector	Use for
Databricks Delta	Databricks Delta	Targets in initial load, incremental load, and combined initial and incremental load operations
Google Analytics	Google Analytics Mass Ingestion	Sources in initial load, incremental load, and combined initial and incremental load operations
Google BigQuery	Google BigQuery V2	Targets in initial load, incremental load, and combined initial and incremental load operations
Google Cloud Storage	Google Cloud Storage V2	Targets in initial load, incremental load, and combined initial and incremental load operations
Marketo	Marketo V3	Sources in initial load operations
Microsoft Azure Data Lake Storage Gen2	Microsoft Azure Data Lake Storage Gen2	Targets in initial load, incremental load, and combined initial and incremental load operations
Microsoft Azure Synapse Analytics <sup>1</sup>	Microsoft Azure Synapse Analytics Database Ingestion	Targets in initial load, incremental load, and combined initial and incremental load operations
Microsoft Dynamics 365	Microsoft Dynamics 365 Mass Ingestion	Sources in initial load, incremental load, and combined initial and incremental load operations
NetSuite	NetSuite Mass Ingestion	Sources in initial load, incremental load, and combined initial and incremental load operations
Oracle Fusion Cloud Applications	Oracle Fusion Cloud Mass Ingestion	Sources in initial load operations
Salesforce	Salesforce Mass Ingestion	Sources in initial load, incremental load, and combined initial and incremental load operations
SAP ECC	SAP ODP Extractor	Sources in initial load, incremental load, and combined initial and incremental load operations
SAP S/4HANA	SAP ODP Extractor	Sources in initial load, incremental load, and combined initial and incremental load operations
ServiceNow	ServiceNow Mass Ingestion	Sources in initial load, incremental load, and combined initial and incremental load operations
Snowflake	Snowflake Cloud Data Warehouse V2	Targets in initial load, incremental load, and combined initial and incremental load operations
Workday	Workday Mass Ingestion	Sources in initial load, incremental load, and combined initial and incremental load operations
Zendesk	Zendesk Mass Ingestion	Sources in initial load, incremental load, and combined initial and incremental load operations
<p>1. For Microsoft Azure Synapse Analytics targets, Mass Ingestion Applications uses Microsoft Azure SQL Data Lake Storage Gen2 to store staging files. Before you configure a connection for a Microsoft Azure Synapse Analytics target, ensure that you have Microsoft Azure SQL Data Lake Storage Gen2 installed.</p>		

## Mass Ingestion Databases connectors

Before you begin defining connections for database ingestion tasks, verify that the connectors for your source and target types are available in Informatica Intelligent Cloud Services.

The following table lists the connectors that Mass Ingestion Databases requires to connect to a source or target that can be configured in a database ingestion task:

Source or target type	Connector	Use for
Amazon Redshift	Amazon Redshift V2	Targets in initial load, incremental load, and initial and incremental load operations
Amazon S3	Amazon S3 V2	Targets in initial load and incremental load operations
Databricks Delta	Databricks Delta	Targets in initial load, incremental load, and initial and incremental load operations
Db2 for i	Db2 for i Database Ingestion	Sources in initial load, incremental load, and initial and incremental load operations
Db2 for Linux, UNIX, and Windows	Db2 for LUW Database Ingestion	Sources in initial load operations
Db2 for z/OS	Db2 for zOS Database Ingestion	Sources in initial load and incremental load operations
Flat file	No connector required	Targets in initial load and incremental load operations.
Google BigQuery	Google BigQuery V2	Targets in initial load, incremental load, and initial and incremental load operations
Google Cloud Storage	Google Cloud Storage V2	Targets in initial load and incremental load operations
Kafka, including Apache Kafka, Confluent Kafka, Amazon Managed Streaming for Apache Kafka, and Kafka-enabled Azure Event Hubs	Kafka	Targets in incremental load operations
Microsoft Azure Data Lake Storage Gen2	Microsoft Azure Data Lake Storage Gen2	Targets in initial load and incremental load operations
Microsoft Azure SQL Database	Microsoft SQL Server	Sources in initial load operations
Microsoft Azure Synapse Analytics <sup>1</sup>	Microsoft Azure Synapse Analytics Database Ingestion	Targets in initial load, incremental load, and initial and incremental load operations
Microsoft SQL Server	Microsoft SQL Server	Sources in initial load, incremental load, and initial and incremental load operations
MongoDB	MongoDB Mass Ingestion	Sources in initial load and incremental load operations



Source or target type	Connector	Use for
MySQL, including RDS for MySQL	MySQL	Sources in initial load operations
Netezza	Netezza	Sources in initial load operations
Oracle, including RDS for Oracle	Oracle Database Ingestion	Sources in initial load, incremental load, and initial and incremental load operations Targets in initial load, incremental load, and initial and incremental load operations
PostgreSQL, including RDS for PostgreSQL and Amazon Aurora PostgreSQL	PostgreSQL	Sources in initial load and incremental load operations
SAP HANA	SAP HANA Database Ingestion	Sources in initial load and incremental load operations
Snowflake	Snowflake Data Cloud Connector	Targets in initial load, incremental load, and initial and incremental load operations
Teradata Data Warehouse Appliance	Teradata	Sources in initial load operations
1. For this target type, Mass Ingestion Databases uses Microsoft Azure SQL Data Lake Storage Gen2 to store staging files. Ensure that you have Microsoft Azure SQL Data Lake Storage Gen2 installed.		

## Mock connectors

Mass Ingestion Databases supports mock, or sample, connections for some of the sources and targets. Use mock connections to learn how to create initial load database ingestion tasks without creating real connections to the database.

A mock connector does not connect to a real database. Instead, a source mock connector uses flat files with sample data. A target mock connector reports the information about processed source data to Mass Ingestion Databases user interface, but it does not write any data to the target.

The sample connections appear in the source and target connection lists in Mass Ingestion Databases if you have the MockConnector license.

The following table lists mock connections that you can use for Mass Ingestion Databases sources and targets:

Connection name	Source or Target
Sample Oracle Connection	Source
Sample SQL Server Connection	Source
Sample S3 Connection	Target
Sample ADLS Gen2 Connection	Target

**Note:** You must use sample connections for both the source and target databases. You cannot use a sample connection for only one of them, for example, for the source but not for the target.

## Source data

The source data for sample connections is stored in CVS files in the following directory:

```
Secure_Agent_installation/downloads/package-MockConnector.version/package/sampleData/  
source/database_type/
```

Each file represents a single table. A mock table name matches the file name. The first line in a file determines column headers, and the subsequent lines contain row data.

## Mass Ingestion Files connectors

Before you define connections for file ingestion tasks, ensure that you have a license for the connectors that Mass Ingestion Files requires for the source and target types.

The following table lists the connectors that a file ingestion task supports based on the source and target types:

Source or target name	Connector	Source or target type
Local folder	No connector required	Source and Target
Advanced FTP	Advanced FTP V2 (add-on)	Source and Target
Advanced FTPS	Advanced FTPS V2 (add-on)	Source and Target
Advanced SFTP	Advanced SFTP V2 (add-on)	Source and Target
Amazon S3	Amazon S3 V2 (add-on)	Source and Target
Amazon Redshift	Amazon Redshift V2 (add-on)	target
Databricks Delta	Databricks Delta (add-on)	Source and Target
Google BigQuery	Google BigQuery V2 (add-on)	Target
Google Cloud Storage	Google Cloud Storage V2 (add-on)	Source and Target
Hadoop Files	Hadoop Files V2 (add-on)	Source and Target
Microsoft Azure Blob Storage	Microsoft Azure Blob Storage V3 (add-on)	Source and Target
Microsoft Azure Data Lake Store	Microsoft Azure Data Lake Store Gen2 (add-on)	Source and Target
Microsoft Azure Data Lake Store	Microsoft Azure Data Lake Store V3 (add-on)	Source and Target
Microsoft Azure Synapse SQL	Microsoft Azure Synapse SQL (add-on)	Target
Snowflake	Snowflake Cloud Data Warehouse V2 (add-on)	Target

## Mass Ingestion Streaming connectors

Before you define connections for the streaming ingestion tasks, ensure that you have a license for the required connectors for your source and target types.

The following table lists the connectors that a streaming ingestion task supports based on the source and target type:

Source or target name	Connector	Source or target type
Amazon Kinesis Data Firehose	Amazon Kinesis (add-on)	Target
Amazon Kinesis Data Streams	Amazon Kinesis (add-on)	Source and Target
Amazon Managed Streaming for Apache Kafka (Amazon MSK)	Kafka (add-on)	Source and Target
Amazon S3	Amazon S3 V2 (add-on)	Target
AMQP	AMQP (add-on)	Source
Apache Kafka	Kafka (add-on)	Source and Target
Azure Event Hubs Kafka	Kafka (add-on)	Source
Confluent Kafka	Kafka (add-on)	Source and Target
Databricks Delta	Databricks Delta (add-on)	Target
Flat file	No connector required	Source and Target
Google Cloud Storage	Google Cloud Storage V2 (add-on)	Target
Google PubSub	Google PubSub (add-on)	Source and Target
JDBC V2	JDBC V2 (add-on)	Target
JMS	JMS (add-on)	Source
Microsoft Azure Data Lake Storage	Azure Data Lake Store Gen2 (add-on)	Target
Microsoft Azure Event Hub	Azure Event Hubs (add-on)	Target
MQTT	MQTT (add-on)	Source
OPC UA	OPCUA (add-on)	Source
REST V2	REST V2 (add-on)	Source

**Note:** While importing a streaming ingestion task, both read and write connection types appear in the drop-down list on the **Import Review** page. You can also see connections to connectors that are not supported by Mass Ingestion Streaming.

# Mass Ingestion connection properties

Connections provide access to data in cloud and on-premises applications, platforms, databases, and flat files. Connection definitions include the location of the source or target, the runtime environment, and the other properties specific to the connection type.

Before you can create a connection, ensure that the correct connectors for your sources and targets are available in Informatica Intelligent Cloud Services. The supported connectors vary by type of ingestion task.

To create a connection or search for an existing connection, use the Administrator service.

After you configure connection properties, the connection becomes available for use within the organization.

## Configuring a connection

Configure a source or target connection on the **Connections** page in Administrator.

1. In Administrator, select **Connections**.
2. On the **Connections** page, click **New Connection**.
3. Configure the following connection details:

Property	Description
Connection Name	The name of the connection. Each connection name must be unique within the organization. Connection names can contain alphanumeric characters, spaces, and the following special characters: _ . + - Maximum length is 100 characters. Connection names are not case sensitive.
Description	A description of the connection.
Type	The type of connection, such as Amazon S3.

After you select the connection type, additional properties that are specific to that type appear.

4. Configure the connection-specific properties.  
For example, if you are configuring an Amazon S3 connection, enter the Amazon S3 connection properties. Click the Help icon for a description of each connection property.
5. To test the connection, click **Test Connection**.  
The results of the test are displayed at the top of the page.  
If a connection fails, contact the database administrator, or recheck your settings and verify that the selected runtime environment has the status of Up and Running.
6. Click **Save** to save the connection.

## Adobe Analytics Mass Ingestion connection properties

When you set up an Adobe Analytics Mass Ingestion connection, you must configure the connection properties.

Adobe Analytics uses a JSON Web Token (JWT) to authenticate the Adobe Analytics Mass Ingestion connection. To use an Adobe Analytics Mass Ingestion connection, you must create a Service Account Integration on Adobe Developer Console and then specify the service integration details in the connection

properties. For more information about creating a Service Account Integration on Adobe Developer Console, see the [Adobe documentation](#).

The following table describes the connection properties for an Adobe Analytics Mass Ingestion connection:

Connection property	Description
Runtime Environment	Name of the runtime environment where you want to run the ingestion tasks. You must specify a Secure Agent as the runtime environment. <b>Note:</b> You cannot run application ingestion tasks on a Hosted Agent or serverless runtime environment.
Client ID	Client ID of the Service Account that you created on Adobe Developer Console.
Client Secret	Client secret of the Service Account that you created on Adobe Developer Console.
Technical Account ID	Technical account ID of the Service Account.
Organization ID	Organization ID of the Service Account.
Private Key	Private key that is generated when you create the Service Account Integration. The private key is required to generate the JWT.
IMS Host	Base URL of Adobe Identity Management System (IMS). The default value is: <code>ims-na1.adobelogin.com</code>
IMS Exchange	Exchange URL of IMS. The connection use the JWT to obtain an access token from Adobe by making a POST request to the exchange URL. The default value is: <code>https://ims-na1.adobelogin.com/ims/exchange/jwt</code>

## Advanced FTP V2 connection properties

When you set up an Advanced FTP V2 connection, you must configure the connection properties.

The following table describes the Advanced FTP V2 connection properties:

Connection property	Description
Connection Name	The name of the connection. The name is not case sensitive and must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters:~`!\$%^&*()-+= { }   \ : ; " ' < , > . ? /
Description	Optional. The description of the connection. The description cannot exceed 4,000 characters.
Type	Select the <b>Advanced FTP V2</b> connection type.
Runtime Environment	The name of the runtime environment where you want to run the tasks. Specify a Secure Agent.
Host	The host name or IP address of the FTP server.

Connection property	Description
Port	The port number to use for connecting to the FTP server. If left blank, the default port number 21 is used.
Username	User name to connect to the FTP server.
Password	Password to connect to the FTP server.
Folder Path	The directory to use after connecting to the FTP server.
Use passive mode	<p>Indicates whether the connection uses <b>Passive</b> or <b>Active</b> mode. Specify <b>Yes</b> to use <b>Passive</b> mode. Specify <b>No</b> to use <b>Active</b> mode.</p> <p>The default value is <b>Yes</b>.</p> <p>In Passive mode, the server does not need to connect back to a port on the connection client, which is a firewall-friendly mode. If you have problems with connecting to the server, you might want to change the mode to Passive by selecting <b>Yes</b> for this option.</p> <p>In Active mode, the server attempts to connect back to a port on the connection client to perform the data transfer.</p>
Data Connection Start Port	The starting port number to use for the data connection.
Data Connection End Port	The ending port number to use for the data connection.
Timeout	The number of seconds to wait when attempting to connect to the server. A timeout occurs if the connection cannot be established in the specified amount of time. If left blank, the default value of 120 seconds is used.
Connection Retry Attempts	The number of times to connect to retry the FTP connection if a connection cannot be established. This setting is used for both the initial connection and any reconnect attempts due to lost connections. If left blank, no retries will be attempted.
Connection Retry Interval	<p>The number of seconds to wait between each connection retry attempt.</p> <p><b>Note:</b> For instance, if you want to retry to connect up to 10 times with a five second delay between retries, then specify <b>10</b> for the <b>Connection Retry Attempts</b> and <b>5</b> for the <b>Connection Retry Interval</b>.</p>
Control Encoding	If left blank, the connection uses the ISO standard ISO-8859-1. If supported by the server, other encodings such as UTF-8 can be specified to support international characters.
List Parser	The list parser to use for the server connection. If the field is left blank, the Advanced FTP V2 Connector attempts to use the MLSD parser. If the MLSD parser is not supported by the server, the UNIX parser is used. If you experience problems listing directories, select a different list parser.

Connection property	Description
Date Format	This date format is applied if the server returns a date that is different from the selected list parser default. If your location requires a different date format (for example, d MMM yyyy), specify the date format in this field. Not all list parsers support the date format setting. List parsers that do not support the date format setting ignores any user specified value.
Recent Date Format	Specify the date format to use when parsing the recent last modified date for each file. The recent date format applies in UNIX-based systems and appears on entries less than a year old. If your location requires a specific date format (for example, d MMM HH:mm), specify that pattern in this field. Not all list parsers support the recent date format setting. List parsers that do not support the recent date format setting ignores any user-specified value.

## Advanced FTPS V2 connection properties

When you set up an Advanced FTPS V2 connection, you must configure the connection properties.

The following table describes the Advanced FTPS V2 connection properties:

Connection property	Description
Connection Name	The name of the connection. The name is not case sensitive and must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters:~ `! \$ % ^ & * ( ) - + = { [ ]   \ ; ' " < , > . ? /
Description	Optional. The description of the connection. The description cannot exceed 4,000 characters.
Type	Select the <b>Advanced FTPS V2</b> connection type.
Runtime Environment	The name of the runtime environment that contains the Secure Agent that you want to run the tasks.
Host	The host name or IP address of the server.
Port	The port number to use for connecting to the server. If left blank, the default port number is 21.
Username	User name to connect to the FTPS server.
Password	Password to connect to the FTPS server.
Folder Path	The directory to use after connecting to the server.
Use passive mode	Indicates whether the connection uses <b>Passive</b> or <b>Active</b> mode. Specify <b>Yes</b> to use <b>Passive</b> mode. Specify <b>No</b> to use <b>Active</b> mode. The default value is <b>Yes</b> . In Passive mode, the server does not need to connect back to a port on the connection client, which is a firewall-friendly mode. If you have problems with connecting to the server, you might want to change the mode to Passive by selecting <b>Yes</b> for this option. In Active mode, the server attempts to connect back to a port on the connection client to perform the data transfer.

Connection property	Description
Data Connection Start Port	The starting port number to use for the data connection.
Data Connection End Port	The ending port number to use for the data connection.
Timeout	The number of seconds to wait when attempting to connect to the server. A timeout occurs will if the connection cannot be established in the specified amount of time. If left blank, the default value of 120 seconds is used.
Connection Retry Attempts	The number of times to connect to retry the Advanced FTP V2 connection if a connection cannot be established. This setting is used for both the initial connection and any reconnect attempts due to lost connections. If left blank, no retries will be attempted.
Connection Retry Interval	The number of seconds to wait between each connection retry attempt. <b>Note:</b> For instance, if you want to retry to connect up to 10 times with a five second delay between retries, then specify <b>10</b> for the <b>Connection Retry Attempts</b> and <b>5</b> for the <b>Connection Retry Interval</b> .
Control Encoding	If left blank, the connection uses the ISO standard ISO-8859-1. If supported by the server, other encodings like UTF-8 can be specified to support international characters.
Trusted Server	Specify whether the FTPS server is a trusted server. The Advanced FTP V2 Connector only supports a trusted server.
List Parser	The list parser to use for the server connection. If the field is empty, the Advanced FTP V2 Connector tries to use the MLSD parser. If the server does not support the MLSD parser, the connector uses the UNIX parser. If you experience problems listing directories, select a different list parser.
Date Format	This date format is applied if the server returns a date that is different from the selected list parser default. If your location requires a different date format (for example, d MMM yyyy), specify the date format in this field. Not all list parsers support the date format setting. List parsers that do not support the date format setting ignores any user specified values.
Recent Date Format	Specify the date format to use when parsing the recent last modified date for each file. The recent date format applies in UNIX-based systems and appears on entries less than a year old. If your location requires a specific date format (for example, d MMM HH:mm), specify that pattern in this field. Not all list parsers support the recent date format setting. List parsers that do not support the recent date format setting ignores any user-specified values.
Connection Type	Indicates if the connection type is IMPLICIT_SSL or EXPLICIT_SSL. - IMPLICIT_SSL. The connection automatically starts as an SSL connection. - EXPLICIT_SSL. After initial authentication with the FTPS server, the connection is encrypted with SSL or TLS depending on the security protocol you select. Default is IMPLICIT_SSL.
SecurityProtocol	Indicates whether SSL or TLS is used for EXPLICIT_SSL connections. Default is SSL.
Key Store File	The path and file name of the keystore file. The keystore file contains the certificates to authenticate the FTPS server.
Key Store Password	The password for the keystore file required to access the Trusted Server Certificate Store.



Connection property	Description
Key Alias	The alias of the individual key.
Key Store Type	Indicates if the type of the keystore is Java KeyStore (JKS) or Public Key Cryptology Standard (PKCS12). Default is JKS.

## Advanced SFTP V2 connection properties

When you set up an Advanced SFTP V2 connection, you must configure the connection properties.

The following table describes the Advanced SFTP V2 connection properties:

Connection property	Description
Connection Name	The name of the connection. The name is not case sensitive and must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ & * ( ) - + = { [ ]   \ : ; " ' < , > . ? /
Description	Optional. The description of the connection. The description cannot exceed 4,000 characters.
Type	Select the <b>Advanced SFTP V2</b> connection type.
Runtime Environment	The name of the runtime environment that contains the Secure Agent that you want to run the tasks.
Host	The host name or IP address of the server.
Port	The port number to use for connecting to the server. If left blank, the default port number is 21.
Username	User name to connect to the SFTP server.
Password	Password to connect to the SFTP server.
Folder Path	The directory to use after connecting to the server.
Timeout	The number of seconds to wait when attempting to connect to the server. A timeout occurs if the connection cannot be established in the specified amount of time. If left blank, the default value of 120 seconds is used.
Connection Retry Attempts	The number of times to connect to retry the SFTP connection if a connection cannot be established. This setting is used for both the initial connection and any reconnect attempts due to lost connections. If left blank, no retries will be attempted.
Connection Retry Interval	The number of seconds to wait between each connection retry attempt. <b>Note:</b> For instance, if you want to retry to connect up to 10 times with a five second delay between retries, then specify <b>10</b> for the <b>Connection Retry Attempts</b> and <b>5</b> for the <b>Connection Retry Interval</b> .

Connection property	Description
Private Key File	The name of the SSH private key file along with the path where the file is stored. Ensure that the file path is on the machine that hosts the Secure Agent. For example, <code>C:/SSH/my_keys/key.ppk</code>
Private Key Passphrase	Specify the passphrase to encrypt the SSH private key.
Use Curve Kex Algorithm	Enable additional key exchange algorithms such as curve, and keyed-hash algorithm such as, - hmac-sha2-512, and -hmac-sha2-256.
Use File Integration Proxy Server	The connector connects to the SFTP server through the file integration proxy server. <b>Note:</b> <ul style="list-style-type: none"> <li>- You must have the File Integration Service license to use this option.</li> <li>- You must define a proxy server in File Servers.</li> <li>- If you don't have the File Integration Service proxy, you need to use the agent proxy through the proxy.ini file.</li> </ul>
Proxy Server Host Name	Host name or IP address of the outgoing File Integration Service proxy server.
Proxy Server Port	Port number of the outgoing File Integration Service proxy server.

## Amazon Redshift V2 connection properties

When you set up an Amazon Redshift V2 connection, you will need to configure the connection properties.

The following table describes the Amazon Redshift V2 connection properties:

Connection property	Description
Runtime Environment	Name of the runtime environment where you want to run the tasks. You cannot run an application ingestion task, database ingestion task, file ingestion task, or streaming ingestion task on a Hosted Agent or serverless runtime environment.
Username	User name of the Amazon Redshift account.
Password	Password for the Amazon Redshift account.
Access Key ID	Access key to access the Amazon S3 staging bucket. Provide the access key value based on the following authentication methods: <ul style="list-style-type: none"> <li>- Basic authentication: Provide the actual access key value.</li> <li>- IAM authentication: Do not provide the access key value.</li> <li>- Temporary security credentials via assume role: Provide access key of an IAM user with no permissions to access the Amazon S3 staging bucket.</li> <li>- Assume role for EC2: Do not provide the access key value.</li> </ul> <b>Note:</b> If you want to use the connection for an application ingestion or database ingestion task, you must use the basic authentication method to provide the access key value.

Connection property	Description
Secret Access Key	<p>Secret access key to access the Amazon S3 staging bucket.</p> <p>The secret key is associated with the access key and uniquely identifies the account. Provide the access key value based on the following authentication methods:</p> <ul style="list-style-type: none"> <li>- Basic authentication: Provide the actual access secret value.</li> <li>- IAM authentication: Do not provide the access secret value.</li> <li>- Temporary security credentials via assume role: provide access secret of an IAM user with no permissions to access Amazon S3 staging bucket.</li> <li>- Assume role for EC2: Do not provide the access secret value.</li> </ul> <p><b>Note:</b> If you want to use the connection for an application ingestion or database ingestion task, you must provide the actual access secret value.</p>
IAM Role ARN	<p>The Amazon Resource Number (ARN) of the IAM role assumed by the user to use the dynamically generated temporary security credentials.</p> <p>Set the value of this property if you want to use the temporary security credentials to access the Amazon S3 staging bucket.</p> <p><b>Note:</b> You cannot use the temporary security credentials in streaming ingestion tasks.</p> <p>For more information about how to obtain the ARN of the IAM role, see the AWS documentation.</p>
External Id	<p>Optional. Specify the external ID for a more secure access to the Amazon S3 bucket when the Amazon S3 staging bucket is in a different AWS account.</p>
Use EC2 Role to Assume Role	<p>Optional. Select the check box to enable the EC2 role to assume another IAM role specified in the IAM Role ARN option.</p> <p><b>Note:</b> The EC2 role must have a policy attached with a permission to assume an IAM role from the same or different account.</p> <p>By default, the Use EC2 Role to Assume Role check box is not selected.</p>
Master Symmetric Key	<p>Optional. Provide a 256-bit AES encryption key in the Base64 format when you enable client-side encryption. You can generate a key using a third-party tool.</p>
JDBC URL	<p>The URL of the Amazon Redshift V2 connection.</p> <p>Enter the JDBC URL in the following format: <code>jdbc:redshift://&lt;amazon_redshift_host&gt;:&lt;port_number&gt;/&lt;database_name&gt;</code></p>

Connection property	Description
Cluster Region	<p>Optional. The AWS cluster region in which the bucket you want to access resides.</p> <p>Select a cluster region if you choose to provide a custom JDBC URL that does not contain a cluster region name in the <b>JDBC URL</b> connection property.</p> <p>If you specify a cluster region in both <b>Cluster Region</b> and <b>JDBC URL</b> connection properties, the agent ignores the cluster region that you specify in the <b>JDBC URL</b> connection property.</p> <p>To use the cluster region name that you specify in the <b>JDBC URL</b> connection property, select <b>None</b> as the cluster region in this property.</p> <p>Select one of the following cluster regions:</p> <ul style="list-style-type: none"> <li>- None</li> <li>- Asia Pacific(Mumbai)</li> <li>- Asia Pacific(Seoul)</li> <li>- Asia Pacific(Singapore)</li> <li>- Asia Pacific(Sydney)</li> <li>- Asia Pacific(Tokyo)</li> <li>- Asia Pacific(Hong Kong)</li> <li>- AWS GovCloud (US)</li> <li>- AWS GovCloud (US-East)</li> <li>- Canada(Central)</li> <li>- China(Beijing)</li> <li>- China(Ningxia)</li> <li>- EU(Ireland)</li> <li>- EU(Frankfurt)</li> <li>- EU(Paris)</li> <li>- EU(Stockholm)</li> <li>- South America(Sao Paulo)</li> <li>- Middle East(Bahrain)</li> <li>- US East(N. Virginia)</li> <li>- US East(Ohio)</li> <li>- US West(N. California)</li> <li>- US West(Oregon)</li> </ul> <p>Default is <b>None</b>. You can only read data from or write data to the cluster regions supported by AWS SDK used by the connector.</p>
Customer Master Key ID	<p>Optional. Specify the customer master key ID generated by AWS Key Management Service (AWS KMS) or the ARN of your custom key for cross-account access.</p> <p>You must generate the customer master key ID for the same region where your Amazon S3 staging bucket resides. You can either specify the customer-generated customer master key ID or the default customer master key ID.</p>

## Amazon Kinesis connection properties

The Amazon Kinesis connection is a messaging connection. Use the Amazon Kinesis connection to access Amazon Kinesis Data Streams or Amazon Kinesis Data Firehose as targets.

## Amazon Kinesis Firehose connection properties

When you set up an Amazon Kinesis Firehose connection, you must configure the connection properties.

The following table describes the Amazon Kinesis Firehose connection properties:

Property	Description
Connection Name	<p>Name of the connection.</p> <p>The name is not case sensitive and must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters:</p> <p>~ ` ! \$ % ^ &amp; * ( ) - + = { [ ]   \ : ; " ' &lt; , &gt; . ? /</p>
Description	<p>Optional. Description that you can use to identity the connection.</p> <p>The description cannot exceed 4,000 characters.</p>
Type	<p>The Amazon Kinesis connection type.</p> <p>If you do not see the Amazon Kinesis connection type, go to the <b>Add-On Connectors</b> page to enable the connector.</p>
Runtime Environment	Name of the runtime environment where you want to run the tasks.
Service	The type of Kinesis Service that you want to use. Select <b>Kinesis Firehose</b> .
AWS Access Key ID	The access key ID of the Amazon AWS user account.
AWS Secret Access Key	The secret access key for the Amazon AWS user account.
Region	<p>Region where the endpoint for your service is available. You can select one of the following values:</p> <ul style="list-style-type: none"> <li>- us-east-2. Indicates the US East (Ohio) region.</li> <li>- us-east-1. Indicates the US East (N. Virginia) region.</li> <li>- us-west-1. Indicates the US West (N. California) region.</li> <li>- us-west-2. Indicates the US West (Oregon) region.</li> <li>- ap-northeast-1. Indicates the Asia Pacific (Tokyo) region.</li> <li>- ap-northeast-2. Indicates the Asia Pacific (Seoul) region.</li> <li>- ap-northeast-3. Indicates the Asia Pacific (Osaka-Local) region.</li> <li>- ap-south-1. Indicates the Asia Pacific (Mumbai) region.</li> <li>- ap-southeast-1. Indicates the Asia Pacific (Singapore) region.</li> <li>- ap-southeast-2. Indicates the Asia Pacific (Sydney) region.</li> <li>- ca-central-1. Indicates the Canada (Central) region.</li> <li>- cn-north-1. Indicates the China (Beijing) region.</li> <li>- cn-northwest-1. Indicates the China (Ningxia) region.</li> <li>- eu-central-1. Indicates the EU (Frankfurt) region.</li> <li>- eu-west-1. Indicates the EU (Ireland) region.</li> <li>- eu-west-2. Indicates the EU (London) region.</li> <li>- eu-west-3. Indicates the EU (Paris) region.</li> <li>- sa-east-1. Indicates the South America (São Paulo) region.</li> <li>- us-gov-west-1. Indicates AWS GovCloud (US-West) region.</li> <li>- us-gov-east-1. Indicates AWS GovCloud (US-East) region.</li> </ul> <p>A streaming ingestion task does not support ap-northeast-3 region.</p>
Connection TimeOut (ms)	<p>Optional. Number of milliseconds that the Mass Ingestion service waits to establish a connection to the Kinesis Firehose after which it times out.</p> <p>Default is 10,000 milliseconds.</p>

Property	Description
Authentication Type	<p>The type of authentication.</p> <p>Select one of the following values:</p> <ul style="list-style-type: none"> <li>- AWS Credential Profile</li> <li>- Cross-account IAM Role</li> </ul> <p>Default is AWS Credential Profile.</p> <p>Cross-account IAM Role is not applicable for a streaming ingestion task.</p>
AWS Credential Profile Name	<p>An AWS credential profile defined in the credentials file.</p> <p>Required if you use the AWS credential profile authentication type.</p> <p>A mapping accesses the AWS credentials through the profile name at run time. If you do not provide an AWS credential profile name, the mapping uses the access key ID and secret access key that you specify when you create the connection.</p>
ARN of IAM Role	<p>The Amazon Resource Name specifying the role of an IAM user.</p> <p>Required if you use the cross-account IAM role authentication type.</p> <p>Not applicable for a streaming ingestion task.</p>
External ID	<p>The external ID for an IAM role is an additional restriction that you can use in an IAM role trust policy to designate who can assume the IAM role.</p> <p>Required if you use the cross-account IAM role authentication type and if the external ID is defined by the AWS account.</p> <p>Not applicable for a streaming ingestion task.</p>

## Amazon Kinesis Streams connection properties

When you set up an Amazon Kinesis Streams connection, you must configure the connection properties.

The following table describes the Amazon Kinesis Streams connection properties:

Property	Description
Connection Name	<p>Name of the connection.</p> <p>The name is not case sensitive and must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters:</p> <p>~ ` ! \$ % ^ &amp; * ( ) - + = { [ ]   \ : ; " ' &lt; , &gt; . ? /</p>
Description	<p>Optional. Description that you can use to identify the connection.</p> <p>The description cannot exceed 4,000 characters.</p>
Type	<p>The Amazon Kinesis connection type.</p> <p>If you do not see the Amazon Kinesis connection type, go to the <b>Add-On Connectors</b> page to install the connector.</p>
Runtime Environment	<p>Name of the runtime environment where you want to run the tasks.</p>
Service	<p>The type of Kinesis Service that you want to use. Select <b>Kinesis Streams</b>.</p>
AWS Access Key ID	<p>The access key ID of the Amazon AWS user account.</p>

Property	Description
AWS Secret Access Key	The secret access key for your Amazon AWS user account.
Region	<p>Region where the endpoint for your service is available. You can select one of the following values:</p> <ul style="list-style-type: none"> <li>- us-east-2. Indicates the US East (Ohio) region.</li> <li>- us-east-1. Indicates the US East (N. Virginia) region.</li> <li>- us-west-1. Indicates the US West (N. California) region.</li> <li>- us-west-2. Indicates the US West (Oregon) region.</li> <li>- ap-northeast-1. Indicates the Asia Pacific (Tokyo) region.</li> <li>- ap-northeast-2. Indicates the Asia Pacific (Seoul) region.</li> <li>- ap-northeast-3. Indicates the Asia Pacific (Osaka-Local) region.</li> <li>- ap-south-1. Indicates the Asia Pacific (Mumbai) region.</li> <li>- ap-southeast-1. Indicates the Asia Pacific (Singapore) region.</li> <li>- ap-southeast-2. Indicates the Asia Pacific (Sydney) region.</li> <li>- ca-central-1. Indicates the Canada (Central) region.</li> <li>- cn-north-1. Indicates the China (Beijing) region.</li> <li>- cn-northwest-1. Indicates the China (Ningxia) region.</li> <li>- eu-central-1. Indicates the EU (Frankfurt) region.</li> <li>- eu-west-1. Indicates the EU (Ireland) region.</li> <li>- eu-west-2. Indicates the EU (London) region.</li> <li>- eu-west-3. Indicates the EU (Paris) region.</li> <li>- sa-east-1. Indicates the South America (São Paulo) region.</li> <li>- us-gov-west-1. Indicates AWS GovCloud (US-West) region.</li> <li>- us-gov-east-1. Indicates AWS GovCloud (US-East) region.</li> </ul> <p>A streaming ingestion task does not support ap-northeast-3 region.</p>
Connection TimeOut (ms)	<p>Optional. Number of milliseconds that the Mass Ingestion service waits to establish a connection to the Kinesis Streams after which it times out.</p> <p>Default is 10,000 milliseconds.</p>
Authentication Type	<p>The type of authentication.</p> <p>Select one of the following values:</p> <ul style="list-style-type: none"> <li>- AWS Credential Profile.</li> <li>- Cross-account IAM Role</li> </ul> <p>Default is AWS Credential Profile.</p> <p>Cross-account IAM Role is not applicable for a streaming ingestion task.</p>
AWS Credential Profile Name	<p>An AWS credential profile defined in the credentials file.</p> <p>Required if you use the AWS credential profile authentication type.</p> <p>A mapping accesses the AWS credentials through the profile name at run time. If you do not provide an AWS credential profile name, the mapping uses the access key ID and secret access key that you specify when you create the connection.</p>
ARN of IAM Role	<p>The Amazon Resource Name specifying the role of an IAM user.</p> <p>Required if you use the cross-account IAM role authentication type.</p> <p>Not applicable for a streaming ingestion task.</p>
External ID	<p>The external ID for an IAM role is an additional restriction that you can use in an IAM role trust policy to designate who can assume the IAM role.</p> <p>Required if you use the cross-account IAM role authentication type and if the external ID is defined by the AWS account.</p> <p>Not applicable for a streaming ingestion task.</p>

## AWS Credential Profile

You can define AWS credential profiles in the credentials file. Each credential profile consists of secret access key and access key ID.

Users can use the AWS credential profile names to use different AWS credentials at run time than the AWS credentials that they specify when they create an Amazon Kinesis connection with an Amazon Kinesis Streams as a source and target and Amazon Kinesis Firehose as a target.

Create AWS credentials for the users, such as access key ID and secret access key. Users can select an authentication type while creating an Amazon Kinesis connection, such as AWS credential profile. The default authentication type is AWS credential profile.

Generate an Access Key ID and Secret Access Key for the users in AWS.

## Amazon S3 V2 connection properties

When you set up an Amazon S3 V2 connection, you must configure the connection properties.

The following table describes the Amazon S3 V2 connection properties:

Property	Description
Connection Name	Name of the connection. The name is not case sensitive and must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ & * ( ) - + = { [ ] }   \ : ; " ' < , > . ? /
Description	Optional. Description of the connection. The description cannot exceed 4,000 characters.
Type	The Amazon S3 V2 connection type.
Runtime Environment	Name of the runtime environment where you want to run the tasks. You cannot run a database ingestion task on a Hosted Agent or serverless runtime environment.
Access Key	Access key to access the Amazon S3 bucket. Provide the access key value based on the following authentication methods: <ul style="list-style-type: none"><li>- Basic authentication. Provide the actual access key value.</li><li>- IAM authentication. Do not provide the access key value..</li><li>- Temporary security credentials via assume role. Provide the secret access key of an IAM user with no permissions to access Amazon S3 bucket.</li><li>- Assume role for EC2. Do not provide the access key value.</li><li>- Credential profile file authentication. Do not provide the access key value.</li><li>- Federated user single sign-on. Do not provide the secret access key value.</li></ul>



Property	Description
Secret Key	<p>Secret access key to access the Amazon S3 bucket.</p> <p>The secret key is associated with the access key and uniquely identifies the account. Provide the secret access key value based on the following authentication methods:</p> <ul style="list-style-type: none"> <li>- Basic authentication. Provide the actual access secret value.</li> <li>- IAM authentication. Do not provide the access secret value.</li> <li>- Temporary security credentials via assume role. Provide access secret of an IAM user with no permissions to access Amazon S3 bucket.</li> <li>- Assume role for EC2. Do not provide the access key value.</li> <li>- Credential profile file authentication. Do not provide the access secret value.</li> <li>- Federated user single sign-on. Do not provide the access secret value.</li> </ul>
IAM Role ARN	<p>The Amazon Resource Name (ARN) of the AWS Identity and Access Management (IAM) role assumed by the user to use the dynamically generated temporary security credentials.</p> <p>Enter the value of this property if you want to use the temporary security credentials to access the AWS resources.</p> <p>This property is not applicable to an application ingestion task and database ingestion task.</p> <p><b>Note:</b> Even if you remove the IAM role that enables the agent to access the Amazon S3 bucket, and create a connection, the test connection is successful.</p> <p>For more information about how to obtain the ARN of the IAM role, see the AWS documentation.</p>
External Id	Optional. Specify the external ID for a more secure access to the Amazon S3 bucket when the Amazon S3 bucket is in a different AWS account.
Use EC2 Role to Assume Role	<p>Optional. Select the check box to enable the EC2 role to assume another IAM role specified in the IAM Role ARN option.</p> <p><b>Note:</b> The EC2 role must have a policy attached with a permission to assume an IAM role from the same or different account.</p> <p>By default, the Use EC2 Role to Assume Role check box is not selected.</p>
Folder Path	<p>Bucket name or complete folder path to the Amazon S3 objects.</p> <p>Do not use a slash at the end of the folder path. For example:</p> <p><code>&lt;bucket name&gt;/&lt;my folder name&gt;.</code></p>
Master Symmetric Key	<p>Optional. Provide a 256-bit AES encryption key in the Base64 format when you enable client-side encryption. You can generate a key using a third-party tool.</p> <p>Not applicable for an application ingestion task, database ingestion task, or streaming ingestion task.</p>

Property	Description
Customer Master Key ID	<p>Optional. Specify the customer master key ID or alias name generated by AWS Key Management Service (AWS KMS) or the Amazon Resource Name (ARN) of your custom key for cross-account access.</p> <p>You must generate the customer master key for the same region where Amazon S3 bucket resides. You can specify the following master keys:</p> <p><b>Customer generated customer master key</b></p> <p>Enables client-side or server-side encryption.</p> <p><b>Default customer master key</b></p> <p>Enables client-side or server-side encryption. Only the administrator user of the account can use the default customer master key ID to enable client-side encryption.</p> <p>Not applicable for an application ingestion task, database ingestion task, or streaming ingestion task.</p>
S3 Account Type	<p>The type of the Amazon S3 account.</p> <p>Select from the following options:</p> <ul style="list-style-type: none"> <li>- Amazon S3 Storage. Enables you to use the Amazon S3 services.</li> <li>- S3 Compatible Storage. Specify the endpoint for a third-party storage provider such as Scalify RING or MinIO.</li> </ul> <p>By default, Amazon S3 storage is selected.</p>
REST Endpoint	<p>The S3 storage endpoint.</p> <p>Specify the S3 storage endpoint in HTTP or HTTPs format when you select the S3 compatible storage option. For example, <code>http://s3.isv.scality.com</code></p>
Region Name	<p>The AWS region of the bucket that you want to access.</p> <p>Select one of the following regions:</p> <ul style="list-style-type: none"> <li>- Asia Pacific (Mumbai)</li> <li>- Asia Pacific (Seoul)</li> <li>- Asia Pacific (Singapore)</li> <li>- Asia Pacific (Sydney)</li> <li>- Asia Pacific (Tokyo)</li> <li>- Asia Pacific (Hong Kong)</li> <li>- AWS GovCloud (US)</li> <li>- AWS GovCloud (US-East)</li> <li>- Canada (Central)</li> <li>- China (Beijing)</li> <li>- China (Ningxia)</li> <li>- EU (Ireland)</li> <li>- EU (Frankfurt)</li> <li>- EU (London)</li> <li>- EU (Paris)</li> <li>- EU (Stockholm)</li> <li>- South America (Sao Paulo)</li> <li>- Middle East (Bahrain)</li> <li>- US East (Ohio)</li> <li>- US East (N. Virginia)</li> <li>- US West (N. California)</li> <li>- US West (Oregon)</li> </ul> <p>Default is US East (N. Virginia).</p>

Property	Description
Federated SSO IdP	<p>SAML 2.0-enabled identity provider for the federated user single sign-on to use with the AWS account. Amazon S3 V2 connector supports only ADFS 3.0 identity provider. Select <code>None</code> if you do not want to use federated user single sign-on.</p> <p><b>Note:</b> Federated user single sign-on is not applicable to application ingestion tasks, database ingestion tasks, and streaming ingestion tasks.</p>
Other Authentication Type	<p>Select one the following authentication types:</p> <ul style="list-style-type: none"> <li>- NONE</li> <li>- Credential Profile File Authentication</li> </ul> <p>Select the Credential Profile File Authentication option to access the Amazon S3 credentials from a credential file that contains the access key and secret key.</p> <p>Specify the credential profile file path and the profile name to establish the connection with Amazon S3.</p> <p>You can use permanent IAM credentials or temporary session tokens when you configure the Credential Profile File Authentication.</p> <p>Default is NONE.</p>
Credential Profile File Path	<p>Optional. Specify the credential profile file path.</p> <p>If you do not specify the credential profile path, the Secure Agent uses the credential profile file present in the following default location in your home directory:</p> <pre>~/aws/credentials</pre> <p><b>Note:</b> Mass Ingestion Databases has not been certified with the <b>Credential Profile File Path</b> and <b>Profile Name</b> connection properties. Mass Ingestion Databases finds AWS credentials by using the default credential provider chain that is implemented by the <code>DefaultAWSCredentialsProviderChain</code> class, which includes the credential profile file.</p>
Profile Name	<p>Optional. Name of the profile in the credential profile file used to get the credentials.</p> <p>If you do not specify the profile name, the credentials from the default profile in the credential profile file are used.</p>

## Federated user single sign-on connection properties

Configure the following properties when you select ADFS 3.0 in **Federated SSO IdP**:

Property	Description
Federated User Name	User name of the federated user to access the AWS account through the identity provider.
Federated User Password	Password for the federated user to access the AWS account through the identity provider.
IdP SSO URL	<p>Single sign-on URL of the identity provider for AWS.</p> <p>Not applicable for a streaming ingestion task.</p>
SAML Identity Provider ARN	ARN of the SAML identity provider that the AWS administrator created to register the identity provider as a trusted provider.
Role ARN	ARN of the IAM role assumed by the federated user.

## Credential Profile File Authentication

You can provide the credentials required to establish the connection with Amazon S3 through the credential profile file that contains an access key and secret key. The credential profile file contains an access key, a secret key, and a session token when you use temporary security credentials.

You can use permanent IAM credentials or temporary security credentials with a session token when you use credential profile file authentication.

If you do not specify the credential profile file path, the default credential file path is used. If you do not specify the profile name, the credentials are used from the default profile in the credential file.

Consider the following rules for a credential profile file:

- The credential file must be on the same machine where you installed the Secure Agent.
- The credential profile file name must end with `.credentials`.
- If you do not specify the credential profile path, the Secure Agent uses the credential profile file present in the following default location in your home directory:

`~/.aws/credentials`

**Note:** On Windows, you can refer to your home directory by using the environment variable `%UserProfile` `%`. On Unix-like systems, you can use the environment variable `$HOME`.

A sample credential profile file:

```
[default]

aws_access_key_id = 12333333

aws_secret_access_key = abcabcabc


[test-profile]

aws_access_key_id = 12333333

aws_secret_access_key = abcabcabc

aws_session_token = jahaheieomdrftflmlioerp
```

The `aws_access_key_id` and `aws_secret_access_key` specify the AWS access key and secret key used as part of credentials to authenticate the user.

The `aws_session_token` specifies an AWS session token used as part of the credentials to authenticate the user. A session token is required only if you specify temporary security credentials.

## AMQP connection properties

When you set up an AMQP connection, you must configure the connection properties.

The following table describes the AMQP connection properties:

Property	Description
Connection Name	Name of the connection. The name is not case sensitive and must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ & * ( ) - + = { [ ] }   \ : ; " ' < , > . ? /
Description	Optional. Description that you can use to identify the connection. The description cannot exceed 4,000 characters.
Type	The AMQP connection type. If you do not see the connection type, go to the <b>Add-On Connectors</b> page to install the connector.
Runtime Environment	Name of the runtime environment where you want to run the tasks.
Host Name	Network address of the AMQP broker.
Port	Port number of the AMQP broker to which the underlying TCP connection is made. Default is 5672.
Virtual Host	Virtual host name that identifies the AMQP system. Use the virtual host name for enhanced security.
Username	Username for the AMQP broker.
Password	Password for the AMQP broker.
Use SSL	Enable this option to use SSL for secure transmission. If you enable the SSL authentication, ensure that you provide both keystore and truststore details for using the AMQP connection in a streaming ingestion task.
Keystore File Name	Contains the keys and certificates required for secure communication.
Keystore Password	Password for the keystore filename.
Keystore Type	Type of keystore that you want to use. Keystore type defines the storage and data format of the keystore information and the algorithms used to protect private keys in the keystore. Use one of the following types: - JKS. Stores private keys and certificates. - PKCS12. Stores private keys, secret keys, and certificates.
Truststore File Name	Name of the truststore file.
Truststore Password	Password for the truststore file.

Property	Description
Truststore Type	Type of truststore that you want to use. Use one of the following types: <ul style="list-style-type: none"> <li>- JKS</li> <li>- PKCS 12</li> </ul>
TLS Protocol	Transport protocols that you want to use. Use one of the following types: <ul style="list-style-type: none"> <li>- SSL</li> <li>- SSLv2Hello</li> <li>- SSLv3</li> <li>- TLS</li> <li>- TLSv1</li> <li>- TLSv1.1</li> <li>- TLSv1.2</li> </ul>
Client Authentication	Client authentication policy when connecting to the secured AMQP broker. Use one of the following property values when you define and enable an SSL context. <ul style="list-style-type: none"> <li>- WANT</li> <li>- REQUIRED</li> <li>- NONE</li> </ul>

## Db2 for i Database Ingestion connection properties

When you define a Db2 for i Database Ingestion connection, you must configure connection properties. You can use this connection type in database ingestion tasks, which you configure in the Mass Ingestion service.

The following table describes the connection properties:

Property	Description
Connection Name	A name for the connection. This name must be unique within the organization. Connection names can contain alphanumeric characters, spaces, and the following special characters: _ . + - Spaces at the beginning or end of the name are trimmed and are not saved as part of the name. Maximum length is 100 characters. Connection names are not case sensitive.
Description	An optional description for the connection. Maximum length is 255 characters.
Type	The type of connection. Ensure that the type is Db2 for i Database Ingestion.
Runtime Environment	The name of the runtime environment where you want to run database ingestion tasks. You define runtime environments in Administrator.
User Name	The user name to use for connecting to the Db2 for i instance.
Password	The password to use for connecting to the Db2 for i instance.
Host	The name of the machine that hosts the database server.
Port	The network port number used to connect to the database server.

Property	Description
Location Name	The name of the Db2 for i location that you want to access. Your system administrator can determine the name of the Db2 location by using the WRKRDBDIRE command. In the output, find the name of the database that is listed as *LOCAL and then use that value as the value of this property.
Code Page for Bit Data	The code page that Mass Ingestion Databases uses to read character data that is stored as bit data. This value must be a canonical name for the java.io API and java.lang API. For more information, see the supported encodings in the Oracle Java documentation. Specify this property if you have FOR BIT DATA source columns.
Advanced Connection Properties	Advanced properties for the Progress DataDirect JDBC DB2 driver, which is used to connect to the Db2 for i source. If you specify more than one <i>property=value</i> entry, separate them with a semicolon (;).  The driver properties that you can enter in this field are described in the Progress DataDirect documentation at <a href="https://docs.progress.com/bundle/datadirect-connect-jdbc-51/page/Connection-Properties_10.html">https://docs.progress.com/bundle/datadirect-connect-jdbc-51/page/Connection-Properties_10.html</a> . For example, you can set the ConnectionRetryCount property to control the number of times the driver retries attempts to connect to the primary database server.

## Db2 for LUW Database Ingestion connection properties

When you define a Db2 for LUW Database Ingestion connection, you must configure connection properties. You can use this connection type in database ingestion tasks, which you configure in the Mass Ingestion service.

The following table describes the connection properties:

Property	Description
Connection Name	A name for the connection. This name must be unique within the organization. Connection names can contain alphanumeric characters, spaces, and the following special characters: _ . + - Spaces at the beginning or end of the name are trimmed and are not saved as part of the name. Maximum length is 100 characters. Connection names are not case sensitive.
Description	An optional description for the connection. Maximum length is 255 characters.
Type	The type of connection. Ensure that the type is Db2 for LUW Database Ingestion.
Runtime Environment	The name of the runtime environment where you want to run database ingestion tasks. You define runtime environments in Administrator.
User Name	The user name to use for connecting to the Db2 for LUW instance.
Password	The password to use for connecting to the Db2 for LUW instance.
Host	The name of the machine that hosts the database server.
Port	The network port number used to connect to the database server.

Property	Description
Database Name	The name of the Db2 for LUW database that you want to access.
Advanced Connection Properties	<p>Advanced properties for the Progress DataDirect JDBC DB2 driver, which is used to connect to the Db2 for LUW source. If you specify more than one <i>property=value</i> entry, separate them with a semicolon (;).</p> <p>The driver properties that you can enter in this field are described in the Progress DataDirect documentation at <a href="https://docs.progress.com/bundle/datadirect-connect-jdbc-51/page/Connection-Properties_10.html">https://docs.progress.com/bundle/datadirect-connect-jdbc-51/page/Connection-Properties_10.html</a>. For example, you can set the EncryptionMethod property to control whether data is encrypted and decrypted when transmitted over the network between the driver and database server.</p>

## Db2 for zOS Database Ingestion connection properties

When you define a Db2 for zOS Database Ingestion connection, you must configure connection properties. You can use this connection type in database ingestion tasks, which you configure in the Mass Ingestion service.

The following table describes the connection properties:

Property	Description
Connection Name	A name for the connection. This name must be unique within the organization. Connection names can contain alphanumeric characters, spaces, and the following special characters: _ . + - Spaces at the beginning or end of the name are trimmed and are not saved as part of the name. Maximum length is 100 characters. Connection names are not case sensitive.
Description	An optional description for the connection. Maximum length is 255 characters.
Type	The type of connection. Ensure that the type is Db2 for zOS Database Ingestion.
Runtime Environment	The name of the runtime environment where you want to run database ingestion tasks. You define runtime environments in Administrator.
User Name	The user name to use for connecting to the Db2 for zOS instance.
Password	The password to use for connecting to the Db2 for zOS instance.
Host	The name of the machine that hosts the database server.
Port	The network port number used to connect to the database server.
Location Name	The name of the Db2 for zOS location that you want to access. For DB2 for z/OS, your system administrator can determine the name of your DB2 location using the command DISPLAY DDF.
Code Page for Bit Data	The code page that Mass Ingestion Databases uses to read character data that is stored as bit data. This value must be a canonical name for the java.io API and java.lang API. For more information, see the supported encodings in the Oracle Java documentation. Specify this property if you have FOR BIT DATA source columns.
CDC Stored Procedure Schema	For incremental change data capture processing, the name of the schema for the z/OS stored procedure that is required to collect change data from the Db2 log. This value is specified in the #STPINST data set that you customized when setting up the stored procedure on z/OS. No default value is provided.



Property	Description
CDC Stored Procedure Name	For incremental change data capture processing, the name of the z/OS stored procedure that is required to collect change data from the Db2 log. This value is specified in the #STPINST data set that you customized when setting up the stored procedure on z/OS. The default value is INFALOG.
Advanced Connection Properties	<p>Advanced properties for the Progress DataDirect JDBC DB2 driver, which is used to connect to the Db2 for z/OS source. If you specify more than one <i>property=value</i> entry, separate them with a semicolon (;).</p> <p>The driver properties that you can enter in this field are described in the Progress DataDirect documentation at <a href="https://docs.progress.com/bundle/datadirect-connect-jdbc-51/page/Connection-Properties_10.html">https://docs.progress.com/bundle/datadirect-connect-jdbc-51/page/Connection-Properties_10.html</a>. For example, you can set the ConnectionRetryCount property to control the number of times the driver retries attempts to connect to the primary database server.</p>

## Databricks Delta connection properties

When you create a Databricks Delta connection, you must configure the connection properties.

The following table describes the Databricks Delta connection properties:

Property	Description
Connection Name	<p>Required. The name of the connection. The name is not case sensitive and must be unique within the domain.</p> <p>You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ &amp; * ( ) - + = { [ ]   \ : ; " ' &lt; , &gt; . ? /</p>
Description	<p>Description of the connection.</p> <p>The description cannot exceed 4,000 characters.</p>
Type	Required. Select Databricks Delta.
Runtime Environment	Required. Name of the runtime environment where you want to run the tasks.
Databricks Host	<p>Required. The host name of the endpoint the Databricks account belongs to.</p> <p>Use the following syntax:</p> <pre>jdbc:spark://&lt;Databricks Host&gt;:443/default;transportMode=http;ssl=1;httpPath=sq l/protocolv1/o/&lt;Org Id&gt;/&lt;Cluster ID&gt;;AuthMech=3;UID=token;PWD=&lt;personal-access-token&gt;</pre> <p><b>Note:</b> You can get the URL from the Databricks Delta analytics cluster or all purpose cluster -&gt; <b>Advanced Options</b> -&gt; <b>JDBC / ODBC</b>.</p> <p>The value of PWD in Databricks Host, Org Id, and Cluster ID is always &lt;personal-access-token&gt;.</p>

Property	Description
Org Id	<p>Required. The unique organization ID for the workspace in Databricks.</p> <p>Use the following syntax:</p> <pre>jdbc:spark://&lt;Databricks Host&gt;:443/default;transportMode=http;ssl=1;httpPath=sql/protocolv1/o/&lt;Org Id&gt;/&lt;Cluster ID&gt;;AuthMech=3;UID=token;PWD=&lt;personal-access-token&gt;</pre>
Cluster ID	<p>Required. The ID of the Databricks analytics cluster. You can obtain the cluster ID from the JDBC URL.</p> <p>Use the following syntax:</p> <pre>jdbc:spark://&lt;Databricks Host&gt;:443/default;transportMode=http;ssl=1;httpPath=sql/protocolv1/o/&lt;Org Id&gt;/&lt;Cluster ID&gt;;AuthMech=3;UID=token;PWD=&lt;personal-access-token&gt;</pre>
Databricks Token	<p>Required. Personal access token to access Databricks. You must have permissions to attach to the cluster identified in the <b>Cluster ID</b> property.</p>
SQL Endpoint JDBC URL	<p>Databricks SQL endpoint JDBC connection URL.</p> <p>Use the following syntax:</p> <pre>jdbc:spark://&lt;Databricks Host&gt;:443/default;transportMode=http;ssl=1;AuthMech=3;httpPath=/sql/1.0/endpoints/&lt;SQL endpoint cluster ID&gt;;</pre> <p><b>Note:</b> The Databricks Host, Org ID, and Cluster ID properties are not considered if you configure the SQL Endpoint JDBC URL property.</p> <p>For more information on Databricks Delta SQL endpoint, contact Informatica Global Customer Support.</p>
Database	The database in Databricks Delta that you want to connect to.
JDBC Driver Class Name	Required. The name of the JDBC driver class.
Cluster Environment	<p>The cloud provider where the Databricks cluster is deployed.</p> <p>You can select from the following options:</p> <ul style="list-style-type: none"> <li>- AWS</li> <li>- Azure</li> </ul> <p>Default is AWS.</p> <p>The connection attributes differ depending on the cluster environment you select. For more information, see the AWS and Azure cluster properties sections.</p>
Min Workers	The minimum number of worker nodes to be used for the Spark job.

Property	Description
Max Workers	The maximum number of worker nodes to be used for the Spark job. If you do not want autoscale, set Max Workers = Min Workers or do not set Max Workers.
DB Runtime Version	The Databricks runtime version. Select 7.3 LTS from the list.
Worker Node Type	<i>Required.</i> The instance type of the machine used for the Spark worker node.
Driver Node Type	The instance type of the machine used for the Spark driver node. If not provided, the value as in worker node type is used.
Instance Pool ID	The instance pool used for the Spark cluster.
Enable Elastic Disk	Enable this option for the cluster to dynamically acquire additional disk space when the Spark workers are running low on disk space.
Spark Configuration	The Spark configuration to be used in the Databricks cluster. The configuration must be in the following format: <code>"key1"="value1";"key2"="value2";...</code> For example: <code>"spark.executor.userClassPathFirst"="False"</code>
Spark Environment Variables	The environment variables that you need to export before launching the Spark driver and workers. The variables must be in the following format: <code>"key1"="value1";"key2"="value2";...</code> For example: <code>"MY_ENVIRONMENT_VARIABLE"="true"</code>

## Flat file connection properties

The following table describes the flat file connection properties:

Connection Property	Description
Runtime Environment	Runtime environment that contains the Secure Agent to use to access the flat files. <b>Note:</b> Do not select a runtime environment with Secure Agents that run on NTT. A flat file connection cannot use a Secure Agent that runs on NTT.
Directory	<p>Directory where the flat file is stored. Must be accessible by all Secure Agents in the selected runtime environment.</p> <p>Enter the full directory or click <b>Browse</b> to locate and select the directory.</p> <p>When you use the connection, you can select a file that's contained in the directory or in any of its subdirectories.</p> <p>Maximum length is 100 characters. Directory names can contain alphanumeric characters, spaces, and the following special characters:</p> <p>/ \ : _ ~</p> <p>The directory is the service URL for this connection type.</p> <p><b>Note:</b> On Windows, the <b>Browse for Directory</b> dialog box does not display mapped drives. You can browse My Network Places to locate the directory or enter the directory name in the following format: \\&lt;server_name&gt;\&lt;directory_path&gt;. If network directories do not display, you can configure a login for the Secure Agent service.</p> <p>Do not include the name of the flat file. You specify the file name when you create the task.</p>
Browse button	Use to locate and select the directory where flat files are stored.
Date Format	Date format for date fields in the flat file. Default date format is: MM/dd/yyyy HH:mm:ss
Code Page	<p>The code page of the system that hosts the flat file. Select one of the following code pages:</p> <ul style="list-style-type: none"> <li>- MS Windows Latin 1. Select for ISO 8859-1 Western European data.</li> <li>- UTF-8. Select for Unicode data.</li> <li>- Shift-JIS. Select for double-byte character data.</li> <li>- ISO 8859-15 Latin 9 (Western European).</li> <li>- ISO 8859-2 Eastern European.</li> <li>- ISO 8859-3 Southeast European.</li> <li>- ISO 8859-5 Cyrillic.</li> <li>- ISO 8859-9 Latin 5 (Turkish).</li> <li>- IBM EBCDIC International Latin-1.</li> <li>- Japanese EUC (with \ &lt;-&gt; Yen mapping)</li> <li>- IBM EBCDIC Japanese</li> <li>- IBM EBCDIC Japanese CP939</li> <li>- PC Japanese SJIS-78 syntax (IBM-942)</li> <li>- PC Japanese SJIS-90 (IBM-943)</li> <li>- MS Windows Traditional Chinese, superset of Big 5</li> <li>- Taiwan Big-5 (w/o euro update)</li> <li>- Chinese EUC</li> <li>- ISO 8859-8 Hebrew</li> <li>- PC Hebrew (old)</li> <li>- PC Hebrew (w/o euro update)</li> <li>- EBCDIC Hebrew (updated with new sheqel, control characters)</li> <li>- IBM EBCDIC US English IBM037</li> </ul> <p><b>Note:</b> When you use a flat file connection with the Shift-JIS code page and a UTF data object, be sure to install fonts that fully support Unicode.</p>

## Google Analytics Mass Ingestion connection properties

When you set up a Google Analytics Mass Ingestion connection, you must configure the connection properties.

The following table describes the connection properties for a Google Analytics Mass Ingestion connection:

Connection property	Description
Runtime Environment	Name of the runtime environment where you want to run the ingestion tasks. You must specify a Secure Agent as the runtime environment. <b>Note:</b> You cannot run application ingestion tasks on a Hosted Agent or serverless runtime environment.
Service Account ID	Specifies the client_email value present in the JSON file that you download after you create a service account.
Service Account Key	Specifies the private_key value present in the JSON file that you download after you create a service account.

## Google BigQuery V2 connection properties

When you create a Google BigQuery V2 connection, configure the connection properties.

The following table describes the Google BigQuery V2 connection properties:

Property	Description
Connection Name	The name of the connection. The name is not case sensitive and must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ & * ( ) - + = { } }   \ : ; " ' < , > . ? /
Description	Optional. The description of the connection. The description cannot exceed 4,000 characters.
Type	The <b>Google Big Query V2</b> connection type.
Runtime Environment	Name of the runtime environment where you want to run the tasks. You cannot run a database ingestion task on a Hosted Agent or serverless runtime environment.
Service Account ID	Specifies the client_email value present in the JSON file that you download after you create a service account.
Service Account Key	Specifies the private_key value present in the JSON file that you download after you create a service account.
Project ID	Specifies the project_id value present in the JSON file that you download after you create a service account. If you have created multiple projects with the same service account, enter the ID of the project that contains the dataset that you want to connect to.

Property	Description
Storage Path	<p>This property applies when you read or write large volumes of data. Required if you read data in staging mode or write data in bulk mode.</p> <p>Path in Google Cloud Storage where the agent creates a local stage file to store the data temporarily.</p> <p>You can either enter the bucket name or the bucket name and folder name.</p> <p>For example, enter <code>gs://&lt;bucket_name&gt;</code> or <code>gs://&lt;bucket_name&gt;/&lt;folder_name&gt;</code></p>
Connection mode	<p>The mode that you want to use to read data from or write data to Google BigQuery.</p> <p>Select one of the following connection modes:</p> <ul style="list-style-type: none"> <li>- Simple. Flattens each field within the Record data type field as a separate field in the mapping.</li> <li>- Hybrid<sup>1</sup>. Displays all the top-level fields in the Google BigQuery table including Record data type fields. Google BigQuery V2 Connector displays the top-level Record data type field as a single field of the String data type in the mapping.</li> <li>- Complex<sup>1</sup>. Displays all the columns in the Google BigQuery table as a single field of the String data type in the mapping.</li> </ul> <p>Default is Simple.</p>
Schema Definition File Path <sup>1</sup>	<p>Specifies a directory on the Secure Agent machine where the Secure Agent must create a JSON file with the sample schema of the Google BigQuery table. The JSON file name is the same as the Google BigQuery table name.</p> <p>Alternatively, you can specify a storage path in Google Cloud Storage where the Secure Agent must create a JSON file with the sample schema of the Google BigQuery table. You can download the JSON file from the specified storage path in Google Cloud Storage to a local machine.</p> <p>The schema definition file is required if you configure complex connection mode in the following scenarios:</p> <ul style="list-style-type: none"> <li>- You add a Hierarchy Builder transformation in a mapping to read data from relational sources and write data to a Google BigQuery target.</li> <li>- You add a Hierarchy Parser transformation in a mapping to read data from a Google BigQuery source and write data to relational targets.</li> </ul> <p><b>Note:</b> When you use a serverless runtime environment, you must specify a storage path in Google Cloud Storage.</p>
Use Legacy SQL For Custom Query <sup>1</sup>	<p>Select this option to use a legacy SQL to define a custom query. If you clear this option, you must use a standard SQL to define a custom query.</p> <p><b>Note:</b> Not applicable when you configure the Google BigQuery V2 connection in hybrid or complex mode</p>
Dataset Name for Custom Query <sup>1</sup>	<p>When you define a custom query, you must specify a Google BigQuery dataset.</p>
Region Id	<p>Specify the region name where the Google BigQuery dataset that you want to access resides.</p> <p><b>Note:</b> You must ensure that you specify a bucket name or the bucket name and folder name in the <b>Storage Path</b> property that resides in the specified region.</p> <p>For more information about the regions supported by Google BigQuery, see the following Google BigQuery documentation: <a href="https://cloud.google.com/bigquery/docs/locations">https://cloud.google.com/bigquery/docs/locations</a></p>

Property	Description
Optional Properties <sup>1</sup>	<p>Specifies whether you can configure certain source and target functionalities through custom properties.</p> <p>You can select one of the following options:</p> <ul style="list-style-type: none"> <li>- None. If you do not want to configure any custom properties, select None.</li> <li>- Required. If you want to specify custom properties to configure the source and target functionalities.</li> </ul> <p>Default is None.</p>
Provide Optional Properties <sup>1</sup>	<p>Comma-separated key-value pairs of custom properties in the Google BigQuery V2 connection to configure certain source and target functionalities.</p> <p>Appears only when you select <b>Required</b> in the Optional Properties.</p> <p>For more information about the list of custom properties that you can specify, see the Informatica Knowledge Base article: <a href="https://kb.informatica.com/faq/7/Pages/26/632722.aspx">https://kb.informatica.com/faq/7/Pages/26/632722.aspx</a></p>

**Note:** Ensure that you specify valid credentials in the connection properties. The test connection is successful even if you specify incorrect credentials in the connection properties.

## Google Cloud Storage V2 connection properties

When you create a Google Cloud Storage V2 connection, configure the connection properties.

The following table describes the Google Cloud Storage connection properties:

Property	Description
Connection Name	<p>The name of the connection.</p> <p>The name is not case sensitive and must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ &amp; * ( ) - + = { [ ]   \ : ; " ' &lt; , &gt; . ? /</p>
Description	<p>Optional. Description that you use to identity the connection.</p> <p>The description cannot exceed 4,000 characters.</p>
Type	The Google Cloud Storage V2 connection type.
Runtime Environment	<p>Name of the runtime environment where you want to run the tasks.</p> <p>You cannot run a database ingestion task or streaming ingestion task on a Hosted Agent or serverless runtime environment.</p>
Service Account ID	The <code>client_email</code> value in the JSON file that you download after you create a service account.
Service Account Key	The <code>private_key</code> value in the JSON file that you download after you create a service account.
Project ID	<p>The <code>project_id</code> value in the JSON file that you download after you create a service account.</p> <p>If you have created multiple projects with the same service account, enter the ID of the project that contains the bucket that you want to connect to.</p>

Property	Description
Private Key ID	The <code>private_key_id</code> value in the JSON file that you download after you create a service account. This property applies only to a database ingestion or streaming ingestion task.
Client ID	The <code>client_id</code> value in the JSON file that you download after you create a service account. This property applies only to a database ingestion or streaming ingestion task.
Bucket Name	The Google Cloud Storage bucket name that you want to connect to. When you select a source object, the Package Explorer lists files and folder available in the specified Google Cloud Storage bucket. If you do not specify a bucket name, you can select a bucket from the Package Explorer to select a source.

## Configuring the proxy settings on Windows

If your organization uses an outgoing proxy server to connect to the Internet, the Secure Agent connects to Informatica Intelligent Cloud Services using the proxy server.

To configure the proxy server settings for the Secure Agent on a Windows machine, you must configure the proxy server settings through the Secure Agent Manager and the JVM options of the Secure Agent.

**Restriction:** These steps do not work for Mass Ingestion Databases.

Contact your network administrator for the proxy settings.

1. Click **Start > All Programs > Informatica Cloud Secure Agent > Informatica Cloud Secure Agent** to launch the Secure Agent Manager.

The **Secure Agent Manager** displays the **Secure Agent** status.

2. Click **Proxy** on the Secure Agent Manager page.
3. Click **Use a Proxy Server** to enter the proxy server settings.
4. Configure the following proxy server details:

Field	Description
Proxy Host	Host name of the outgoing proxy server that the Secure Agent uses.
Proxy Port	Port number of the outgoing proxy server.
User Name	User name to connect to the outgoing proxy server.
Password	Password to connect to the outgoing proxy server.

5. Click **OK**.
6. Log in to Informatica Intelligent Cloud Services.
7. Open Administrator and select **Runtime Environments**.
8. Select the Secure Agent for which you want to configure a proxy server.
9. On the upper-right corner of the page, click **Edit**.
10. In the **System Configuration Details** section, select the **Type** as **Agent** for the CMI Streaming Agent Service.



11. To use a proxy server, add the following parameters in any **JVMOption** field and specify appropriate values for each parameter:

Parameter	Description
-Dproxy.host=	Host name of the outgoing HTTPS proxy server.
-Dproxy.port=	Port number of the outgoing HTTPS proxy server.
-Dproxy.user=	User name for the HTTPS proxy server.
-Dproxy.password=	Password for the HTTPS proxy server.

**Note:** You must specify the parameter and the value for the parameter enclosed in single quotation marks.

For example,

```
JVMOption1='-Dproxy.host=INPQ8583WI29'
```

```
JVMOption2='-Dproxy.port=8081'
```

```
JVMOption3='-Dproxy.user=adminuser'
```

```
JVMOption4='-Dproxy.password=password'
```

**Note:** You can configure only five **JVMOption** fields in the **System Configuration Details** section. To configure the remaining parameters, you must add the **JVMOption** fields in the **Custom Configuration Details** section. In the **Custom Configuration Details** section, select the **Type** as **Agent** for the CMI Streaming Agent Service, add the **JVMOption** fields, and specify the remaining parameters and appropriate values for each parameter.

12. Click **Save**.

The Secure Agent restarts to apply the settings.

**Note:** The session log does not record the proxy server details even if you have configured a proxy server.

## Configuring the proxy settings on Linux

If your organization uses an outgoing proxy server to connect to the Internet, the Secure Agent connects to Informatica Intelligent Cloud Services through the proxy server.

You can update the proxy server settings defined for the Secure Agent from the command line. To configure the proxy server settings for the Secure Agent on a Linux machine, you must update the `proxy.ini` file and configure the JVM options of the Secure Agent.

**Restriction:** These steps do not work for Mass Ingestion Databases.

Contact your network administrator for the proxy settings.

1. Navigate to the following directory:

```
<Secure Agent installation directory>/apps/agentcore/conf
```

2. To update the `proxy.ini` file, add the following parameters and specify appropriate values for each parameter:

```
InfaAgent.ProxyHost=<proxy_server_hostname>
InfaAgent.ProxyPort=<proxy_server_port>
InfaAgent.ProxyUser=<user_name>
```

```
InfaAgent.ProxyPassword=<password>
InfaAgent.ProxyPasswordEncrypted=false
```

For example,

```
InfaAgent.ProxyHost=INW2PF0MT01V
InfaAgent.ProxyPort=808
InfaAgent.ProxyUser=user06
InfaAgent.ProxyPassword=user06
InfaAgent.ProxyPasswordEncrypted=false
```

3. Log in to Informatica Intelligent Cloud Services.
4. Open Administrator and select **Runtime Environments**.
5. Select the Secure Agent for which you want to configure a proxy server.
6. On the upper-right corner of the page, click **Edit**.
7. In the **System Configuration Details** section, select the **Type** as **Agent** for the CMI Streaming Agent Service.
8. To use a proxy server, add the following parameters in any **JVMOption** field and specify appropriate values for each parameter:

Parameter	Description
-Dproxy.host=	Host name of the outgoing HTTPS proxy server.
-Dproxy.port=	Port number of the outgoing HTTPS proxy server.
-Dproxy.user=	User name for the HTTPS proxy server.
-Dproxy.password=	Password for the HTTPS proxy server.

**Note:** You must specify the parameter and the value for the parameter enclosed in single quotation marks.

For example,

```
JVMOption1='-Dproxy.host=INPQ8583WI29'
JVMOption2='-Dproxy.port=8081'
JVMOption3='-Dproxy.user=adminuser'
JVMOption4='-Dproxy.password=password'
```

**Note:** You can configure only five **JVMOption** fields in the **System Configuration Details** section. To configure the remaining parameters, you must add the **JVMOption** fields in the **Custom Configuration Details** section. In the **Custom Configuration Details** section, select the **Type** as **Agent** for the CMI Streaming Agent Service, add the **JVMOption** fields, and specify the remaining parameters and appropriate values for each parameter.

9. Click **Save**.

The Secure Agent restarts to apply the settings.

**Note:** The session log does not record the proxy server details even if you have configured a proxy server.

## Google PubSub - Mass Ingestion Streaming connection properties

When you define a Google PubSub Mass Ingestion Streaming connection, you must configure connection properties. You can use this connection type in streaming ingestion tasks, which you configure in the Mass Ingestion service.

The following table describes the Google PubSub connection properties:

Property	Description
Connection Name	The name of the connection. The name is not case sensitive and must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ & * ( ) - + = { [ ]   \ : ; " ' < , > . ? /
Description	Optional. Description that you use to identify the connection. The description must not exceed 4,000 characters.
Type	The <b>Google PubSub</b> connection type.
Runtime Environment	Name of the runtime environment where you want to run the tasks.
Client Email	The <code>client_email</code> value available in the JSON file that you download after you create a service account.
Client ID	The <code>client_id</code> value available in the JSON file that you download after you create a service account.
Private Key ID	The <code>private_key_id</code> value available in the JSON file that you download after you create a service account.
Private Key	The <code>private_key</code> value available in the JSON file that you download after you create a service account.
Project ID	The <code>project_id</code> value available in the JSON file that you download after you create a service account.

**Note:** The test connection for the Google PubSub connector does not fail even if you enter incorrect values for **Client ID** and **Private Key ID**.

## Hadoop Files V2 connection properties

When you set up a Hadoop Files V2 connection, you must configure the connection properties.

The following table describes the Hadoop Files V2 connection properties:

Connection property	Description
Connection Name	Name of the Hadoop Files V2 connection.
Description	Description of the connection. The description cannot exceed 765 characters.
Type	Type of connection. Select <b>Hadoop Files V2</b> .

Connection property	Description
Runtime Environment	The name of the runtime environment where you want to run the tasks.
User Name	Required to read data from HDFS. Enter a user name that has access to the single-node HDFS location to read data from or write data to.
NameNode URI	<p>The URI to access HDFS.</p> <p>Use the following format to specify the name node URI in Cloudera, Amazon EMR, and Hortonworks distributions:</p> <pre>hdfs://&lt;namenode&gt;:&lt;port&gt;/</pre> <p>Where</p> <ul style="list-style-type: none"> <li>- &lt;namenode&gt; is the host name or IP address of the name node.</li> <li>- &lt;port&gt; is the port that the name node listens for remote procedure calls (RPC).</li> </ul> <p>If the Hadoop cluster is configured for high availability, you must copy the <code>fs.defaultFS</code> value in the <code>core-site.xml</code> file and append <code>/</code> to specify the name node URI.</p> <p>For example, the following snippet shows the <code>fs.defaultFS</code> value in a sample <code>core-site.xml</code> file:</p> <pre>&lt;property&gt;   &lt;name&gt;fs.defaultFS&lt;/name&gt;   &lt;value&gt;hdfs://nameservice1&lt;/value&gt;   &lt;source&gt;core-site.xml&lt;/source&gt; &lt;/property&gt;</pre> <p>In the above snippet, the <code>fs.defaultFS</code> value is</p> <pre>hdfs://nameservice1</pre> <p>and the corresponding name node URI is</p> <pre>hdfs://nameservice1/</pre> <p><b>Note:</b> Specify either the name node URI or the local path. Do not specify the name node URI if you want to read data from or write data to a local file system path.</p>
Local Path	<p>A local file system path to read and write data. Read the following conditions to specify the local path:</p> <ul style="list-style-type: none"> <li>- You must enter <b>NA</b> in local path if you specify the name node URI. If the local path does not contain <b>NA</b>, the name node URI does not work.</li> <li>- If you specify the name node URI and local path, the local path takes the preference. The connection uses the local path to run all tasks.</li> <li>- If you leave the local path blank, the agent configures the root directory (<code>/</code>) in the connection. The connection uses the local path to run all tasks.</li> <li>- If the file or directory is in the local system, enter the fully qualified path of the file or directory.</li> </ul> <p>For example, <code>/user/testdir</code> specifies the location of a directory in the local system.</p> <p>Default value for Local Path is NA.</p>
Configuration Files Path	<p>The directory that contains the Hadoop configuration files.</p> <p><b>Note:</b> Copy the <code>core-site.xml</code>, <code>hdfs-site.xml</code>, and <code>hive-site.xml</code> from the Hadoop cluster and add them to a folder in Linux Box.</p>
Keytab File	The file that contains encrypted keys and Kerberos principals to authenticate the machine.

Connection property	Description
Principal Name	Users assigned to the superuser privilege can perform all the tasks that a user with the administrator privilege can perform.
Impersonation Username	You can enable different users to run mappings in a Hadoop cluster that uses Kerberos authentication or connect to sources and targets that use Kerberos authentication. To enable different users to run mappings or connect to big data sources and targets, you must configure user impersonation.

**Note:** When you read from or write to remote files, the **Name Node URI** and **Configuration Files Path** fields are mandatory. When you read from or write to local files only **Local Path** field is required.

## JDBC V2 connection properties

When you set up a JDBC V2 connection, you must configure the connection properties.

The following table describes JDBC V2 connection properties:

Connection property	Description
Runtime Environment	The name of the runtime environment where you want to run the tasks. You can specify a Secure Agent for a mapping. Specify a Secure Agent or a serverless runtime environment for an elastic mapping.
User Name	The user name to connect to the database.
Password	The password for the database user name.
Schema Name	Optional. The schema name. If you do not specify the schema name, all the schemas available in the database are listed.
JDBC Driver Class Name	Name of the JDBC driver class. To connect to Aurora PostgreSQL, specify the following driver class name: <code>org.postgresql.Driver</code> For more information about which driver class to use with specific databases, see the corresponding third-party vendor documentation.
Connection String	Connection string to connect to the database. Use the following format to specify the connection string: <code>jdbc:&lt;subprotocol&gt;:&lt;subname&gt;</code> For example, the connection string for the Aurora PostgreSQL database type is <code>jdbc:postgresql://&lt;host&gt;:&lt;port&gt;[/dbname]</code> For more information about the connection string to use with specific drivers, see the corresponding third-party vendor documentation.

Connection property	Description
Database Type	<p>The database type to which you want to connect.</p> <p>You can select from the following database types:</p> <ul style="list-style-type: none"> <li>- PostgreSQL. Connect to the Aurora PostgreSQL database hosted in the Amazon Web Services or the Microsoft Azure environment.</li> <li>- Azure SQL Database. Connect to Azure SQL Database hosted in the Microsoft Azure environment.</li> <li>- Others. Connect to any database that supports the Type 4 JDBC driver.</li> </ul>
Support Mixed-Case Identifiers	<p>Indicates whether the database supports case-sensitive identifiers.</p> <p>When enabled, the Secure Agent encloses all identifiers within the character selected for the SQL Identifier Character property.</p>
SQL Identifier Character	<p>Type of character that the database uses to enclose delimited identifiers in SQL queries. The available characters depend on the database type.</p> <p>Select <b>None</b> if the database uses regular identifiers. When the Secure Agent generates SQL queries, it does not place delimited characters around any identifiers.</p> <p>Select a character if the database uses delimited identifiers. When the Secure Agent generates SQL queries, it encloses delimited identifiers within this character.</p>

## JMS connection properties

When you set up a JMS connection, you must configure the connection properties.

The following table describes the connection properties for the JMS connection:

Property	Description
Connection Name	<p>Name of the connection.</p> <p>The name is not case sensitive and must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters:</p> <p>~ ` ! \$ % ^ &amp; * ( ) - + = { [ ]   \ : ; " ' &lt; , &gt; . ? /</p>
Description	<p>Optional. Description that you can use to identify the connection.</p> <p>The description cannot exceed 4,000 characters.</p>
Type	<p>The JMS connection type.</p> <p>If you do not see the connection type, go to the <b>Add-On Connectors</b> page to install the connector.</p>
Runtime Environment	<p>Name of the runtime environment where you want to run the tasks.</p>
Connection URL	<p>URL of the JNDI naming provider.</p> <p>For example, in IBM MQ it is the directory location that contains the .bindings file.</p>
JNDI User Name	<p>Optional. User name to connect to the JNDI context factory.</p>
JNDI Password	<p>Optional. The password of the user account that you use to connect to the JNDI context factory.</p>

Property	Description
JNDI Context Factory	The JMS provider specific initial JNDI context factory implementation for connecting to the JNDI service. This value is a fully qualified class name of the Initial Context Factory. For example, the class name of the Initial Context Factory for ActiveMQ is <code>org.apache.activemq.jndi.ActiveMQInitialContextFactory</code> For more information, see the documentation of the JMS provider.
JNDI Package Prefixes	A colon-delimited list of package prefixes to use when loading URL context factories. These are the package prefixes for the name of the factory class that will create a URL context factory. For more information about the values, see the documentation of the JMS provider.
JMS Connection Factory	The name of the object in the JNDI server that enables the JMS Client to create JMS connections. For example, <code>jms/QCF</code> or <code>jmsSalesSystem</code> .
JMS Connection User Name	Optional. User name to connect to the JMS connection factory.
JMS Connection Password	Optional. The password of the user account that you use to connect to the JMS connection factory.

**Note:** Ensure to copy the external JMS JAR files to the following location:

<Secure\_Agent\_home>/ext/connectors/thirdparty/infa.jms

After copying the external JMS JAR files, restart the Secure Agent.

## Kafka connection properties

When you set up a Kafka connection, you must configure the connection properties.

The following table describes the Kafka connection properties:

Property	Description
Connection Name	Name of the connection. The name is not case sensitive. It must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: <code>~ ` ! \$ % ^ &amp; * ( ) - + = { [ ]   \ : ; " ' &lt; , &gt; . ? /</code>
Description	Optional. Description that you use to identity the connection. The description cannot exceed 4,000 characters.
Type	The Kafka connection type. If you do not see the connection type, go to the <b>Add-On Connectors</b> page in Administrator to install the connector.
Runtime Environment	Name of the runtime environment where you want to run the tasks.

Property	Description
Kafka Broker List	<p>Comma-separated list of the Kafka brokers.</p> <p>To list a Kafka broker, use the following format:</p> <pre>&lt;HostName&gt;:&lt;PortNumber&gt;</pre> <p><b>Note:</b> When you connect to a Kafka broker over SSL, you must specify the fully qualified domain name for the host name. Otherwise, the test connection fails with SSL handshake error.</p>
Retry Timeout	<p>Optional. Number of seconds after which the Secure Agent attempts to reconnect to the Kafka broker to read or write data.</p> <p>Default is 180 seconds.</p> <p>This property is not used by Mass Ingestion Databases. You can specify an equivalent Kafka property in <b>Additional Connection Properties</b>.</p>
Kafka Broker Version	<p>Kafka message broker version. The only valid value is Apache 0.10.1.1 and above.</p> <p>Optional for a streaming ingestion task.</p>
Additional Connection Properties	<p>Optional. Comma-separated list of additional configuration properties of the Kafka producer or consumer.</p> <p>For a streaming ingestion task, ensure that you set the <code>&lt;kerberos name&gt;</code> property if you configure <code>&lt;Security Protocol&gt;</code> as <code>SASL_PLAINTEXT</code> or <code>SASL_SSL</code>.</p>
Schema Registry URL <sup>1</sup>	<p>Location and port of the Confluent schema registry service to access Avro sources and targets in Kafka.</p> <p>To list a schema registry URL, use the following format:</p> <pre>&lt;https&gt;://&lt;HostName or IP&gt;:&lt;PortNumber&gt;</pre> <p>or</p> <pre>&lt;http&gt;://&lt;HostName or IP&gt;:&lt;PortNumber&gt;</pre> <p>Example for the schema registry URL:</p> <pre>https://kafkarnd.informatica.com:8082</pre> <p>or</p> <pre>http://10.65.146.181:8084</pre> <p>Applies only when you import a Kafka topic in Avro format that uses the Confluent schema registry to store the metadata.</p> <p>This property is not used by Mass Ingestion Databases. You can specify an equivalent Kafka property in <b>Additional Connection Properties</b>.</p>
SSL Mode	<p>Required. Determines the encryption type to use for the connection.</p> <p>You can choose a mode from the following SSL modes:</p> <ul style="list-style-type: none"> <li>- Disabled. Establishes an unencrypted connection to the Kafka broker.</li> <li>- One-way. Establishes an encrypted connection to the Kafka broker using truststore file and truststore password.</li> <li>- Two-way. Establishes an encrypted connection to the Kafka broker using truststore file, truststore password, keystore file, and keystore password.</li> </ul> <p>This property is not used by Mass Ingestion Databases. You can specify an equivalent Kafka property in <b>Additional Connection Properties</b>.</p>
SSL TrustStore File Path	<p>Required when you use the one-way or two-way SSL mode.</p> <p>Absolute path and file name of the SSL truststore file. The truststore file contains the SSL certificate that the Kafka cluster validates against the Kafka broker certificate.</p>
SSL TrustStore Password	<p>Required when you use the one-way or two-way SSL mode.</p> <p>Password for the SSL truststore.</p>



Property	Description
SSL KeyStore File Path	Required when you use the two-way SSL mode. Absolute path and file name of the SSL keystore file that contains private keys and certificates that the Kafka broker validates against the Kafka cluster certificate.
SSL KeyStore Password	Required when you use the two-way SSL mode. Password for the SSL keystore.
Additional Security Properties	Optional. Comma-separated list of additional configuration properties to connect to the Kafka broker in a secured way. If you specify two different values for the same property in <b>Additional Connection Properties</b> and <b>Additional Security Properties</b> , the value in <b>Additional Security Properties</b> overrides the value in <b>Additional Connection Properties</b> . This property is not used by Mass Ingestion Databases.
<sup>1</sup> Does not apply to elastic mappings. The remaining properties are applicable for both mappings and elastic mappings.	

## Configuring the krb5.conf file to read data from or write to a Kerberised Kafka cluster

To read from or write to a Kerberised Kafka cluster, configure the default realm, KDC, and Kafka advanced source or target properties.

You can configure Kerberos authentication for a Kafka client by placing the required Kerberos configuration files on the Secure Agent machine and specifying the required JAAS configuration in the Kafka connection. The JAAS configuration defines the keytab and principal details that the Kafka broker must use to authenticate the Kafka client.

**Note:** This topic is not applicable to Mass Ingestion Applications and Mass Ingestion Databases. Mass Ingestion Applications and Mass Ingestion Databases does not yet support this functionality.

Before you read from or write to a Kerberised Kafka cluster, perform the following tasks:

1. Ensure that you have the `krb5.conf` file for the Kerberised Kafka cluster.
2. Configure the default realm and KDC. If the default `/etc/krb5.conf` file is not configured or you want to change the configuration, add the following lines to the `/etc/krb5.conf` file:

```
[libdefaults]
default_realm = <REALM NAME>
dns_lookup_realm = false
dns_lookup_kdc = false
ticket_lifetime = 24h
renew_lifetime = 7d
forwardable = true

[realms]
<REALM NAME> = {
kdc = <Location where KDC is installed>
admin_server = <Location where KDC is installed>
}

[domain_realm]
.<domain name or hostname> = <KERBEROS DOMAIN NAME>
<domain name or hostname> = <KERBEROS DOMAIN NAME>
```

3. To pass a static JAAS configuration file into the JVM using the `java.security.auth.login.config` property at runtime, perform the following tasks:

- a. Ensure that you have JAAS configuration file.

For information about creating JAAS configuration and configuring keytab for Kafka clients, see the Apache Kafka documentation at <https://kafka.apache.org/0101/documentation/#security>

For example, the JAAS configuration file can contain the following lines of configuration:

```
//Kafka Client Authentication. Used for client to kafka broker connection
KafkaClient {
  com.sun.security.auth.module.Krb5LoginModule required
  doNotPrompt=true
  useKeyTab=true
  storeKey=true
  keyTab="<path to Kafka keytab file>/<Kafka keytab file name>"
  principal="<principal name>"
  client=true
};
```

- b. Place the JAAS config file and keytab file in the same location on all the secure agents.

Informatica recommends that you place the files in a location that is accessible by all the secure agents in the runtime environment. For example, `/etc` or `/temp`.

- c. Configure the following properties:

#### Kafka connection

Configure the **Additional Connection Properties** in a Kafka connection and specify the value in the following format:

```
security.protocol=SASL_PLAINTEXT,sasl.kerberos.service.name=kafka,sasl.mechanism=GSSAPI
```

#### Sources

Configure the **Consumer Configuration Properties** in the advanced source properties to override the value specified in **Additional Connection Properties** in a Kafka connection. Specify the value in the following format:

```
security.protocol=SASL_PLAINTEXT,sasl.kerberos.service.name=kafka,sasl.mechanism=GSSAPI
```

#### Targets

Configure the **Producer Configuration Properties** in the advanced target properties to override the value specified in **Additional Connection Properties** in a Kafka connection. Specify the value in the following format:

```
security.protocol=SASL_PLAINTEXT,sasl.kerberos.service.name=kafka,sasl.mechanism=GSSAPI
```

4. To embed the JAAS configuration in the `sasl.jaas.config` configuration property, configure the following properties:

#### Kafka connection

Configure the **Additional Connection Properties** in a Kafka connection and specify the value in the following format:

```
security.protocol=SASL_PLAINTEXT,sasl.kerberos.service.name=kafka,sasl.mechanism=GSSAPI,
sasl.jaas.config=com.sun.security.auth.module.Krb5LoginModule required
useKeyTab=true
storeKey=true doNotPrompt=true serviceName="<service_name>" keyTab="<location of
keytab file>"
client=true principal="<principal_name>;"
```

## Sources

Configure the **Consumer Configuration Properties** in the advanced source properties to override the value specified in **Kerberos Configuration Properties** in a Kafka connection. Specify the value in the following format:

```
security.protocol=SASL_PLAINTEXT,sasl.kerberos.service.name=kafka,sasl.mechanism=
GSSAPI,
sasl.jaas.config=com.sun.security.auth.module.Krb5LoginModule required
useKeyTab=true
storeKey=true doNotPrompt=true serviceName="<service_name>" keyTab="<location of
keytab file>"
client=true principal="<principal_name>;
```

## Targets

Configure the **Producer Configuration Properties** in the advanced target properties to override the value specified in **Kerberos Configuration Properties** in a Kafka connection. Specify the value in the following format:

```
security.protocol=SASL_PLAINTEXT,sasl.kerberos.service.name=kafka,sasl.mechanism=
GSSAPI,
sasl.jaas.config=com.sun.security.auth.module.Krb5LoginModule required
useKeyTab=true
storeKey=true doNotPrompt=true serviceName="<service_name>" keyTab="<location of
keytab file>"
client=true principal="<principal_name>;
```

## Configuring SASL PLAIN authentication for a Kafka cluster

In the Kafka connection, you can configure PLAIN security for the Kafka broker to connect to a Kafka broker. To read data from or write data to a Kafka broker with SASL PLAIN authentication, configure the Kafka connection properties. To override the properties defined in the Kafka connection, you can configure the advanced source or target properties.

You can configure SASL PLAIN authentication so that the Kafka broker can authenticate the Kafka producer and the Kafka consumer. Kafka uses the Java Authentication and Authorization Service (JAAS) for SASL PLAIN authentication. To enable SASL PLAIN authentication, you must specify the SASL mechanism as PLAIN. You must also provide the formatted JAAS configuration that the Kafka broker must use for authentication. The JAAS configuration defines the username, password, that the Kafka broker must use to authenticate the Kafka client.

**Note:** This topic is not applicable to Mass Ingestion Applications and Mass Ingestion Databases. Mass Ingestion Applications and Mass Ingestion Databases does not yet support this functionality.

Configure the following properties:

### Kafka connection

Configure the **Additional Connection Properties** or **Additional Security Properties** property in the Kafka connection and specify the value in the following format:

```
security.protocol=SASL_SSL,sasl.mechanism=PLAIN,sasl.jaas.config=org.apache.kafka.com
mon.security.plain.PlainLoginModule required username=<username> password=<password>
```

In the **Security Configuration Section**, select **One-Way** as the **SSL Mode** and specify the SSL TrustStore File Path and SSL TrustStore Password.

## Sources

Configure the **Consumer Configuration Properties** property in the advanced source properties to override the value that you specified in the **Additional Connection Properties** property in the Kafka connection. Specify the value in the following format:

```
security.protocol=SASL_SSL,sasl.mechanism=PLAIN,sasl.jaas.config=org.apache.kafka.com
mon.security.plain.PlainLoginModule required username=<username> password=<password>
```

## Targets

Configure the **Producer Configuration Properties** property in the advanced target properties to override the value that you specified in the **Additional Connection Properties** property in the Kafka connection.

Specify the value in the following format:

```
security.protocol=SASL_SSL,sasl.mechanism=PLAIN,sasl.jaas.config=org.apache.kafka.com  
mon.security.plain.PlainLoginModule required username=<username> password=<password>
```

## Configuring SASL PLAIN authentication for an Azure Event Hub Kafka broker

In the Kafka connection, you can configure PLAIN security for the Kafka broker to connect to an Azure Event Hub Kafka broker. When you connect to an Azure Event Hub Kafka broker, the password defines the endpoint URL that contains the fully qualified domain name (FQDN) of the Event Hub namespace, shared access key name, and shared access key required to connect to an Azure Event Hub Kafka broker. Configure the SSL Mode as One-Way and provide the path to a trusted root certificate on your file system for SSL TrustStore File Path.

To connect to an Azure Event Hub Kafka broker, configure any of the above properties and specify the value in the following format:

```
security.protocol=SASL_SSL,sasl.mechanism=PLAIN,sasl.kerberos.service.name=Kafka,sasl.jaa  
s.config=org.apache.kafka.common.security.plain.PlainLoginModule required  
username="$ConnectionString" password="Endpoint=sb://<FQDN>/;SharedAccessKeyName=<key  
name>;SharedAccessKey=<shared access key>=";
```

## Configuring SASL\_SSL authentication for a Cloud Confluent Kafka cluster

In the Kafka connection, you can configure SSL security for encryption and authentication while connecting to a Kafka broker. To read data from or write data to a Confluent Kafka broker with SASL\_SSL authentication, configure the Kafka connection properties. To override the properties defined in the Kafka connection, you can configure the advanced source or target properties.

**Note:** This topic is not applicable to Mass Ingestion Applications and Mass Ingestion Databases. Mass Ingestion Applications and Mass Ingestion Databases does not yet support this functionality.

Configure the following properties:

Property	Values
Additional Connection Properties	security.protocol=SASL_SSL,sasl.kerberos.service.name=kafka,ssl.endpoint.identification.algorithm=required username=<> password=<>
SSL Mode	One-way
SSL TrustStore File Path	Use cacert file of agent JDK. For example: /root/staging/infaagent/jdk/jre/lib/security/cacerts
SSL TrustStore Password	Password for the SSL truststore.

## Marketo V3 connection properties

When you set up a Marketo V3 connection, you must configure the connection properties.

The following table describes the Marketo V3 connection properties:

Connection property	Description
Connection Name	Name of the Marketo V3 connection.
Description	Description of the connection. The description cannot exceed 765 characters.
Type	Type of connection. Select the Marketo V3 connection.
Runtime Environment	The name of the runtime environment where you want to run the tasks.
client_ID	The client ID of the custom service required to generate a valid access token.
client_secret	The client secret of the Marketo custom service required to generate a valid access token.
grant_type	The access permissions for an administrator to invoke the Marketo REST APIs to read data from and write data to Marketo. Marketo supports only the client_credentials grant type.
REST API URL	The URL has the following format: https://<Host name of the Marketo Rest API Server>. Contact the Marketo Administrator for the REST API URL.
Bypass Proxy	Connects to Marketo by using the proxy server settings defined in the proxy.ini file or through the Secure agent manager. When you select Bypass Proxy, you connect to Marketo using the Secure agent manager. When you clear Bypass Proxy, you connect to Marketo using the proxy server. Default is Bypass Proxy.

## Microsoft Azure Blob Storage V3 connection properties

When you create a Microsoft Azure Blob Storage V3 connection, you must configure the connection properties.

The following table describes Microsoft Azure Blob Storage V3 connection properties:

Connection property	Description
Runtime Environment	The name of the runtime environment where you want to run the tasks.
Account Name	Microsoft Azure Blob Storage account name.

Connection property	Description
Authentication Type	Authentication type to access the Microsoft Azure Blob Storage account. Select one of the following options: <ul style="list-style-type: none"> <li>- <b>Shared Key Authentication.</b> Select to use the account key to connect to Microsoft Azure Blob Storage.</li> <li>- <b>Shared Access Signature.</b> Select to use the SAS token to connect to Microsoft Azure Blob Storage. Use the SAS token to grant access to storage account resources for a specific time range without sharing the account key.</li> </ul>
Account Key	Applies to shared key authentication. The account key for the Microsoft Azure Blob Storage account.
SAS Token	Applies to shared access signature. The shared access signature token generated in the Azure portal.
Container Name	Microsoft Azure Blob Storage container name.
Endpoint Suffix	Type of Microsoft Azure end-points. You can select any of the following end-points: <ul style="list-style-type: none"> <li>- <code>core.windows.net</code>: Default</li> <li>- <code>core.usgovcloudapi.net</code>: To select the Azure Government end-points</li> <li>- <code>core.chinacloudapi.cn</code>:</li> </ul>

## Microsoft Azure Data Lake Storage Gen2 connection properties

When you set up a Microsoft Azure Data Lake Storage Gen2 connection, you must configure the connection properties.

The following table describes the Microsoft Azure Data Lake Storage Gen2 connection properties:

Connection property	Description
Connection Name	Name of the connection. The name is not case sensitive. It must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ & * ( ) - + = { [ ]   \ : ; " ' < , > . ? /
Description	Optional. Description that you use to identity the connection. The description cannot exceed 765 characters.
Type	The Microsoft Azure Data Lake Storage Gen2 connection type.
Runtime Environment	Name of the runtime environment where you want to run the tasks. You cannot run a database ingestion or streaming ingestion task on a Hosted Agent or serverless runtime environment.
Account Name	Microsoft Azure Data Lake Storage Gen2 account name or the service name.

Connection property	Description
Authentication Type	<p>Authentication type to access the Microsoft Azure Data Lake Storage Gen2 account.</p> <p>Select one of the following options:</p> <ul style="list-style-type: none"> <li>- <b>Service Principal Authentication.</b> Select to use the client ID, client secret, and tenant ID to connect to Microsoft Azure Data Lake Storage Gen2.</li> <li>- <b>Shared Key Authentication.</b> Select to use the account key to connect to Microsoft Azure Data Lake Storage Gen2.</li> <li>- <b>Managed Identity Authentication.</b> Select to authenticate using identities that are assigned to applications in Azure to access Azure resources in Microsoft Azure Data Lake Storage Gen2.</li> </ul> <p><b>Note:</b> Mass Ingestion Applications and Mass Ingestion Streaming do not support shared key authentication or managed identity authentication.</p>
Client ID	<p>Applies to Service Principal Authentication and Managed Identity Authentication.</p> <p>To use service principal authentication, specify the application ID or client ID for your application registered in the Azure Active Directory.</p> <p>To use managed identity authentication, specify the client ID for the user-assigned managed identity. If the managed identity is system-assigned, leave the field empty.</p>
Client Secret	<p>Applies to Service Principal Authentication.</p> <p>The client secret key to complete the OAuth authentication in the Azure Active Directory.</p>
Tenant ID	<p>Applies to Service Principal Authentication.</p> <p>The directory ID of the Azure Active Directory.</p>
Account Key	<p>Applies to Shared Key Authentication.</p> <p>The account key for the Microsoft Azure Data Lake Storage Gen2 account.</p>
File System Name	<p>The name of an existing file system in the Microsoft Azure Data Lake Storage Gen2 account.</p>
Directory Path	<p>The path of an existing directory without the file system name.</p> <p>You can select one of the following syntax:</p> <ul style="list-style-type: none"> <li>- / for root directory.</li> <li>- /dir1</li> <li>- dir1/dir2</li> </ul> <p>There is no default directory.</p>
Adls Gen2 End-point	<p>The type of Microsoft Azure endpoints.</p> <p>You can choose one of the following endpoints:</p> <ul style="list-style-type: none"> <li>- <code>core.windows.net</code>. Default</li> <li>- <code>core.usgovcloudapi.net</code>. To select the Azure Government endpoints.</li> </ul>

For more information about creating a client ID and a client secret, see *Microsoft Azure Data Lake Storage Gen2 documentation*.

## Microsoft Azure Event Hub connection properties

When you set up an Azure Event Hub connection, you must configure the connection properties.

The following table describes the Azure Event Hub connection properties:

Property	Description
Connection Name	Name of the connection. The name is not case sensitive. It must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ & * ( ) - + = { [ } ]   \ : ; " ' < , > . ? /
Description	Optional. Description that you use to identity the connection. The description cannot exceed 4,000 characters.
Type	The Azure Event Hub connection type. If you do not see the connection type, go to the <b>Add-On Connectors</b> page in Administrator to install the connector.
Runtime Environment	Name of the runtime environment where you want to run the tasks.
Tenant ID	The ID of the tenant that the data belongs to. This ID is the Directory ID of the Azure Active Directory.
Subscription ID	The ID of the Azure subscription.
Resource Group Name	The name of the Azure resource group associated with the Event Hub namespace.
Client Application ID	The ID of the application created under the Azure Active Directory.
Client Secret Key	The secret key generated for the application.
Event Hub Namespace	The name of the Event Hub namespace that is associated with the resource group name.
Shared Access Policy Name	Optional. The name of the Event Hub Namespace Shared Access Policy. The policy must apply to all data objects that are associated with this connection. To read from Event Hubs, you must have Listen permission. To write to an Event Hub, the policy must have Send permission.
Shared Access Policy Primary Key	Optional. The primary key of the Event Hub Namespace Shared Access Policy.

## Microsoft Azure Synapse Analytics Database Ingestion connection properties

When you define a Microsoft Azure Synapse Analytics Database Ingestion connection, you must configure connection properties. You can use this connection type in application ingestion tasks and database ingestion tasks, which you configure in the Mass Ingestion service.

**Note:** Some properties are for Microsoft Azure Data Lake Storage Gen2. Mass Ingestion Applications and Mass Ingestion Databases use Microsoft Azure Data Lake Storage Gen2 to stage data in files before sending the data to the Microsoft Azure Synapse Analytics target tables.



The following table describes the connection properties:

Property	Description
Connection Name	A name for the connection. This name must be unique within the organization. Connection names can contain alphanumeric characters, spaces, and the following special characters: _ . + - Spaces at the beginning or end of the name are trimmed and are not saved as part of the name. Maximum length is 100 characters. Connection names are not case sensitive.
Description	An optional description for the connection. Maximum length is 255 characters.
Type	The type of connection. Ensure that the type is for Microsoft Azure Synapse Analytics - Database Ingestion.
Runtime Environment	The name of the runtime environment where you want to run the application ingestion tasks or database ingestion tasks. You define runtime environments in Administrator. <b>Note:</b> You cannot run application ingestion tasks and database ingestion tasks on a Hosted Agent or serverless runtime environment.
Azure Synapse Analytics JDBC URL	The Microsoft Azure Synapse Analytics (formerly SQL Data Warehouse) JDBC connection string. Example connection string for Microsoft SQL Server authentication: <code>jdbc:sqlserver://server.database.windows.net:1433;database=database</code> Example connection string for Azure Active Directory (AAD) authentication: <code>jdbc:sqlserver://server.database.windows.net:1433;database=database;encrypt=true;trustServerCertificate=false;hostNameInCertificate=*.database.windows.net;loginTimeout=30;Authentication=ActiveDirectoryPassword;</code> <b>Note:</b> The default authentication type is Microsoft SQL Server authentication.
Azure Synapse Analytics JDBC Username	The user name to use for connecting to the Microsoft Azure Synapse Analytics account. Provide the AAD user name for AAD authentication.
Azure Synapse Analytics JDBC Password	The password to use for connecting to the Microsoft Azure Synapse Analytics account.
Azure Synapse Analytics Schema Name	The name of the schema in the Microsoft Azure Synapse Analytics target.
ADLS Gen2 Account Name	The name of the Microsoft Azure Data Lake Storage Gen2 account.
Client Id	The ID of your client application for completing the OAuth Authentication in the Active Directory.
Client Secret	The client secret key for completing the OAuth Authentication in the Active Directory.
Directory	The Microsoft Azure Data Lake Storage Gen2 directory that Mass Ingestion Applications and Mass Ingestion Databases uses to stage data in files. The default is the root directory.
Filesystem Name	The name of an existing file system in the Microsoft Azure Data Lake Storage Gen2 account.
Tenant ID	The Directory ID of the Azure Active Directory.

## Microsoft Azure Synapse SQL connection properties

When you set up a Microsoft Azure Synapse SQL connection, you must configure the connection properties.

The following table describes the Microsoft Azure Synapse SQL connection properties:

Connection property	Description
Runtime Environment	The name of the runtime environment where you want to run the tasks. Select a Secure Agent, Hosted Agent, or serverless runtime environment.
Azure DW JDBC URL	Microsoft Azure Synapse SQL JDBC connection string. Example for Microsoft SQL Server authentication: <code>jdbc:sqlserver://&lt;Server&gt;.database.windows.net:1433;database=&lt;Database&gt;</code> Example for Azure Active Directory (AAD) authentication: <code>jdbc:sqlserver://&lt;Server&gt;.database.windows.net:1433; database=&lt;Database&gt;;encrypt=true;trustServerCertificate=false; hostNameInCertificate=*.database.windows.net;loginTimeout=30; Authentication=ActiveDirectoryPassword;</code> The default authentication is Microsoft SQL Server authentication.
Azure DW JDBC Username	User name to connect to the Microsoft Azure Synapse SQL account. Provide AAD user name for AAD authentication.
Azure DW JDBC Password	Password to connect to the Microsoft Azure Synapse SQL account.
Azure DW Schema Name	Name of the schema in Microsoft Azure Synapse SQL.
Azure Storage Type	Type of Azure storage to stage the files. You can select any of the following storage type: <ul style="list-style-type: none"><li>- Azure Blob. Default. To use Microsoft Azure Blob Storage to stage the files.</li><li>- ADLS Gen2. To use Microsoft Azure Data Lake Storage Gen2 as storage to stage the files.</li></ul>
Authentication Type	Authentication type to connect to Azure storage to stage the files. Select one of the following options: <ul style="list-style-type: none"><li>- <b>Shared Key Authentication.</b> Select to use the account name and account key to connect to Microsoft Azure Blob Storage or Microsoft Azure Data Lake Storage Gen2.</li><li>- <b>Service Principal Authentication.</b> Applies to Microsoft Azure Data Lake Storage Gen2. To use Service Principal authentication, you must register an application in the Azure Active Directory, generate a client secret, and then assign the Storage Blob Contributor role to the application</li><li>- <b>Managed Identity Authentication.</b> Applies to Microsoft Azure Data Lake Storage Gen2. Select to authenticate using identities that are assigned to applications in Azure to access Azure resources in Microsoft Azure Data Lake Storage Gen2.</li></ul> In a file ingestion task, if you select Microsoft Azure Synapse SQL with Managed Identity authentication type as the target, then you must select Microsoft Azure Data Lake Storage Gen2 as the source.
Azure Blob Account Name	Applies to Shared Key Authentication for Microsoft Azure Blob Storage. Name of the Microsoft Azure Blob Storage account to stage the files.
Azure Blob Account Key	Applies to Shared Key Authentication for Microsoft Azure Blob Storage. Microsoft Azure Blob Storage access key to stage the files.

Connection property	Description
ADLS Gen2 Storage Account Name	Applies to Shared Key Authentication and Service Principal Authentication for Microsoft Azure Data Lake Storage Gen2. Name of the Microsoft Azure Data Lake Storage Gen2 account to stage the files.
ADLS Gen2 Account Key	Applies to Shared Key Authentication for Microsoft Azure Data Lake Storage Gen2. Microsoft Azure Data Lake Storage Gen2 access key to stage the files.
Client ID	Applies to Service Principal Authentication and Managed Identity Authentication for Microsoft Azure Data Lake Storage Gen2. To use service principal authentication, specify the application ID or client ID for your application registered in the Azure Active Directory. To use managed identity authentication, specify the client ID for the user-assigned managed identity. If the managed identity is system-assigned, leave the field empty.
Client Secret	Applies to Service Principal Authentication for Microsoft Azure Data Lake Storage Gen2. The client secret for your application.
Tenant ID	Applies to Service Principal Authentication for Microsoft Azure Data Lake Storage Gen2. The directory ID or tenant ID for your application.
Blob End-point	Type of Microsoft Azure endpoints. You can select one of the following endpoints: - <code>core.windows.net</code> . Default. - <code>core.usgovcloudapi.net</code> . To select the Azure Government endpoints.
VNet Rule	Enable to connect to a Microsoft Azure Synapse SQL endpoint residing in a virtual network (VNet). When you use a serverless runtime environment, you cannot connect to a Microsoft Azure Synapse SQL endpoint residing in a virtual network.

## Microsoft Dynamics 365 Mass Ingestion connection properties

When you set up a Microsoft Dynamics 365 Mass Ingestion connection, you must configure the connection properties.

The Microsoft Dynamics 365 Mass Ingestion connection requires a native application that is registered in Azure Active Directory (Azure AD) to access the Microsoft Dynamics 365 data. Before you configure the connection, you must register an application in Azure AD to allow the connection to access the Microsoft Dynamics 365 data. For more information about registering an application in Azure AD, see the [Microsoft documentation](#).

The properties of a Microsoft Dynamics 365 Mass Ingestion connection vary based on the authentication method that you specify for the connection. When you create a connection, you can select one of the following authentication methods:

- **OAuth 2.0 Username-Password Flow:** Authenticates the connection by using the Microsoft Dynamics 365 account login credentials and the client ID of the application registered in Azure AD.
- **OAuth 2.0 Client Secret Flow:** Authenticates the connection by using the client ID and client secret of the application registered in Azure AD.

- **OAuth 2.0 JWT Bearer Flow:** Authenticates the connection by using a X509 Public Key Infrastructure (PKI) certificate and a JSON Web Token (JWT). Use this authentication method to gain secured access to Microsoft Dynamics 365 without sharing sensitive information, such as client secret and Microsoft Dynamics 365 account login credentials.

### Connection properties for OAuth 2.0 Username-Password Flow authentication

The following table describes the connection properties for a Microsoft Dynamics 365 Mass Ingestion connection configured with OAuth 2.0 Username-Password Flow authentication:

Connection property	Description
Runtime Environment	Name of the runtime environment where you want to run the ingestion tasks. You must specify a Secure Agent as the runtime environment. <b>Note:</b> You cannot run application ingestion tasks on a Hosted Agent or serverless runtime environment.
User Name	User name of the Microsoft Dynamics 365 account.
Password	Password for the Microsoft Dynamics 365 account.
Client ID	Client ID of the application registered in Azure AD.
Resource URL	URL of the Microsoft Dynamics 365 organization. You must enter the resource URL in the following format: <code>https://&lt;Microsoft_Dynamics_365_org_name&gt;.api.crm8.dynamics.com</code>
OAuth Token URL	OAuth 2.0 token endpoint of the Microsoft Dynamics 365 organization. The application that is registered in Azure AD sends the access token requests to this endpoint. You must enter the following value in this field: <code>https://login.windows.net/common/oauth2/token</code>

**Note:** For more information about the OAuth 2.0 Username-Password Flow authentication method, see the Microsoft Dynamics 365 documentation.

### Connection properties for OAuth 2.0 Client Secret Flow authentication

The following table describes the connection properties for a Microsoft Dynamics 365 Mass Ingestion connection configured with OAuth 2.0 Client Secret Flow authentication:

Connection property	Description
Runtime Environment	Name of the runtime environment where you want to run the ingestion tasks. You must specify a Secure Agent as the runtime environment. <b>Note:</b> You cannot run application ingestion tasks on a Hosted Agent or serverless runtime environment.
Client ID	Client ID of the application registered in Azure AD.
Client Secret	Client secret of the application registered in Azure AD.

Connection property	Description
Resource URL	URL of the Microsoft Dynamics 365 organization. You must enter the resource URL in the following format: <code>https://&lt;Microsoft_Dynamics_365_org_name&gt;.api.crm8.dynamics.com</code>
OAuth Token URL	OAuth 2.0 token endpoint of the Microsoft Dynamics 365 organization. The application that is registered in Azure AD sends the access token requests to this endpoint. You must enter the following value in this field: <code>https://login.microsoftonline.com/&lt;tentant_id&gt;/oauth2/token</code>

**Note:** For more information about the OAuth 2.0 Client Secret Flow authentication method, see the Microsoft Dynamics 365 documentation.

### Connection properties for OAuth 2.0 JWT Bearer Flow authentication

The following table describes the connection properties for a Microsoft Dynamics 365 Mass Ingestion connection configured with OAuth 2.0 JWT Bearer Flow authentication:

Connection property	Description
Runtime Environment	Name of the runtime environment where you want to run the ingestion tasks. You must specify a Secure Agent as the runtime environment. <b>Note:</b> You cannot run application ingestion tasks on a Hosted Agent or serverless runtime environment.
Client ID	Client ID of the application registered in Azure AD.
Certificate Signature	Base64URL string that encodes the hexadecimal value which represents the SHA-1 thumbprint of the X509 certificate.
Keystore Path	Absolute path to the keystore file that contains the X509 certificate required to validate a JSON Web Token (JWT) and establish a secure connection with Microsoft Dynamics 365. The keystore file must be in the Java KeyStore (JKS) format.
Keystore Password	Password for the keystore file.
Private Key Alias	Alias name of the private key used to sign the JWT.
Private Key Password	Password for the private key.
Audience for JWT	URL of the Microsoft Dynamics 365 resource server to which the application that is registered in Azure AD sends the JWT for validation. You must enter the address in the following format: <code>https://login.microsoftonline.com/&lt;tentant_id&gt;/oauth2/token</code>

Connection property	Description
Resource URL	URL of the Microsoft Dynamics 365 organization. You must enter the resource URL in the following format: <code>https://&lt;Microsoft_Dynamics_365_org_name&gt;.api.crm8.dynamics.com</code>
OAuth Token URL	OAuth 2.0 token endpoint of the Microsoft Dynamics 365 organization. The application that is registered in Azure AD sends the access token requests to this endpoint. You must enter the following value in this field: <code>https://login.microsoftonline.com/&lt;tentant_id&gt;/oauth2/token</code>

**Note:** For more information about the OAuth 2.0 Client Secret Flow authentication method, see the Microsoft Dynamics 365 documentation.

## Microsoft SQL Server connection properties

When you set up a Microsoft SQL Server connection, configure the connection properties.

The following table describes the Microsoft SQL Server connection properties:

Connection property	Description
Runtime Environment	The name of the runtime environment where you want to run the tasks. You cannot run a database ingestion task on a Hosted Agent or serverless runtime environment.
SQL Server Version	Microsoft SQL Server database version.

Connection property	Description
Authentication Mode	<p>Authentication method to access Microsoft SQL Server. Select one of the following methods:</p> <ul style="list-style-type: none"> <li>- <b>SQL Server Authentication.</b> Enter your Microsoft SQL Server user name and password to access Microsoft SQL Server.</li> <li>- <b>Windows Authentication (Deprecated).</b> Use the Microsoft Windows authentication to access Microsoft SQL Server. This option is available when you access Data Integration or Mass Ingestion by using Microsoft Windows. When you choose this option, you do not need to enter credentials to access Microsoft SQL Server. To use Windows authentication in an SQL Server connection, ensure that the user account that starts the Secure Agent service is available in the Microsoft SQL Server database.</li> </ul> <p>If you use Mass Ingestion Databases and want to use Windows authentication, select this option.</p> <p><b>Note:</b> Windows authentication is not certified for Microsoft SQL Server 2017 version hosted on Linux. You cannot configure Windows Authentication when you use a serverless runtime environment.</p> <ul style="list-style-type: none"> <li>- <b>Active Directory Password.</b> Enter the Azure Active Directory user name and password to authenticate and access the Microsoft Azure SQL Database.</li> <li>- <b>Windows Authentication v2.</b> Use this authentication method to access Microsoft SQL Server from Data Integration using the agent hosted on a Linux or Windows machine. When you choose the Windows Authentication v2 mode, enter your domain name and your Microsoft Windows credentials to access Microsoft SQL Server. To use Windows Authentication v2 in an SQL Server connection using the Windows agent, ensure that the user account that starts the Secure Agent service is available in the Microsoft SQL Server database.</li> </ul> <p><b>Note:</b> You cannot configure Windows Authentication when you use a serverless runtime environment.</p>
Domain	<p>Required for Windows Authentication v2.</p> <p>The domain name of the Windows user.</p>
User Name	<p>User name for the database login. The user name cannot contain a semicolon.</p> <p>To connect to Microsoft Azure SQL Database, you must specify the user name in the following format: <code>username@host</code></p> <p>For Windows Authentication v2, you must specify the Windows NT user name.</p> <p><b>Note:</b> This property is not applicable if you use the Windows Authentication mode to access Microsoft SQL Server.</p>
Password	<p>Password for the database login. The password cannot contain a semicolon.</p> <p>For Windows Authentication v2, you must specify the Windows NT password.</p> <p><b>Note:</b> This property is not applicable if you use the Windows Authentication mode to access Microsoft SQL Server.</p>
Host	<p>Name of the machine hosting the database server.</p> <p>To connect to Microsoft Azure SQL Database, specify the fully qualified host name. For example, <code>vmjcmwxsfboheng.westus.cloudapp.azure.com</code></p>
Port	<p>Network port number used to connect to the database server. Default port number is 1433.</p>
Instance Name	<p>Instance name of the Microsoft SQL Server database.</p>
Database Name	<p>Database name for the Microsoft SQL Server target. Database name is case sensitive if the database is case sensitive. Maximum length is 100 characters.</p> <p>Database names can include alphanumeric and underscore characters.</p>

Connection property	Description
Schema	Schema used for the target connection.
Code Page	The code page of the database server.
Encryption Method	<p>The method that the Secure Agent uses to encrypt the data sent between the driver and the database server. You can use the encryption method to connect to Microsoft Azure SQL Database.</p> <p><b>Note:</b> When you use a serverless runtime environment, you cannot configure a Microsoft SQL Server connection to use SSL to securely communicate with the Microsoft SQL Server database.</p>
Crypto Protocol Version	Cryptographic protocols to use when you enable SSL encryption.
Validate Server Certificate	<p>When set to True, Secure Agent validates the certificate that is sent by the database server. If you specify the HostNameInCertificate parameter, Secure Agent also validates the host name in the certificate.</p> <p>When set to false, the Secure Agent does not validate the certificate that is sent by the database server.</p>
Trust Store	<p>The location and name of the trust store file. The trust store file contains a list of Certificate Authorities (CAs) that the driver uses for SSL server authentication.</p> <p>For the serverless runtime environment, specify the following certificate path in the serverless agent directory:</p> <pre>/home/cldagnt/SystemAgent/serverless/configurations/ssl_store/ &lt;TrustStore_filename&gt;</pre>
Trust Store Password	The password to access the contents of the trust store file.
Host Name in Certificate	Host name of the machine that hosts the secure database. If you specify a host name, the Secure Agent validates the host name included in the connection with the host name in the SSL certificate.
Metadata Advanced Connection Properties	The additional properties for the JDBC driver to fetch the metadata. If you specify more than one property, separate each key-value pair with a semicolon.
Runtime Advanced Connection Properties	The additional properties for the ODBC driver for running mappings. If you specify more than one property, separate each key-value pair with a semicolon.



## MongoDB Mass Ingestion connection properties

When you set up a MongoDB Mass Ingestion connection, you must configure the connection properties.

The following table describes MongoDB Mass Ingestion connection properties:

Connection property	Description
Connection Name	The name of the connection. The name is not case sensitive and must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 255 characters, contain spaces, or contain the following special characters: ~`!\$%^&*()-+= {[}]  \ ; " ' < , > . ? /
Description	Optional. A description of the connection. The description cannot exceed 4,000 characters.
Type	Type of connection. You must select <b>MongoDB Mass Ingestion</b> .
Runtime Environment	The name of the runtime environment where you want to run the tasks.
Host and Port	An SRV record or a comma-separated list of <i>host_name:port</i> pairs.
SRV	Select this check box if you specified an SRV record in <b>Host and Port</b> property.
User Name	User name for logging in to the database.
Password	Password for the specified database user.
Authentication Database	The name of the authentication database associated with the specified user.
Replica Set Name	The name of the replica set that is composed of the MongoDB servers with replicas of the source data.
Additional Connection Properties	One or more additional connection properties that you want to use. Specify the properties as key-value pairs. If you specify more than one property, separate them with the ampersand symbol (&). The connection properties are case sensitive. Example: <code>authSource=admin&amp;replicaSet=rsprimary</code>

## MQTT connection properties

When you set up an MQ Telemetry Transport (MQTT) connection, you must configure the connection properties.

The following table describes the MQTT connection properties:

Property	Description
Connection Name	Name of the connection. The name is not case sensitive and must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ & * ( ) - + = { [ ] }   \ : ; " ' < , > . ? /
Description	Optional. Description that you can use to identity the connection. The description cannot exceed 4,000 characters.
Type	The MQTT connection type. If you do not see the connection type, go to the <b>Add-On Connectors</b> page to install the connector.
Runtime Environment	Name of the runtime environment where you want to run the tasks.
Broker URI	The connection URL of the MQTT broker. If specified, this value overrides the URL specified in the main portion of the URL. Sample URL: <code>tcp://&lt;IP Address&gt;:&lt;port&gt;</code>
Client Id	Client identifier of your MQTT client. If this value is left blank, the MQTT server assigns a unique value. This property value must be unique for each MQTT client connecting to a specific MQTT server. Sharing projects without changing the Client ID can lead to connection issues, including disconnections and missing updates.
Username	Username to use when connecting to the broker.
Password	Password to use when connecting to the broker.
Connection Timeout	Maximum time interval the client will wait for the connection to the MQTT server to be established. Default timeout is 30 seconds. A value of 0 disables timeout processing. That is, the client waits until the network connection is made successfully or fails.
Use SSL	Enable this option to use SSL for secure transmission. If you enable the SSL authentication, ensure to provide both keystore and truststore details for using the MQTT connection in a streaming ingestion task.
Keystore Filename	Contains the keys and certificates required for secure communication.
Keystore Password	Password for the keystore filename.

Property	Description
Keystore Type	Type of keystore to use. Keystore type defines the storage and data format of the keystore information and the algorithms used to protect private keys in the keystore. Use one of the following types: <ul style="list-style-type: none"> <li>- JKS. Stores private keys and certificates.</li> <li>- PKCS12. Stores private keys, secret keys. and certificates.</li> </ul>
Truststore Filename	File name of the truststore file.
Truststore Password	Password for the truststore file name.
Truststore Type	Type of truststore to use. Use one of the following types: <ul style="list-style-type: none"> <li>- JKS</li> <li>- PKCS 12</li> </ul>
TLS Protocol	Transport protocols to use. Use one of the following types: <ul style="list-style-type: none"> <li>- SSL</li> <li>- SSLv3</li> <li>- TLS</li> <li>- TLSv1</li> <li>- TLSv1.1</li> <li>- TLSv1.2</li> </ul>

## MySQL connection properties

When you set up a MySQL connection, configure the connection properties.

The following table describes MySQL connection properties:

Connection property	Description
Runtime Environment	The name of the runtime environment where you want to run the tasks. <b>Note:</b> You cannot run a database ingestion task on a Hosted Agent or serverless runtime environment.
User Name	User name for the database login. The user name cannot contain a semicolon.
Password	Password for the database login. The password cannot contain a semicolon.
Host	Name of the machine that hosts the database server.
Port	Network port number used to connect to the database server. Default is 3306.
Database Name	Name of the MySQL database that you want to connect to. <b>Note:</b> The database name is case sensitive. Maximum length is 64 characters. Valid characters are alphanumeric and underscore characters.
Code Page	The code page of the database server.

Connection property	Description
Metadata Advanced Connection Properties	The additional properties for the JDBC driver to fetch the metadata. If you specify more than one property, separate each key-value pair with a semicolon.
Runtime Advanced Connection Properties	The additional properties for the ODBC driver for running mappings. If you specify more than one property, separate each key-value pair with a semicolon.

## Netezza connection properties

When you set up a Netezza connection, you must configure the connection properties.

The following table describes the Netezza connection properties:

Connection property	Description
Runtime Environment	The name of the runtime environment where you want to run the tasks.
Database	The name of the Netezza database.
Schemaname	The schema used for the Netezza source or target.
Servername	The Netezza database host name.
Port	Network port number used to connect to the database server. Default is 1521.
Driver	The Netezza ODBC driver name, NetezzaSQL, used to connect to the Netezza database.
Runtime Additional Connection Configuration	Additional run-time attributes required to fetch data. For example, <code>securityLevel=preferredUnSecured;caCertFile =</code>
Metadata Additional Connection Configuration	The values to set the optional properties of the JDBC driver to fetch the metadata.
Username	Database user name with the appropriate read and write database permissions to access the database.
Password	Password for the database user name.

## NetSuite Mass Ingestion connection properties

When you set up a NetSuite Mass Ingestion connection, you must configure the connection properties.

**Note:** Before you configure the connection properties, install the SuiteAnalytics Connect JDBC driver and copy the NQjc.jar file to the following directory: `<Secure_Agent>\ext\connectors\thirdparty\informatica.netsuiteami`

For more information about installing the SuiteAnalytics Connect JDBC driver, see the [SuiteAnalytics Connect documentation](#).

The following table describes the connection properties for a NetSuite Mass Ingestion connection:

Connection property	Description
Runtime Environment	Name of the runtime environment where you want to run the ingestion tasks. You must specify a Secure Agent as the runtime environment. <b>Note:</b> You cannot run application ingestion tasks on a Hosted Agent or serverless runtime environment.
Email ID	User name of the NetSuite account. The user name is an email address.
Password	Password for the NetSuite account.
Service Host	Name of the SuiteAnalytics Connect Service host. The value in this field must match the value specified in the <b>Service Host</b> field under the <b>Your Configuration</b> section of the <b>SuiteAnalytics Connect Driver Download</b> page in NetSuite. To access the <b>SuiteAnalytics Connect Driver Download</b> page, log in to NetSuite and click the Set Up SuiteAnalytics Connect link in the Settings portlet.
Service Port	TCP/IP port on which the SuiteAnalytics Connect server is listening. Default is 1708.
Account ID	NetSuite account ID. To find your account ID, log in to NetSuite and click <b>Setup &gt; Integration &gt; Web Services Preferences</b> . If you cannot access the <b>Setup</b> menu, navigate to <b>Support &gt; Go to Suite Answers &gt; Contact support by phone</b> . The page displays your account ID.
Role ID	Role ID associated with the NetSuite account.

## OPC UA connection properties

When you set up an OPC UA connection, you must configure the connection properties.

The following table describes the OPC UA connection properties:

Property	Description
Connection Name	Name of the connection. The name is not case sensitive and must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ & * ( ) - + = { [ ]   \ : ; " ' < , > . ? /
Description	Optional. Description of the connection. The description cannot exceed 4,000 characters.
Type	The OPC UA connection type.
Runtime Environment	Name of the runtime environment where you want to run the tasks.

Property	Description
Endpoint URL	<p>The unique URL to connect to the OPC UA server.</p> <p>The endpoint URL identifies the specific instance of a server and a security policy type. A valid endpoint URL consists of the endpoint type (opc.tcp), the endpoint host name (IP address, URL, or DSN), and the endpoint port number.</p> <p>For example, <code>opc.tcp://opcuaserver.com:48010</code></p>
Security Policy	<p>The security policy used to connect to the OPC UA server.</p> <p>The security policy parameters specify the security algorithms that the OPC UA server supports.</p> <p>You can choose one of the following security policies:</p> <ul style="list-style-type: none"> <li>- None. No security provided.</li> <li>- Basic128Rsa15</li> <li>- Basic256</li> <li>- Basic256Sha256</li> <li>- Aes128_Sha256_RsaOaep</li> <li>- Aes256_Sha256_RsaPss</li> </ul> <p><b>Note:</b> The OPC Foundation deprecated the security policies, Basic128Rsa15 and Basic256 as of OPC UA specification version 1.04. The encryption provided by these policies is less secure. Use these security policies only to provide backward compatibility.</p>
Security Mode	<p>The security mode used to connect to the OPC UA server.</p> <p>The security mode is valid only when security policy is not set to None. You can choose one of the following security policies:</p> <ul style="list-style-type: none"> <li>- Sign. Transfer unencrypted data, but with digital signatures that allow verification of data integrity.</li> <li>- SignAndEncrypt. Transfer signed and encrypted data.</li> </ul>
Application URI	<p>Optional. A unique identifier that the OPC UA application can use to connect to the OPC UA server.</p> <p>Enter a unique ID in the following format:</p> <p><code>urn:aaa:bbb</code></p> <p>For example, <code>urn:nifi:opcua</code></p> <p>The unique identifier must match the URI of the Subject Alternative Name of your OPC UA client certificate.</p>
Client Keystore Location	<p>Optional. Absolute path and file name of the keystore file that contains private keys and certificates for the OPC UA server.</p> <p>Enter the path in the following format:</p> <p><code>/root/opcua/client.jks</code></p> <p>The keystore must contain only one keypair entry of private key and certificate. If multiple keypair entries exist, the first entry is used.</p>
Client Keystore Password	Optional. Password for the client keystore.
Require server authentication	Optional. Enable if you require server authentication of client certificates, client authentication of server certificates, or both.
Trust store Location	<p>Optional. The absolute path of the truststore file that contains the trusted certificate.</p> <p>Enter the path in the following format:</p> <p><code>/root/opcua/trust.jks</code></p>

Property	Description
Trust store Password	Password for the truststore file.
Authentication Policy	Authentication settings required to establish the connections. You can choose one of the following authentication policies: <ul style="list-style-type: none"> <li>- Anon. Anonymous authentication. Anonymous tokens are associated with servers that do not require user authentication.</li> <li>- UserName. User name and password tokens are associated with servers with any password based system, such as Windows.</li> </ul>
User Name	User name to access the OPC UA server if you choose authentication policy as <b>UserName</b> .
Password	Password to access the OPC UA server if you choose authentication policy as <b>UserName</b> .

## Oracle Database Ingestion connection properties

When you define an Oracle Database Ingestion connection for a database ingestion task, you must configure connection properties.

The following table describes the connection properties:

Property	Description
Connection Name	A name for the connection. This name must be unique within the organization. Connection names can contain alphanumeric characters, spaces, and the following special characters: _ . + - Spaces at the beginning or end of the name are trimmed and are not saved as part of the name. Maximum length is 100 characters. Connection names are not case sensitive.
Description	An optional description for the connection. Maximum length is 255 characters.
Type	The type of connection. For an Oracle Database Ingestion connection, the type must be <b>Oracle Database Ingestion</b> .
Runtime Environment	The name of the runtime environment where you want to run database ingestion tasks. You define runtime environments in Administrator.
User Name	User name for the Oracle database login. The user name cannot contain a semicolon.
Password	Password for the Oracle database login. The password cannot contain a semicolon.
Host	Host name of the database server.
Port	Network port number used to connect to the database server. Default is 1521.

Property	Description
Service Name	Service name or System ID (SID) that uniquely identifies the Oracle database. Specify the SID in the following format to connect to Oracle databases: SID:<ORACLE_SID>
Schema	Schema used for the Oracle connection.
Code Page	The code page of the database server. Database ingestion tasks use the UTF-8 code page. Default is UTF-8.
Encryption Method	For initial load jobs, determines whether the data exchanged between the Secure Agent and the Oracle database server is encrypted: Options are: <ul style="list-style-type: none"> <li>- <b>SSL</b>. Establishes a secure connection using SSL for data encryption. If the Oracle database server cannot configure SSL, the connection fails.</li> <li>- <b>No Encryption</b>. Establishes a connection without using SSL. Data is not encrypted.</li> </ul> Default is No Encryption.
Crypto Protocol Version	If you selected SSL as the encryption method, you must specify a cryptographic protocol or a list of cryptographic protocols supported by your server to use with an encrypted connection. Options are: <ul style="list-style-type: none"> <li>- SSLv2</li> <li>- SSLv3</li> <li>- TLSv1</li> <li>- TLSv1.1</li> <li>- TLSv1.2</li> </ul> Default is TLSv1.2.
Validate Server Certificate	If you selected SSL as the encryption method, controls whether the Secure Agent validates the server certificate that is sent by the Oracle database server. <ul style="list-style-type: none"> <li>- <b>True</b>. Validate the server certificate.</li> <li>- <b>False</b>. Do not validate the server certificate.</li> </ul> Default is False. If you also specify the <b>Host Name in Certificate</b> property, the Secure Agent also validates the host name in the certificate.
Trust Store	If you selected SSL as the encryption method and enabled validation of the server certificate, specify the path and name of the truststore file, which contains the list of the Certificate Authorities (CAs) that the client trusts for SSL authentication.
Trust Store Password	If you selected SSL as the encryption method and enabled validation of the server certificate, specify a password for accessing the contents of the truststore file.



Property	Description
Host Name in Certificate	If you selected SSL as the encryption method and enabled validation of the server certificate, specify the host name of the machine that hosts the Oracle database to provide for additional security. The Secure Agent validates the host name included the connection with the host name in the SSL certificate.
Key Store	If you selected SSL as the encryption method and client authentication is enabled on the Oracle database server, specify the path and name of the keystore file. The keystore file contains the certificates that the client sends to the Oracle server in response to the server's certificate request.
Key Store Password	If you selected SSL as the encryption method and client authentication is enabled on the Oracle database server, specify the password for the keystore file.
Key Password	If you selected SSL as the encryption method and client authentication is enabled on the Oracle database server, specify the password for the keys in the keystore file. Use this property when the keys have a different password than the keystore file.
Database Connect String	An Oracle connection string, defined in TNS, that database ingestion tasks use to connect to the Oracle database.
TDE Wallet Directory	The path and file name for the Oracle wallet file that is used for Oracle Transparent Data Encryption (TDE). Specify this property value only if you capture change data from TDE-encrypted tablespaces and one of the following conditions are true: <ul style="list-style-type: none"> <li>- The Oracle wallet is not available to the database.</li> <li>- The Oracle database is running on a server that is remote from Oracle redo logs.</li> <li>- The wallet directory is not in the default location on the database host or the wallet name is not the default name of ewallet.p12.</li> <li>- The wallet directory is not available to the Secure Agent host.</li> </ul>
TDE Wallet Password	A clear text password that is required to access the Oracle TDE wallet and get the master key. This property value is required if you need to read and decrypt data from TDE-encrypted tablespaces in the Oracle source database.

Property	Description
Directory Substitution	<p>A local path prefix to substitute for the server path prefix of the redo logs on the Oracle server. This substitute local path is required when the log reader runs on a system other than the Oracle server and uses a different mapping to access the redo log files. Use this property in the following situations:</p> <ul style="list-style-type: none"> <li>- The redo logs reside on shared disk.</li> <li>- The redo logs have been copied to a system other than the Oracle system.</li> <li>- The archived redo logs are accessed by using a different NFS mount.</li> </ul> <p><b>Note:</b> Do not use this statement if you use Oracle Automatic Storage Management (ASM) to manage the redo logs.</p> <p>You can define one or more substitutions. Use the following format:</p> <pre>server_path_prefix,local_path_prefix;server_path_prefix,local_path_prefix;...</pre>
Reader Active Log Mask	<p>A mask that the log reader uses for selecting active redo logs when the Oracle database uses multiplexing of redo logs. The log reader compares the mask against the member names in an active redo log group to determine which log to read. In the mask, you can use the asterisk (*) wildcard to represent zero or more characters.</p> <p>The mask can be up to 128 characters in length. It is case-sensitive on Linux or UNIX systems but not on Windows systems.</p>
Reader Archive Destination 1	<p>The primary log destination from which the log reader reads archived logs, when Oracle is configured to write more than one copy of each archived redo log. Enter a number that corresponds to a <i>n</i> value in an Oracle LOG_ARCHIVE_DEST_<i>n</i> initialization parameter, where <i>n</i> is a value from 1 to 10.</p> <p>If you set only one of the Reader Archive Destination 1 and Destination 2 properties, the log reader uses that property setting. If you specify neither property, the archive log queries are not filtered by the log destination.</p>
Reader Archive Destination 2	<p>The secondary log destination from which the log reader reads archived logs when the primary destination becomes unavailable or when the logs at the primary destination cannot be read. For example, logs might have been corrupted or deleted. Enter a number that corresponds to the <i>n</i> value in an Oracle LOG_ARCHIVE_DEST_<i>n</i> initialization parameter, where <i>n</i> is a value from 1 to 10. Usually, this value is a number greater than 1.</p>
Reader ASM Connect String	<p>In an Oracle ASM environment, the Oracle connection string, defined in TNS, that the log reader uses to connect to the ASM instance that manages storage of active and archived redo logs for the source database.</p>

Property	Description
Reader ASM User Name	In an Oracle ASM environment, an Oracle user ID that the log reader uses to connect to the ASM instance that manages storage of active and archived redo logs for the source database. This user ID must have SYSDBA or SYSASM authority. To use SYSASM authority, set the <b>Reader ASM Connect As SYSASM</b> property to Y.
Reader ASM Password	In an Oracle ASM environment, a clear text password for the user that is specified in the <b>Reader ASM User Name</b> property. The log reader uses this password and the ASM user name to connect to the ASM instance that manages storage of active and archived redo logs for the source database.
Reader ASM Connect As SYSASM	If you use Oracle 11g ASM or later and want the log reader to use a user ID that has SYSASM authority to connect to the ASM instance, select this check box. Also specify a user ID that has SYSASM authority in the <b>Reader ASM User Name</b> property. To use a user ID that has SYSDBA authority, clear this check box. By default, this check box is cleared.
Reader Mode	<p>Indicates the source of and types of Oracle redo logs that the log reader reads. Valid options are:</p> <ul style="list-style-type: none"> <li>- <b>ACTIVE</b>. Read active and archived redo logs from the Oracle online system. Optionally, you can use the <b>Reader Active Log Mask</b> property to filter the active redo logs and use the <b>Reader Archive Destination 1</b> and <b>Reader Archive Destination 2</b> properties to limit the archived log destinations from which to read archived logs.</li> <li>- <b>ARCHIVEONLY</b>. Read only archived redo logs. Optionally, you can use the <b>Reader Archive Destination 1</b> and <b>Reader Archive Destination 2</b> properties to limit the archived log destinations from which to read archived logs.</li> <li>- <b>ARCHIVECOPY</b>. Read archived redo logs that have been copied to an alternate file system. Use this option in the following situations: <ul style="list-style-type: none"> <li>- You do not have the authority to access the Oracle archived redo logs directly.</li> <li>- The archived redo logs are written to ASM, but you do not have access to ASM.</li> <li>- The archived log retention policy for the database server causes the archived logs to not be retained long enough.</li> </ul> <p>With this option, the <b>Reader Archive Destination 1</b> and <b>Reader Archive Destination 2</b> properties are ignored. Default is <b>ACTIVE</b>.</p> </li> </ul>

Property	Description
Reader Standby Log Mask	<p>A mask that the log reader uses for selecting redo logs for an Oracle standby database when the database uses multiplexing of redo logs. The log reader compares the mask against the member names in an redo log group to determine which log to read. In the mask, you can use the asterisk (*) wildcard to represent zero or more characters.</p> <p>The mask can be up to 128 characters in length. It is case-sensitive on Linux or UNIX systems but not on Windows systems.</p>
Standby Connect String	An Oracle connection string, defined in TNS, that the log reader uses to connect to the Oracle physical standby database for change capture when the database is not open for read-only access.
Standby User Name	A user ID that the log reader uses to connect to the Oracle physical standby database for change capture. This user ID must have SYSDBA authority.
Standby Password	A clear text password that the log reader uses to connect to the Oracle physical standby database for change capture.
RAC Members	<p>The maximum number of active redo log threads, or <i>members</i>, in an Oracle Real Application Cluster (RAC) that can be tracked. For a Data Guard physical standby database that supports a primary database in a RAC environment, this value is the number of active threads for the primary database.</p> <p>Valid values are 1 to 100. Default is 0, which causes an appropriate number of log threads to be determined automatically. If this value is not appropriate for your environment, set this property to a value greater than 0.</p>
BFILE Access	<p>Select this check box in the following circumstances:</p> <ul style="list-style-type: none"> <li>- You use BFILE access to redo logs in physical directories on the local Oracle server file system. BFILE access uses Oracle directory objects to remotely access the redo logs in the file system. This method is an alternative to other log access methods such as ASM or NFS mounts.</li> <li>- You have an Amazon Relational Database Service (RDS) for Oracle source. In this case, this option enables access to the redo logs of a cloud-based database instance deployed in RDS.</li> </ul> <p>By default, this check box is cleared.</p>

## Oracle Fusion Cloud Mass Ingestion connection properties

When you set up an Oracle Fusion Cloud Mass Ingestion connection, you must configure the connection properties.

**Note:** Oracle Fusion Cloud Mass Ingestion connections can access the data of only Enterprise Resource Planning (ERP) and Oracle Supply Chain and Manufacturing (SCM) modules of Oracle Fusion Cloud Applications Suite.

The following table describes the connection properties for an Oracle Fusion Cloud Mass Ingestion connection:

Connection property	Description
Runtime Environment	Name of the runtime environment where you want to run the ingestion tasks. You must specify a Secure Agent as the runtime environment. <b>Note:</b> You cannot run application ingestion tasks on a Hosted Agent or serverless runtime environment.
Authentication	Authentication method of the connection. By default, the connection uses the Basic authentication method.
User Name	User name of the Oracle Cloud account.
Password	Password for the Oracle Cloud account.
Server URL	URL of the Oracle Cloud service that you want to access.
API Version	Version of the Oracle Cloud REST API that you want to use for the connection.

**Note:** Effective in the April 2022 release, Oracle Fusion Cloud Mass Ingestion connector is available for preview.

Preview functionality is supported for evaluation purposes but is unwarranted and is not supported in production environments or any environment that you plan to push to production. Informatica intends to include the preview functionality in an upcoming release for production use, but might choose not to in accordance with changing market or technical circumstances. For more information, contact Informatica Global Customer Support.

## PostgreSQL connection properties

When you set up an PostgreSQL connection, configure the connection properties.

The following table describes the PostgreSQL connection properties:

Connection property	Description
Runtime Environment	The name of the runtime environment where you want to run the tasks. You cannot run a database ingestion task on a Hosted Agent or serverless runtime environment.
Host Name	Host name of the PostgreSQL server to which you want to connect.
Port	Port number for the PostgreSQL server to which you want to connect. Default is 5432.
Schema	Optional. The schema name. If you do not specify the schema name, all the schemas available in the database are listed while importing the source object in Data Integration.
Database	The PostgreSQL database name.
User Name	User name to access the PostgreSQL database.

Connection property	Description
Password	Password for the PostgreSQL database user name.
Encryption Method	<p>Determines whether the data exchanged between the Secure Agent and the PostgreSQL database server is encrypted:</p> <p>Select one of the following encryption methods:</p> <ul style="list-style-type: none"> <li>- noEncryption. Establishes a connection without using SSL. Data is not encrypted.</li> <li>- SSL. Establishes a connection using SSL. Data is encrypted using SSL. If the PostgreSQL database server cannot configure SSL, the connection fails.</li> <li>- requestSSL. Attempts to establish a connection using SSL. If the PostgreSQL database server cannot configure SSL, the Secure Agent establishes an unencrypted connection.</li> </ul> <p>Default is noEncryption.</p> <p><b>Note:</b> SSL is not applicable when you use the Hosted Agent. You can configure SSL when you use the Secure Agent or the serverless runtime environment.</p>
Validate Server Certificate	<p>Applicable if you select SSL or requestSSL as the encryption method.</p> <p>Select the Validate Server Certificate option so that the Secure Agent validates the server certificate that is sent by the PostgreSQL database server. If you specify the Host Name In Certificate property, the Secure Agent also validates the host name in the certificate.</p>
TrustStore	<p>Applicable if you select SSL or requestSSL as the encryption method and the Validate Server Certificate option.</p> <p>The path and name of the truststore file, which contains the list of the Certificate Authorities (CAs) that the PostgreSQL client trusts.</p> <p>For the serverless runtime environment, specify the following certificate path in the serverless agent directory:</p> <pre>/home/cldagnt/SystemAgent/serverless/configurations/ssl_store/ &lt;TrustStore_filename&gt;</pre>
TrustStore Password	<p>Applicable if you select SSL or requestSSL as the encryption method and the Validate Server Certificate option.</p> <p>The password to access the truststore file that contains the SSL certificate.</p>
Host Name In Certificate	<p>Optional when you select SSL or requestSSL as the encryption method and the Validate Server Certificate option.</p> <p>A host name for providing additional security. The Secure Agent validates the host name included in the connection with the host name in the SSL certificate.</p>
KeyStore	<p>Applicable if you select SSL as the encryption method and when client authentication is enabled on the PostgreSQL database server.</p> <p>The path and the file name of the key store. The keystore file contains the certificates that the PostgreSQL client sends to the PostgreSQL server in response to the server's certificate request.</p> <p>For the serverless runtime environment, specify the following certificate path in the serverless agent directory:</p> <pre>/home/cldagnt/SystemAgent/serverless/configurations/ssl_store/ &lt;KeyStore_filename&gt;</pre>
KeyStore Password	<p>Applicable if you select SSL as the encryption method and when client authentication is enabled on the PostgreSQL database server.</p> <p>The password for the keystore file required for secure communication.</p>

Connection property	Description
Key Password	Applicable if you select SSL as the encryption method and when client authentication is enabled on the PostgreSQL database server. Required when individual keys in the keystore file have a different password than the keystore file.
Additional Connection Properties	Additional connection parameters that you want to use. You must provide the connection parameters as semicolon-separated key-value pairs.
Crypto Protocol Versions	Required if you select SSL or requestSSL as the encryption method. A cryptographic protocol or a list of cryptographic protocols to use with an encrypted connection. You can select from the following protocols: <ul style="list-style-type: none"> <li>- SSLv3</li> <li>- TLSv1</li> <li>- TLSv1_1</li> <li>- TLSv1_2</li> </ul> Default is TLSv1_2. The TLSv1 and TLSv1_1 protocols are not applicable.

## Salesforce Mass Ingestion connection properties

When you set up a Salesforce Mass Ingestion connection, you must configure the connection properties.

The Salesforce Mass Ingestion connection uses a connected app to access the Salesforce data. Before you configure the connection, you must configure a connected app in Salesforce to allow the connection to access the Salesforce data.

**Note:** For more information about configuring a connected app, see the Knowledge Base article [000172095](#).

The properties of a Salesforce Mass Ingestion connection vary based on the authentication method that you specify for the connection. When you create a connection, you can select one of the following authentication methods:

- **OAuth 2.0 Username-Password Flow:** Authenticates the connection by using the Salesforce account login credentials and the consumer key and consumer secret that Salesforce generates for the connected app.
- **OAuth 2.0 JWT Bearer Flow:** Authenticates the connection by using the Salesforce account user name, private key alias, private key password, and the consumer key that Salesforce generates for the connected app. Informatica recommends that you use this authentication method because this method provides secured access to Salesforce without sharing sensitive information, such as consumer secret and Salesforce account password.

## Connection properties for OAuth 2.0 Username-Password Flow authentication

The following table describes the connection properties for a Salesforce Mass Ingestion connection configured with OAuth 2.0 Username-Password Flow authentication:

Connection property	Description
Runtime Environment	Name of the runtime environment where you want to run the ingestion tasks. You must specify a Secure Agent as the runtime environment. <b>Note:</b> You cannot run application ingestion tasks on a Hosted Agent or serverless runtime environment.
User Name	User name of the Salesforce account.
Password	Password for the Salesforce account.
Security Token	Security token associated with the Salesforce account. <b>Note:</b> If you do not have the security token, reset the security token in Salesforce. For more information about resetting the security token, see the <a href="#">Salesforce documentation</a> .
Consumer Key	Consumer key that Salesforce generates when you enable OAuth 2.0 authentication for the connected app.
Consumer Secret	Consumer secret that Salesforce generates when you enable OAuth 2.0 authentication for the connected app.
Base URI	URL of the Salesforce organization. You must enter the base URI in the following format: <code>https://&lt;salesforce_org&gt;.salesforce.com</code>
OAuth token URL	OAuth 2.0 token endpoint of the Salesforce organization. The connected app sends the access token requests to this endpoint. Default value is: <code>https://login.salesforce.com/services/oauth2/token</code>

**Note:** For more information about the OAuth 2.0 Username-Password Flow authentication method, see the Salesforce documentation.

## Connection properties for OAuth 2.0 JWT Bearer Flow authentication

The following table describes the connection properties for a Salesforce Mass Ingestion connection configured with OAuth 2.0 JWT Bearer Flow authentication:

Connection property	Description
Runtime Environment	Name of the runtime environment where you want to run the ingestion tasks. You must specify a Secure Agent as the runtime environment. <b>Note:</b> You cannot run application ingestion tasks on a Hosted Agent or serverless runtime environment.
User Name	User name of the Salesforce account.
Consumer Key	Consumer key that Salesforce generates when you enable OAuth 2.0 authentication for the connected app.



Connection property	Description
Keystore Path	Absolute path to the keystore file that contains the X509 certificate required to validate a JSON Web Token (JWT) and establish a secure connection with Salesforce. The keystore file must be in the Java KeyStore (JKS) format.
Keystore Password	Password for the keystore file.
Private Key Alias	Alias name of the private key used to sign the JWT.
Private Key Password	Password for the private key.
Base URI	URL of the Salesforce organization. You must enter the base URI in the following format: <code>https://&lt;salesforce_org&gt;.salesforce.com</code>
OAuth token URL	OAuth 2.0 token endpoint of the Salesforce organization. The connected app sends the access token requests to this endpoint. Default value is: <code>https://login.salesforce.com/services/oauth2/token</code>

**Note:** For more information about the OAuth 2.0 JWT Bearer Flow authentication method, see the Salesforce documentation.

## SAP HANA Database Ingestion connection properties

When you set up an SAP HANA connection, you must configure the connection properties.

The following table describes the SAP HANA connection properties:

Connection property	Description
Connection Name	A name for the connection. This name must be unique within the organization. Connection names can contain alphanumeric characters, spaces, and the following special characters: _ . + - Spaces at the beginning or end of the name are trimmed and are not saved as part of the name. Maximum length is 100 characters. Connection names are not case sensitive.
Description	An optional description for the connection. Maximum length is 255 characters.
Type	The type of connection. Ensure that the type is <b>SAP HANA Database Ingestion</b> .
Runtime Environment	The name of the runtime environment where you want to run database ingestion tasks. You define runtime environments in Administrator.
User Name	The user name to use for connecting to the SAP HANA instance.
Password	The password to use for connecting to the SAP HANA instance.
Host	The name of the machine that hosts the SAP HANA database server.

Connection property	Description
Port	The port number for the SAP HANA server to which you want to connect. Default is 30015.
Database Name	The SAP HANA source database name.
Advanced Connection Properties	Advanced properties for the SAP HANA JDBC driver, which is used to connect to the SAP HANA source. If you specify more than one <i>property=value</i> entry, separate them with an ampersand (&). The JDBC connection properties that you can enter in this field are described in the SAP <a href="#">JDBC Connection Properties</a> documentation. For example: encrypt=true.
Log Clear	<p>Required for incremental loads. The time interval, in days, after which the PKLOG table entries are purged. The purging occurs while an incremental load job is running. Valid values for a database ingestion job are 0 to 366. Default is 14. A value of 0 means that the PKLOG table entries are not purged. Any value outside the range of 0 to 366, including a negative number or non-numeric value, causes database ingestion jobs that use the connection to fail with the following error:</p> <pre>LogClear contains a non numeric number. Caused by: LogClear contains a non numeric number.</pre>
Trigger Prefix	Adds a prefix to the names of the AFTER DELETE, AFTER INSERT, and AFTER UPDATE triggers that the CDC script generates for each source table to get before images and after images of the DML changes. Enter any prefix value up to 16 characters in length. An underscore (_) follows the prefix in the trigger name, for example, <b>TX_SAP_DEMO_TABLE_DBMI_USER_t_d</b> . You can use the prefix to comply with your site's trigger naming conventions.

**Note:** If you test the connection and the test fails, check that the SAP HANA JDBC driver file, ngdbc.jar, has been installed at *Secure Agent installation directory>/ext/connectors/thirdparty/informatica.hanami*.

## SAP ODP Extractor connection properties

Select the **SAP ODP Extractor** connection type and configure the connection properties.

The following table describes the SAP ODP Extractor connection properties:

Connection property	Description
Runtime Environment	Runtime environment that contains the Secure Agent that you want to use to access SAP S/4HANA or SAP ECC.
SAP Server Connection Type	<p>The SAP server connection type to use.</p> <p>Select from the following options:</p> <ul style="list-style-type: none"><li>- <b>Application Server Connection.</b> Connect to an SAP Application Server using the SAP user name and password.</li><li>- <b>Application Server SNC Connection.</b> Connect to an SAP Application Server using the secured network connection:<ul style="list-style-type: none"><li>- With X.509 Certificate. You do not need to specify the SAP user name and password explicitly. You must provide the path of the x.509 certificate file.</li><li>- Without X.509 Certificate. You must provide the SAP user name.</li></ul></li><li>- <b>Load Balancing Server Connection.</b> Connect to an SAP Application Server with the least load at run time.</li><li>- <b>Load Balancing Server SNC Connection.</b> Connect to an SAP Application Server using SNC with the least load at run time.</li></ul> <p><b>Note:</b> Before you use an SNC connection, you must verify that SNC is configured both on the SAP Server and the machine where the Secure Agent runs.</p>

The following table describes the properties that must configure when you select **Application Server Connection** as the connection type:

Connection property	Description
SAP Client Number	The client number of the SAP Server.
SAP Language	Language code that corresponds to the SAP language.
SAP Application Server	The host name of the SAP Application Server.
SAP System Number	The system number of the SAP Server to connect.
SAP Username	The SAP user name with the appropriate user authorization.
SAP Password	The SAP password.

Connection property	Description
Subscriber Name	A name which defines the Secure Agent as a unique subscriber in the SAP system. SAP uses this name to define unique operational delta queue (ODQ) in case of delta read from ODP.
Additional Parameters	<p>Additional SAP parameters that the Secure Agent uses to connect to the SAP system. For example, to generate SAP JCo and SAP CPIC trace, specify the following properties:</p> <pre>jco.client.trace="1"; jco.client.cpic_trace="3";</pre> <p>During the runtime, the JCo and CPIC traces file are generated in the following location:        &lt;Informatica Secure Agent installation directory&gt;\apps        \Data_Integration_Server\&lt;DIS version&gt;\ICS\main\bin\rdtm</p> <p>During the design time, the CPIC traces are generated in the tomcat.out files at the following location:        &lt;Informatica Secure Agent installation directory&gt;\apps        \Data_Integration_Server\&lt;DIS version&gt;tomcat.out</p>

The following table describes the properties that must configure when you select **Load Balancing Server Connection** as the connection type:

Connection property	Description
SAP Client Number	The client number of the SAP Server.
SAP Language	Language code that corresponds to the SAP language.
SAP Message Server	Host name of the SAP Message Server.
SAP System ID	The system ID of the SAP Message Server.
SAP Group	The login group name, for example, PUBLIC.
SAP Username	The SAP user name with the appropriate user authorization.
SAP Password	The SAP password.

Connection property	Description
Subscriber Name	A name which defines the Secure Agent as a unique subscriber in the SAP system. SAP uses this name to define unique operational delta queue (ODQ) in case of delta read from ODP.
Additional Parameters	<p>Additional SAP parameters that the Secure Agent uses to connect to the SAP system. For example, to generate SAP JCo and SAP CPIC trace, specify the following properties:</p> <pre>jco.client.trace="1"; jco.client.cpic_trace="3";</pre> <p>During the runtime, the JCo and CPIC traces file are generated in the following location: &lt;Informatica Secure Agent installation directory&gt;\apps \Data_Integration_Server\&lt;DIS version&gt;\ICS\main\bin\rdtm</p> <p>During the design time, the CPIC traces are generated in the tomcat.out files at the following location: &lt;Informatica Secure Agent installation directory&gt;\apps \Data_Integration_Server\&lt;DIS version&gt;tomcat.out</p>

The following table describes the properties that must configure when you select **Application Server SNC Connection** as the connection type:

Connection property	Description
SAP Client Number	The client number of the SAP Server.
SAP Language	Language code that corresponds to the SAP language.
SAP Application Server	The host name of the SAP Application Server.
SAP System Number	The system number of the SAP Server to connect.
SNC My Name	Optional. The Informatica client Personal Security Environment (PSE) or certificate name. Default length is 256.
SNC Partner Name	The Informatica client PSE or certificate name. Default length is 256.
SNC Quality of Protection (QoP)	<p>Specifies the SAP PSE or certificate name. You can select from the following options:</p> <ul style="list-style-type: none"> <li>- 1 - Apply authentication only.</li> <li>- 2 - Apply integrity protection (authentication).</li> <li>- 3 - Apply privacy protection (integrity and authentication).</li> <li>- 8 - Apply the default protection.</li> <li>- 9 - Apply the maximum protection.</li> </ul> <p>Default is 3 - <i>Apply privacy protection (integrity and authentication)</i>.</p>
SAP Cryptographic Library Path	<p>The path to the cryptographic library. Specify <code>sapcrypto.dll</code> for Windows or <code>libsapcrypto.so</code> for Linux.</p>
Use X509 Certificate	Specifies the quality of protection. Select to use X509 Certificate based SNC connection.

Connection property	Description
X509 Certificate Path or SAP Username	<p>The path to the X509 certificate file.</p> <p>If you select to use the X509 certificate, specify the path to the X509 certificate file with .crt extension. You do not need to specify the SAP user name and password.</p> <p>If you do not want to use the X509 certificate, specify the SAP user name for which SNC is configured in SAP Server.</p>
Subscriber Name	<p>A name which defines the Informatica Secure Agent as a unique subscriber in the SAP system.</p> <p>SAP uses this name to define unique operational delta queue (ODQ) when the Secure Agent reads delta data from ODP.</p>
Additional Parameters	<p>Additional SAP parameters that the Secure Agent uses to connect to the SAP system.</p> <p>For example, to generate SAP JCo and SAP CPIC trace, specify the following properties:</p> <pre>jco.client.trace="1"; jco.client.cpic_trace="3";</pre> <p>During the runtime, the JCo and CPIC traces file are generated in the following location:</p> <pre>&lt;Informatica Secure Agent installation directory&gt;\apps \Data_Integration_Server\&lt;DIS version&gt;\ICS\main\bin\rdtm</pre> <p>During the design time, the CPIC traces are generated in the tomcat.out files at the following location:</p> <pre>&lt;Informatica Secure Agent installation directory&gt;\apps \Data_Integration_Server\&lt;DIS version&gt;tomcat.out</pre>

The following table describes the properties that must configure when you select **Load Balancing Server SNC Connection** as the connection type:

Connection property	Description
SAP Client Number	The client number of the SAP Server.
SAP Language	Language code that corresponds to the SAP language.
SAP Message Server	Host name of the SAP Message Server.
SAP System ID	The system ID of the SAP Message Server.
SAP Group	The login group name, for example, PUBLIC.
SNC My Name	<p>Optional. The Informatica client PSE or certificate name generated on the Secure Agent machine.</p> <p>Default length is 256.</p>
SNC Partner Name	<p>The Informatica client PSE or certificate name generated on the SAP Server.</p> <p>Default length is 256.</p>

Connection property	Description
SNC Quality of Protection (QoP)	<p>Specifies the SAP PSE or certificate name.</p> <p>You can select from the following options:</p> <ul style="list-style-type: none"> <li>- 1 - Apply authentication only.</li> <li>- 2 - Apply integrity protection (authentication).</li> <li>- 3 - Apply privacy protection (integrity and authentication).</li> <li>- 8 - Apply the default protection.</li> <li>- 9 - Apply the maximum protection.</li> </ul> <p>Default is 3 - <i>Apply privacy protection (integrity and authentication)</i>.</p>
SAP Cryptographic Library Path	<p>The path to the cryptographic library.</p> <p>Specify <code>sapcrypto.dll</code> for Windows or <code>libsapcrypto.so</code> for Linux.</p>
Use X509 Certificate	<p>Specifies the quality of protection. Select to use X509 Certificate based SNC connection.</p>
X509 Certificate Path or SAP Username	<p>The path to the X509 certificate file.</p> <p>If you select to use the X509 certificate, specify the path to the X509 certificate file with <code>.crt</code> extension. You do not need to specify the SAP user name and password.</p> <p>If you do not want to use the X509 certificate, specify the SAP user name for which SNC is configured in SAP Server.</p>
Subscriber Name	<p>A name which defines the Informatica Secure Agent as a unique subscriber in the SAP system.</p> <p>SAP uses this name to define unique operational delta queue (ODQ) when the Secure Agent reads delta data from ODP.</p>
Additional Parameters	<p>Additional SAP parameters that the Secure Agent uses to connect to the SAP system.</p> <p>For example, to generate SAP JCo and SAP CPIC trace, specify the following properties:</p> <pre>jco.client.trace="1"; jco.client.cpic_trace="3";</pre> <p>During the runtime, the JCo and CPIC traces file are generated in the following location:</p> <pre>&lt;Informatica Secure Agent installation directory&gt;\apps \Data_Integration_Server\&lt;DIS version&gt;\ICS\main\bin\rdtm</pre> <p>During the design time, the CPIC traces are generated in the <code>tomcat.out</code> files at the following location:</p> <pre>&lt;Informatica Secure Agent installation directory&gt;\apps \Data_Integration_Server\&lt;DIS version&gt;tomcat.out</pre>

## ServiceNow Mass Ingestion connection properties

When you set up a ServiceNow Mass Ingestion connection, you must configure the connection properties.

The properties of a ServiceNow Mass Ingestion connection vary based on the authentication method that you specify for the connection. When you create a connection, you can select one of the following authentication methods:

- **OAuth 2.0:** Authenticates the connection by using the details of the OAuth API endpoint that is created for the connection in ServiceNow. To use this method, you must create OAuth API endpoint in ServiceNow and then specify the client ID and client secret of the API endpoint in the connection properties. For more information about creating an OAuth API endpoint in ServiceNow, see the [ServiceNow documentation](#).
- **Basic:** Authenticates the connection by validating the login credentials of the ServiceNow account.

## Connection properties for OAuth 2.0 authentication

The following table describes the connection properties for a ServiceNow Mass Ingestion connection configured with OAuth 2.0 authentication:

Connection property	Description
Runtime Environment	Name of the runtime environment where you want to run the ingestion tasks. You must specify a Secure Agent as the runtime environment. <b>Note:</b> You cannot run application ingestion tasks on a Hosted Agent or serverless runtime environment.
User Name	User name of the ServiceNow account.
Password	Password for the ServiceNow account.
Client Secret	Client secret of the API endpoint created for the connection in ServiceNow.
Client ID	Client ID of the API endpoint created for the connection in ServiceNow.
Base URI	URL of the ServiceNow instance. You must enter the base URI in the following format: <code>https://{your_servicenow_instance}.service-now.com/</code>
OAuth Token URL	OAuth token endpoint of the ServiceNow instance. The API client associated with the connection sends the access token requests to this endpoint.

## Connection properties for Basic authentication

The following table describes the connection properties for a ServiceNow Mass Ingestion connection configured with Basic authentication:

Connection property	Description
Runtime Environment	Name of the runtime environment where you want to run the ingestion tasks. You must specify a Secure Agent as the runtime environment. <b>Note:</b> You cannot run application ingestion tasks on a Hosted Agent or serverless runtime environment.
User Name	User name of the ServiceNow account.
Password	Password for the ServiceNow account.
Base URI	URL of the ServiceNow instance. You must enter the base URI in the following format: <code>https://{your_servicenow_instance}.service-now.com/</code>



## Snowflake Data Cloud connection properties

When you set up a Snowflake Data Cloud connection, configure the connection properties.

You can use the following authentication methods to connect to Snowflake:

- Standard. Uses Snowflake account user name and password credentials to connect to Snowflake.

**Note:** For application ingestion tasks, you can use only the Standard authentication method.

- Authorization Code. Uses the OAuth 2.0 protocol with Authorization Code grant type to connect to Snowflake. Authorization Code allows authorized access to Snowflake without sharing or storing your login credentials.
- KeyPair. Uses the private key file and private key file password, along with the existing Snowflake account user name to connect to Snowflake.

You create a Snowflake Data Cloud connection on the Connections page. You can then use the connection when you read from or write data to Snowflake.

### Standard authentication

When you set up a Snowflake Data Cloud connection, configure the connection properties.

The following table describes the Snowflake Data Cloud connection properties for the Standard authentication mode:

Connection property	Description
Connection Name	Name of the connection. The name is not case sensitive. It must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ & * ( ) - + = { [ ] }   \ : ; " ' < , > . ? /
Description	Optional. Description that you use to identify the connection. The description cannot exceed 765 characters.
Type	The Snowflake Data Cloud connection type.
Runtime Environment	The name of the runtime environment where you want to run the tasks. You cannot run application ingestion tasks and database ingestion tasks on a Hosted Agent or serverless runtime environment.
Authentication	Select the authentication method that the connector must use to log in to Snowflake. Select <b>Standard</b> . Default is <b>Standard</b> .
Username	The user name to connect to the Snowflake account.
Password	The password to connect to the Snowflake account.

Connection property	Description
Account	<p>The name of the Snowflake account.</p> <p>For example, if the Snowflake URL is <a href="https://123abc.us-east-2.aws.snowflakecomputing.com/console/login#/">https://123abc.us-east-2.aws.snowflakecomputing.com/console/login#/</a>, your account name is the first segment in the URL. Here, <i>123abc.us-east-2</i> is your account name.</p> <p>For the Snowsight URL, for example, <a href="https://app.snowflake.com/us-east-2.aws/123abc/dashboard">https://app.snowflake.com/us-east-2.aws/123abc/dashboard</a>, your account name is <i>123abc.us-east-2</i>.</p> <p><b>Note:</b> Ensure that the account name does not contain underscores. To use an alias name, contact Snowflake Customer Support.</p>
Warehouse	The Snowflake warehouse name. You must specify the warehouse name.
Role	The Snowflake role assigned to the user.
Additional JDBC URL Parameters	<p>Optional. The additional JDBC connection parameters.</p> <p>Enter one or more JDBC connection parameters in the following format:</p> <pre>&lt;param1&gt;=&lt;value&gt;&amp;&lt;param2&gt;=&lt;value&gt;&amp;&lt;param3&gt;=&lt;value&gt;...</pre> <p>For example:</p> <pre>user=jon&amp;warehouse=mywh&amp;db=mydb&amp;schema=public</pre> <p><b>Important:</b> Ensure that there is no space before and after = when you add the parameters.</p>

## OAuth 2.0 authorization code authentication

The following table describes the Snowflake Data Cloud connection properties for an OAuth 2.0 - AuthorizationCode type connection:

Connection property	Description
Connection Name	<p>Name of the connection.</p> <p>The name is not case sensitive. It must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ &amp; * ( ) - + = { [ ]   \ : ; " ' &lt; , &gt; . ? /</p>
Description	<p>Optional. Description that you use to identify the connection.</p> <p>The description cannot exceed 765 characters.</p>
Type	The Snowflake Data Cloud connection type.
Runtime Environment	<p>The name of the runtime environment where you want to run the tasks.</p> <p>You cannot run application ingestion tasks and database ingestion tasks on a Hosted Agent or serverless runtime environment.</p>
Authentication	The authentication method that Snowflake Data Cloud Connector must use to log in to Snowflake. Select <b>AuthorizationCode</b> .

Connection property	Description
Account	<p>The name of the Snowflake account.</p> <p>For example, if the Snowflake URL is <a href="https://123abc.us-east-2.aws.snowflakecomputing.com/console/login#/">https://123abc.us-east-2.aws.snowflakecomputing.com/console/login#/</a>, your account name is the first segment in the URL. Here, <i>123abc.us-east-2</i> is your account name.</p> <p>For the Snowsight URL, for example, <a href="https://app.snowflake.com/us-east-2.aws/123abc/dashboard">https://app.snowflake.com/us-east-2.aws/123abc/dashboard</a>, your account name is <i>123abc.us-east-2</i>.</p> <p><b>Note:</b> Ensure that the account name does not contain underscores. To use an alias name, contact Snowflake Customer Support.</p>
Warehouse	The Snowflake warehouse name.
Additional JDBC URL Parameters	<p>Optional. The additional JDBC connection parameters.</p> <p>Enter one or more JDBC connection parameters in the following format:</p> <pre>&lt;param1&gt;=&lt;value&gt;&amp;&lt;param2&gt;=&lt;value&gt;&amp;&lt;param3&gt;=&lt;value&gt;...</pre> <p>For example:</p> <pre>user=jon&amp;warehouse=mywh&amp;db=mydb&amp;schema=public</pre> <p>Ensure that there is no space before and after = when you add the parameters.</p>
Authorization URL	<p>The Snowflake server endpoint that is used to authorize the user request.</p> <p>The authorization URL is <a href="https://&lt;account_name&gt;.snowflakecomputing.com/oauth/authorize">https://&lt;account_name&gt;.snowflakecomputing.com/oauth/authorize</a>, where &lt;account_name&gt; specifies the full name of your account provided by Snowflake.</p> <p>For example, <a href="https://informatica.snowflakecomputing.com/oauth/authorize">https://informatica.snowflakecomputing.com/oauth/authorize</a></p> <p><b>Note:</b> If the account name contains underscores, use the alias name.</p> <p>You can also use the Authorization Code grant type that supports the authorization server in a Virtual Private Cloud network.</p>
Access Token URL	<p>The Snowflake access token endpoint that is used to exchange the authorization code for an access token.</p> <p>The access token URL is <a href="https://&lt;account_name&gt;.snowflakecomputing.com/oauth/token-request">https://&lt;account_name&gt;.snowflakecomputing.com/oauth/token-request</a>, where &lt;account_name&gt; specifies the full name of your account provided by Snowflake.</p> <p>For example, <a href="https://informatica.snowflakecomputing.com/oauth/token-request">https://informatica.snowflakecomputing.com/oauth/token-request</a></p> <p><b>Note:</b> If the account name contains underscores, use the alias name.</p>
Client ID	Client ID of your application that Snowflake provides during the registration process.
Client Secret	Client secret of your application.
Scope	<p>Specifies access control if the API endpoint has defined custom scopes.</p> <p>Enter space separated scope attributes.</p> <p>For example, specify <code>session:role:CQA_GCP</code> as the scope to override the value of the default user role. The value must be one of the roles assigned in Security Integration.</p>
Access Token Parameters	<p>Additional parameters to use with the access token URL. Define the parameters in the JSON format.</p> <p>For example, define the following parameters:</p> <pre>[{"Name": "code_verifier", "Value": "5PMddu6Zcg6Tc4sbg"}]</pre>

Connection property	Description
Authorization Code Parameters	<p>Additional parameters to use with the authorization token URL. Define the parameters in the JSON format.</p> <p>For example, define the following parameters:</p> <pre>[{"Name": "code_challenge", "Value": "Ikr-vv52th0UeVRi4"}, {"Name": "code_challenge_method", "Value": "S256"}]</pre>
Access Token	<p>Populates the access token value.</p> <p>Enter the access token value or click <b>Generate Access Token</b> to populate the access token value.</p>
Generate Token	Generates the access token and refresh token based on the OAuth attributes you specified.
Refresh Token	<p>Populates the refresh token value.</p> <p>Enter the refresh token value or click <b>Generate Access Token</b> to populate the refresh token value. If the access token is not valid or expires, the Secure Agent fetches a new access token with the help of refresh token.</p> <p><b>Note:</b> If the refresh token expires, provide a valid refresh token or regenerate a new refresh token by clicking <b>Generate Access Token</b>.</p>

## Key pair authentication

The following table describes the Snowflake Data Cloud connection properties for the KeyPair authentication type connection:

Connection property	Description
Runtime Environment	<p>The name of the runtime environment where you want to run the tasks.</p> <p>You cannot run application ingestion tasks and database ingestion tasks on a Hosted Agent or serverless runtime environment.</p>
Authentication	<p>The authentication method to log in to Snowflake.</p> <p>Select <b>KeyPair</b>.</p>
Username	The user name to connect to the Snowflake account.
Account	<p>The name of the Snowflake account.</p> <p>For example, if the Snowflake URL is <a href="https://123abc.us-east-2.aws.snowflakecomputing.com/console/login#/">https://123abc.us-east-2.aws.snowflakecomputing.com/console/login#/</a>, your account name is the first segment in the URL.</p> <p>Here, <i>123abc.us-east-2</i> is your account name.</p> <p>For the Snowsight URL, for example, <a href="https://app.snowflake.com/us-east-2.aws/123abc/dashboard">https://app.snowflake.com/us-east-2.aws/123abc/dashboard</a>, your account name is <i>123abc.us-east-2</i>.</p> <p><b>Note:</b> Ensure that the account name does not contain underscores. To use an alias name, contact Snowflake Customer Support.</p>
Warehouse	The Snowflake warehouse name.

Connection property	Description
Additional JDBC URL Parameters	<p>Optional. The additional JDBC connection parameters.</p> <p>Enter one or more JDBC connection parameters in the following format:</p> <pre>&lt;param1&gt;=&lt;value&gt;&amp;&lt;param2&gt;=&lt;value&gt;&amp;&lt;param3&gt;=&lt;value&gt;...</pre> <p>For example:</p> <pre>user=jon&amp;warehouse=mywh&amp;db=mydb&amp;schema=public</pre> <p><b>Important:</b> Ensure that there is no space before and after = when you add the parameters.</p>
Private Key File	<p>Path to the private key file, including the private key file name, that the Secure Agent uses to access Snowflake.</p> <p><b>Note:</b> Verify that the keystore is FIPS-certified.</p>
Private Key Password	Password for the private key file.

## REST V2 connection properties

When you set up a REST V2 connection, you must configure the connection properties.

The following table describes the REST V2 connection properties for a standard authentication type connection:

Connection property	Description
Runtime Environment	<p>Name of the runtime environment where you want to run the tasks.</p> <p>Specify a Secure Agent, Hosted Agent, or serverless runtime environment.</p> <p>You cannot run a streaming ingestion task on a Hosted Agent or serverless runtime environment.</p>
Authentication Type	If required, select the authentication method that the connector must use to login to the web service application. Default is none.
Auth User ID	The user name to login to the web service application when you select the Basic authentication. Digest authentication is not applicable.
Auth Password	The password associated with the user name when you select the Basic authentication. Digest authentication is not applicable.
OAuth Consumer Key	The client key associated with the web service application. Required only for OAuth authentication type.
OAuth Consumer Secret	The client password to connect to the web service application. Required only for OAuth authentication type.
OAuth Token	The access token to connect to the web service application. Required only for OAuth authentication type.

Connection property	Description
OAuth Token Secret	The password associated with the OAuth token. Required only for OAuth authentication type.
Swagger File Path	<p>The absolute path along with the file name or the hosted URL of the swagger specification file. The hosted URL must return the content of the file without prompting for further authentication and redirection.</p> <p>If you provide the absolute path of the swagger specification file, the swagger specification file must be located on the machine that hosts the Secure Agent. The user must have the read permission for the folder and the specification file. Example:</p> <p><code>C:\swagger\sampleSwagger.json</code></p> <p><b>Note:</b> In a streaming ingestion task, use only a hosted URL of the swagger specification file as the swagger file path.</p>
TrustStore File Path	<p>The absolute path of the truststore file that contains the TLS certificate to establish a one-way or two-way secure connection with the REST API. Specify a directory path that is available on each Secure Agent machine in the runtime environment.</p> <p>You can also configure the truststore file name and password as a JVM option or import the certificate to the following directory:</p> <p><code>&lt;Secure Agent installation directory&gt;\jre\lib\security\cacerts.</code></p> <p>For the serverless runtime environment, specify the truststore file path in the serverless agent directory.</p> <p>For example: <code>/home/cldagnt/SystemAgent/serverless/configurations/ssl_store/&lt;cert_name&gt;.jks</code></p>
TrustStore Password	<p>The password for the truststore file that contains the SSL certificate.</p> <p>You can also configure the truststore password as a JVM option.</p>
KeyStore File Name	<p>The absolute path of the keystore file that contains the keys and certificates required to establish a two-way secure communication with the REST API. Specify a directory path that is available on each Secure Agent machine in the runtime environment.</p> <p>You can also configure the keystore file name and location as a JVM option or import the certificate to any directory.</p> <p>For the serverless runtime environment, specify the keystore file path in the serverless agent directory.</p> <p>For example: <code>/home/cldagnt/SystemAgent/serverless/configurations/ssl_store/&lt;cert_name&gt;.jks</code></p>
KeyStore Password	<p>The password for the keystore file required for secure communication.</p> <p>You can also configure the keystore password as a JVM option.</p>
Proxy Type	<p>Type of proxy. You can select one of the following options:</p> <ul style="list-style-type: none"> <li>- No Proxy. Bypasses the proxy server configured at the agent or the connection level.</li> <li>- Platform Proxy. Proxy configured at the agent level is considered.</li> <li>- Custom Proxy. Proxy configured at the connection level is considered.</li> </ul>

Connection property	Description
Proxy Configuration	The proxy configuration format: <host>:<port> You cannot configure an authenticated proxy server.
Advanced Fields	<p>Enter the arguments that the Secure Agent uses when connecting to a REST endpoint. You can specify the following arguments, each separated by a semicolon (;):</p> <p><code>ConnectionTimeout</code>. The wait time in milliseconds to get a response from a REST endpoint. The connection ends after the connection timeout is over. Default is the timeout defined in the endpoint API.</p> <p><b>Note:</b> If you define both the REST V2 connection timeout and the endpoint API timeout, the connection ends at the shortest defined timeout.</p> <p><code>connectiondelaytime</code>. The delay time in milliseconds to send a request to a REST endpoint. Default is 10000.</p> <p><code>retryattempts</code>. Number of times the connection is attempted when 400 and 500 series error codes are returned in the response. Default is 3. Specify 0 to disable the retry attempts.</p> <p><code>qualifiedSchema</code>. Specifies if the schema selected is qualified or unqualified. Default is false.</p> <p>Example:  <code>connectiondelaytime:10000;retryattempts:5</code></p> <p><b>Note:</b> In a streaming ingestion task, only <code>ConnectionTimeout</code> and <code>retryattempts</code> are applicable.</p>

## Teradata connection properties

When you set up a Teradata connection, you must configure the connection properties.

The following table describes the Teradata connection properties:

Connection property	Description
Connection Name	Name of the connection.
Description	Description of the connection.
Type	The type of connection. Select Teradata.
Runtime Environment	The name of the run-time environment where you want to run the tasks. You cannot use the Hosted Agent for Teradata Connector.
TDPID	The name or IP address of the Teradata database machine.
Tenacity	Amount of time, in hours, that Teradata PT API continues trying to log on when the maximum number of operations runs on the Teradata database. Specify a positive integer. Default is 4.
Database Name	The Teradata database name. If you do not enter a database name, Teradata PT API uses the default login database name.

Connection property	Description
Code Page	<p>Code page associated with the Teradata database.</p> <p>Select one the following code pages:</p> <ul style="list-style-type: none"> <li>- MS Windows Latin 1. Select for ISO 8859-1 Western European data.</li> <li>- UTF-8. Select for Unicode and non-Unicode data.</li> </ul> <p>When you run a task that extracts data from a Teradata source, the code page of the Teradata PT API connection must be the same as the code page of the Teradata source.</p>
Max Sessions	<p>Maximum number of sessions that Teradata PT API establishes with the Teradata database. Specify a positive, non-zero integer. Default is 4.</p>
Min Sessions	<p>Minimum number of Teradata PT API sessions required for the Teradata PT API job to continue.</p> <p>Specify a positive integer between 1 and the Max Sessions value. Default is 1.</p>
Sleep	<p>Amount of time, in minutes, that Teradata PT API pauses before it retries to log on when the maximum number of operations runs on the Teradata database.</p> <p>Specify a positive, non-zero integer. Default is 6.</p>
Data Encryption	<p>Enables full security encryption of SQL requests, responses, and data.</p> <p>Default is disabled.</p>
Block Size	<p>Maximum block size, in bytes, Teradata PT API uses when it returns data to the Secure Agent. Minimum is 256. Maximum is 64,000. Default is 64,000.</p>
Authentication Type	<p>Method to authenticate the user. Select one of the following authentication types:</p> <ul style="list-style-type: none"> <li>- Native. Authenticates your user name and password against the Teradata database specified in the connection.</li> <li>- LDAP. Authenticates user credentials against the external LDAP directory service.</li> <li>- KRB5. Authenticates to the Teradata database through Kerberos.</li> </ul> <p>Default is Native.</p>
Kerberos Artifacts Directory	<p>Directory that contains the Kerberos configuration files named <code>krb5.conf</code> and <code>IICSTPT.keytab</code>.</p> <p>Applicable when you select KRB5 as the authentication type.</p>
Metadata Advanced Connection Properties	<p>The values to set the optional properties of the JDBC driver to fetch the metadata. For example, <code>tmode=ANSI</code>.</p>
Enable Metadata Qualification	<p>Select this option to enable the Teradata connection to read reserved words used as table or column names from the Teradata database.</p> <p>By default, the Enable Metadata Qualification checkbox is not selected and the Secure Agent does not read reserved words from Teradata.</p>
User Name	<p>Database user name with the appropriate read and write database permissions to access the database.</p> <p>If you select KRB5 as the authentication type, you must specify the Kerberos user name.</p>
Password	<p>Password for the database user name.</p> <p>If you select KRB5 as the authentication type, you do not need to specify the Kerberos user password.</p>



## Workday Mass Ingestion connection properties

When you set up a Workday Mass Ingestion connection, you must configure the connection properties.

The properties of a Workday Mass Ingestion connection vary based on the authentication method that you specify for the connection. When you create a connection, you can select one of the following authentication methods:

- **Basic:** Authenticates the connection by validating the login credentials of the Workday account.
- **OAuth 2.0 Refresh Token Flow:** Authenticates the connection by using an application that is registered in Workday. To use this method, you must register an application in Workday and then specify the client ID and client secret of the application in the connection properties. For more information about registering an application in Workday, see the [Workday documentation](#).

### Connection properties for Basic authentication

The following table describes the connection properties for a Workday Mass Ingestion connection configured with Basic authentication:

Connection property	Description
Runtime Environment	Name of the runtime environment where you want to run the ingestion tasks. You must specify a Secure Agent as the runtime environment. <b>Note:</b> You cannot run application ingestion tasks on a Hosted Agent or serverless runtime environment.
Domain Name	Name of the Workday domain that contains the resources that you want to access.
Tenant Name	Identifier of the Workday tenant that you want to access.
Version	Web Service Description Language (WSDL) version for the endpoints that the connection must use to retrieve Workday data. The list of operations supported for a web service depends on the WSDL version that you specify in this field. <b>Note:</b> Informatica recommends that you use WSDL v36.2 because Workday Mass Ingestion connections might not read data from the services that are not part of WSDL v36.2. For more information on the WSDL versions, see the <a href="#">Workday Web Services (WWS) documentation</a> .
User Name	User name of the Workday account.
Password	Password for the Workday account.

**Note:** If you configure a connection with the Basic authentication method and then test the connection, the test is always successful even if the connection property values that you specified are incorrect. Therefore, ensure that you specify correct values for the connection properties before you save the connection.

## Connection properties for OAuth 2.0 Refresh Token Flow authentication

The following table describes the connection properties for a Workday Mass Ingestion connection configured with OAuth 2.0 Refresh Token Flow authentication:

Connection property	Description
Runtime Environment	Name of the runtime environment where you want to run the ingestion tasks. You must specify a Secure Agent as the runtime environment. <b>Note:</b> You cannot run application ingestion tasks on a Hosted Agent or serverless runtime environment.
Domain Name	Name of the Workday domain that contains the resources that you want to access.
Tenant Name	Identifier of the Workday tenant that you want to access.
Version	Web Service Description Language (WSDL) version for the endpoints that the connection must use to retrieve Workday data. The list of operations supported for a web service depends on the WSDL version that you specify in this field. <b>Note:</b> Informatica recommends that you use WSDL v36.2 because Workday Mass Ingestion connections might not read data from the services that are not part of WSDL v36.2. For more information on the WSDL versions, see the <a href="#">Workday Web Services (WWS) documentation</a> .
Client ID	Client ID of the application registered in Workday.
Client Secret	Private key of the application registered in Workday.
Refresh Token	Refresh token string that Workday generates for the registered application.
Token Endpoint	OAuth token endpoint of the Workday instance. The registered application sends the access token requests to this endpoint.

## Zendesk Mass Ingestion connection properties

When you set up a Zendesk Mass Ingestion connection, you must configure the connection properties.

The properties of a Zendesk Mass Ingestion connection vary based on the authentication method that you specify for the connection. When you create a connection, you can select one of the following authentication methods:

- **Basic:** Authenticates the connection by using the login credentials and subdomain associated with the Zendesk account. The Basic authentication method does not use any encrypted access token to connect to the data source, which results in quick and easy access to Zendesk data.  
**Note:** You can use the Basic authentication method only if your Zendesk account is not configured with two-factor authentication. If the account is configured with two-factor authentication, you must use the OAuth 2.0 authentication method for the connection.
- **OAuth 2.0:** Authenticates the connection by using an application that is registered in Zendesk along with the login credentials and subdomain associated with the Zendesk account. To use this method, you must register an application in Zendesk and then specify the client ID and client secret of the application in the connection properties. For more information about registering an application in Zendesk, see the [Zendesk documentation](#).

### Connection properties for Basic authentication

The following table describes the connection properties for a Zendesk Mass Ingestion connection configured with Basic authentication:

Connection property	Description
Runtime Environment	Name of the runtime environment where you want to run the ingestion tasks. You must specify a Secure Agent as the runtime environment. <b>Note:</b> You cannot run application ingestion tasks on a Hosted Agent or serverless runtime environment.
Email ID	User name of the Zendesk account. The user name is an email address.
Password	Password for the Zendesk account.
Subdomain	URL of the Zendesk help center that you want to access.

**Note:** For more information about the Basic authentication method, see the Zendesk documentation.

### Connection properties for OAuth 2.0 authentication

The following table describes the connection properties for a Zendesk Mass Ingestion connection configured with OAuth 2.0 authentication:

Connection property	Description
Runtime Environment	Name of the runtime environment where you want to run the ingestion tasks. You must specify a Secure Agent as the runtime environment. <b>Note:</b> You cannot run application ingestion tasks on a Hosted Agent or serverless runtime environment.
Email ID	User name of the Zendesk account. The user name is an email address.
Password	Password for the Zendesk account.
Subdomain	URL of the Zendesk help center that you want the connection to access.
Client ID	Client ID of the application registered in Zendesk.
Client Secret	Client secret of the application registered in Zendesk.
Grant Type	OAuth 2.0 grant type to be used by the connection. By default, Zendesk Mass Ingestion connections are configured to use the password grant type to exchange user names and passwords for access tokens.

**Note:** For more information about the OAuth 2.0 authentication method, see the Zendesk documentation.

## CHAPTER 4

# Mass Ingestion Applications

Mass Ingestion Applications can transfer data from Software-as-a-Service (SaaS) and on-premise applications to cloud-based data lakes, data warehouses, and event streaming platforms.

The SaaS and on-premise applications used in your business or organization store large amounts of business-critical data on a daily basis. You can use Mass Ingestion Applications to transfer the data stored by your applications to targets that can handle large volumes of data. After you transfer the data to the target, you can consolidate the data and use it for various purposes, such as advanced data analytics and data warehousing.

Mass Ingestion Applications can perform the following types of load operations:

- *Initial load.* Loads source data read at a single point in time to a target. After the data is loaded, the ingestion job ends. You can use this load type to materialize a target to which incremental changes will be sent later.
- *Incremental load.* Loads data changes continuously or until the ingestion job is stopped or ends. The ingestion job loads the changes that have occurred since the last time it ran or from a specific start point. You can use this load type to keep data in the target up to date so that you can make informed decisions for your business or organization based on the latest data.
- *Initial and Incremental load.* Performs an initial load of point-in-time data to the target and then automatically switches to propagating incremental data changes made to the same source objects on a continuous basis.

For more information about the sources and targets supported for each load type, see [“Supported sources” on page 134](#) and [“Supported targets” on page 141](#).

## Use cases

You can use Mass Ingestion Applications to solve multiple business problems.

Following are some of the use cases of Mass Ingestion Applications:

- **Data warehousing:** Organize the data of SaaS and on-premise applications by transferring it to a cloud-based data warehouse system. After an initial batch load of data to the data warehouse, Mass Ingestion Applications can replicate data changes continuously from a source application to keep the data up to date in the data warehouse.
- **Advanced data analytics:** Consolidate your application data in data lakes and data warehouses for extensive analysis that helps in making informed business decisions.
- **Utilizing application data in other data processing applications:** Keep data lakes and data warehouses synchronized with SaaS and on-premise sources and provide up-to-date data to other applications for processing.

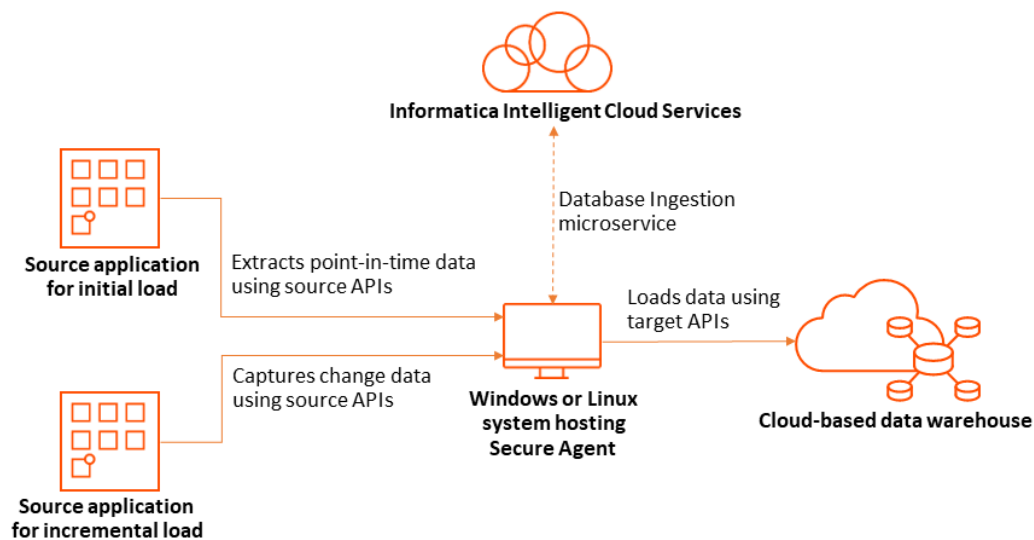
# System requirements

The following table lists the Mass Ingestion Applications minimum system requirements for the Secure Agent:

Component	Minimum requirement
Cores per CPU	8 minimum, 16 recommended if you need to process a large number of source objects in an initial load
Memory	32 GB
Disk space	5 GB per job, based on a row size of 2 KB

## Mass Ingestion Applications architecture

The following diagram shows the Mass Ingestion Applications architecture:



Mass Ingestion Applications requires the following components to run an application ingestion job:

- **Secure Agent:** Secure Agent is the program that runs tasks and enables secure communication between your organization and Informatica Intelligent Cloud Services. You must download the Secure Agent on all the systems from which you want to use Mass Ingestion Applications. When you run an application ingestion job, the metadata of the ingestion task is pushed to the Secure Agent which enables the ingestion job to process the data.
- **Database Ingestion service:** The Database Ingestion service is a microservice that is used by Secure Agent to run both application ingestion and database ingestion tasks. The Database Ingestion service is automatically downloaded to the system when you download the Secure Agent.

- **Informatica Intelligent Cloud Services interface:** Informatica Intelligent Cloud Services provides a web-based interface to create and deploy an application ingestion task. When you deploy the task, an executable ingestion job is created for the task.
- **Source and target APIs:** The application ingestion job uses the source APIs to retrieve the data from the source objects and the target APIs to load the data to the target. For an incremental load operation, the ingestion job identifies the changes that are made to the source objects after a specific date and time and retrieves the changes at intervals defined in the associated ingestion task.

## Supported sources

The sources that Mass Ingestion Applications support depend on whether the application ingestion tasks transfer a point-in-time snapshot of data in a batch initial load operation or load incremental change data from a specific start point.

The following table lists the source types that Mass Ingestion Applications support along with the types of load operations supported for each source type:

Source type	Supported load operations
Adobe Analytics	Initial load, incremental load, and combined initial and incremental load
Google Analytics	Initial load, incremental load, and combined initial and incremental load
Marketo	Initial load, incremental load, and combined initial and incremental load
Microsoft Dynamics 365	Initial load, incremental load, and combined initial and incremental load
NetSuite	Initial load, incremental load, and combined initial and incremental load
Oracle Fusion Cloud Applications	Initial load
Salesforce	Initial load, incremental load, and combined initial and incremental load
SAP ERP Central Component (SAP ECC)	Initial load, incremental load, and combined initial and incremental load
SAP S4/HANA	Initial load, incremental load, and combined initial and incremental load
ServiceNow	Initial load, incremental load, and combined initial and incremental load
Workday	Initial load, incremental load, and combined initial and incremental load
Zendesk	Initial load, incremental load, and combined initial and incremental load

To determine the connectors to use for the source types, see *Connectors and Connections > Mass Ingestion Applications connectors*.

## Guidelines for Marketo sources

Consider the following guidelines when you use Marketo sources:

- The first time an incremental load job runs to capture change data for a Marketo source, the job retrieves and loads only the change records that are created after the date and time specified in the associated ingestion task. You can specify the date and time when you configure the source in the application ingestion task wizard. However, when you resume a stopped or aborted job, the job begins propagating change data from where it last left off.
- Incremental load jobs configured for Marketo sources can replicate only insert and update operations performed on the source objects.
- Mass Ingestion Applications does not replicate the relationships defined between Marketo objects. The target does not contain the relationships that exists between the objects on the source.
- Some Marketo objects do not allow APIs to directly retrieve the data stored in them. To retrieve the data from such objects, APIs require certain filters as input parameters. The filters that the APIs use are fields of other Marketo objects. Mass Ingestion Applications does not replicate the data stored in such Marketo objects that are dependent on other objects.
- Application ingestion jobs do not propagate the data stored in custom objects that are not linked to any lead object. Additionally, application ingestion jobs do not propagate the data stored in custom object fields with the display name ID.
- Mass Ingestion Applications can capture change data only for the following Marketo objects:
  - Program
  - SmartCampaign
  - SmartList
  - StaticList

## Guidelines for Microsoft Dynamics 365 sources

Consider the following guidelines when you use Microsoft Dynamics 365 sources:

- The first time an incremental load job runs to capture change data for a Microsoft Dynamics 365 source, the job retrieves and loads only the change records that are created after the date and time specified in the associated ingestion task. You can specify the date and time when you configure the source in the application ingestion task wizard. However, when you resume a stopped or aborted job, the job begins propagating change data from where it last left off.
- Incremental load jobs for Microsoft Dynamics 365 sources do not capture the change data for objects without the `modifiedon` field.
- Incremental load jobs and combined initial and incremental load job for Microsoft Dynamics 365 sources do not detect and replicate the changes that are made to the source schema.
- If a source object contains locale ID and timezonecode values, Mass Ingestion Applications propagates the values to the target without converting them to the actual values that they represent.
- Mass Ingestion Applications only propagates the fields that are of the primitive data types.
- The fields that are of the Lookup and Customer data types are represented on the target by their unique identifiers.

## Guidelines for NetSuite sources

Consider the following guidelines when you use NetSuite sources:

- The first time an incremental load job runs to capture change data for a NetSuite source, the job retrieves and loads only the change records that are created after the date and time specified in the associated ingestion task. You can specify the date and time when you configure the source in the application ingestion task wizard. However, when you resume a stopped or aborted job, the job begins propagating change data from where it last left off.
- When an application ingestion job retrieves a document or image from the source, the job propagates the unique identifier of the file cabinet that contains the document or image instead of propagating the binary content of the file.
- Mass Ingestion Applications does not capture the deletion of records from the source tables.
- Incremental load jobs configured for NetSuite sources do not capture change data for the tables that do not contain the LAST\_MODIFIED\_DATE column. Additionally, the incremental load jobs do not capture the modification of data stored in the System\_notes and System\_notes\_custom tables.
- Incremental load jobs and combined initial and incremental load jobs configured for NetSuite sources do not detect and replicate the deletion and renaming of source fields.
- Mass Ingestion Applications does not propagate data stored in the following NetSuite tables:
  - Account\_activity
  - Account\_period\_activity
  - Case\_types
  - Case\_origins
  - Deleted\_records

## Guidelines for Oracle Fusion Cloud sources

Consider the following guidelines when you use Oracle Fusion Cloud sources:

- Mass Ingestion Applications can replicate the data of only Enterprise Resource Planning (ERP) and Oracle Supply Chain and Manufacturing (SCM) modules of Oracle Fusion Cloud Applications Suite.
- In the ERP module, Mass Ingestion Applications can replicate the data of the following applications:
  - Oracle Cloud Financials
  - Oracle Procurement
  - Oracle Project Management
  - Oracle Risk Management and Compliance
- In the SCM module, Mass Ingestion Applications can replicate the data of the following applications:
  - Oracle AI Apps
  - Oracle Cloud Service Logistics
  - Oracle Fusion Cloud Inventory Management
  - Oracle Maintenance
  - Oracle Manufacturing
  - Oracle Order Management
  - Oracle Product Lifecycle Management (PLM)
  - Oracle Supply Chain Collaboration



- Oracle Supply Chain Planning

**Note:** Effective in the April 2022 release, support for Oracle Fusion Cloud sources is available for preview.

Preview functionality is supported for evaluation purposes but is unwarranted and is not supported in production environments or any environment that you plan to push to production. Informatica intends to include the preview functionality in an upcoming release for production use, but might choose not to in accordance with changing market or technical circumstances. For more information, contact Informatica Global Customer Support.

## Guidelines for Salesforce sources

Consider the following guidelines when you use Salesforce sources:

- The first time an incremental load job runs to capture change data for a Salesforce source, the job retrieves and loads only the change records that are created after the date and time specified in the associated ingestion task. You can specify the date and time when you configure the source in the application ingestion task wizard. However, when you resume a stopped or aborted job, the job begins propagating change data from where it last left off.
- Mass Ingestion Applications does not capture the relationships defined between Salesforce objects. The target does not contain the relationships between the source objects.
- If an object contains a compound field, Mass Ingestion Applications propagates the component fields that make up the compound field instead of propagating the compound field. For example, if the compound field `geolocation` is a combination of fields `geolocation_longitude` and `geolocation_latitude`, Mass Ingestion Applications separately loads the data stored in the `geolocation_longitude` and `geolocation_latitude` fields to the target instead of propagating the compound field `geolocation`.
- The masking configuration of source fields is retained on the target.
- The lookup fields are represented on the target by their unique alphanumeric identifiers.
- When a new field with a default value of the Function data type is added to an existing source object, Mass Ingestion Applications does not propagate the default value to any existing row on the target. However, when new rows are added to the object, the incremental jobs propagate the values in the field for the newly added rows.

## Guidelines for SAP ECC and SAP S4/HANA sources

Consider the following guidelines when you use SAP ECC or SAP S4/HANA sources:

- When you use an SAP source for the first time, perform the following steps before you configure an application ingestion task for the source:
  1. Verify that the appropriate SAP Notes are available in the SAP server. For more information, see the *SAP Connector Guide*.
  2. Download the SAP Java Connector (SAP JCo) library from the [SAP website](#).
  3. Copy the native and JAR files from the downloaded SAP JCo library to the following directory:  
`<Secure_Agent>\ext\connectors\thirdparty\infa.odp\`
  4. Copy the `sap-adapter-common.jar` file from `<Secure_Agent>\downloads\package-ICSAgent_<version>\package\ICS\main\bin\rdtm\javalib\sap` to the following directory:  
`<Secure_Agent>\ext\connectors\thirdparty\infa.odp\`
  5. Restart the Secure Agent.

- Before you configure an application ingestion task for an SAP source, ensure that you release the datasources from which you want to transfer data. You can use the following functions to release the required datasources:
    - BS\_ANLY\_DS\_RELEASE\_ODP
    - RODPS\_OS\_EXPOSE
  - The first time an incremental load job runs to capture change data for an SAP source, the job retrieves and loads the change records from the latest position in the data stream that contains source data. However, when you resume a stopped or aborted job, the job begins propagating change data from where it last left off.
  - Incremental load jobs for SAP sources do not capture the change data for data sources that do not have a primary key. Additionally, the incremental load jobs do not capture change data for the following SAP datasources:
    - The datasources for which SAP does not support delta.
    - The datasources that are configured to provide change data in the form of additive images.
  - Mass Ingestion Applications does not detect and propagate schema changes in SAP sources.
  - Mass Ingestion Applications does not propagate fields of the following obsolete SAP data types:
    - D16S
    - D34S
    - PREC
    - VARC
- Additionally, application ingestion jobs do not propagate fields of the RSTR data type.

## Guidelines for Workday sources

Consider the following guidelines when you use Workday sources:

- You can use Mass Ingestion Applications to ingest the data of Workday Human Capital Management (HCM). On the **Source** tab of the application ingestion task wizard, you can select the specific HCM services that you want to replicate to your target.
- Mass Ingestion Applications replicates the data of only the source operations whose name start with `Get_`.
- Application ingestion jobs retrieve the source data in an XML structure and then writes the data to the target as a single object in JSON or XML format. When you configure an application ingestion job, on the **Source** tab of the task wizard, you can specify the format of the target data. On the target, each record contains the following fields:
  - WID: Stores the unique identifier or primary key of the record.
  - Data: Stores the content of the record in JSON or XML format.
- Mass Ingestion Applications retains the hierarchical structure of source data on the target.
- Mass Ingestion Applications does not propagate the data stored in custom objects.
- When an application ingestion job ingests an operation under a Workday service to a target, the operation is renamed in the following format: `<Service_Name>__<Operation_Name>`
- The first time an incremental load job runs to capture change data for a Workday source, the job retrieves and loads only the change records that are created after the date and time specified in the application ingestion task. You can specify the date and time when you configure the source in the application ingestion task wizard. However, when you resume a stopped or aborted job, the job begins propagating change data from where it last left off.

- Incremental load jobs can capture only the insert and update operations performed on the source and perform upsert operations on the target to replicate the changes in source data. The application ingestion task wizard does not provide the schema drift options for Workday sources because the upsert operations automatically replicate all changes made to the source schema.
- The following table lists the Workday services and the associated operations for which Mass Ingestion Applications can capture change data:

Service	Operations
Human_Resources	Get_Job_Profiles
	Get_Organizations
	Get_Workers
Recruiting	Get_Evergreen_Requisitions
	Get_Job_Requisitions
	Get_Organizations
	Get_Positions
Staffing	Get_Organizations
	Get_Positions
	Get_Workers

## Guidelines for ServiceNow sources

Consider the following guidelines when you use ServiceNow sources:

- The first time an incremental load job runs to capture change data for a ServiceNow source, the job retrieves and loads only the change records that are created after the date and time specified in the associated ingestion task. You can specify the date and time when you configure the source in the application ingestion task wizard. However, when you resume a stopped or aborted job, the job begins propagating change data from where it last left off.
- The data stored in some ServiceNow objects can be retrieved only if the API client is configured with the maint and nobody roles. Mass Ingestion Applications does not retrieve data from such source objects that require the ServiceNow Mass Ingestion connection to be configured with the maint and nobody roles.
- Incremental load jobs configured for ServiceNow sources do not capture change data for the objects that do not contain the sys\_updated\_on field.
- Incremental load jobs and combined initial and incremental load jobs configured for ServiceNow sources do not detect and replicate the renaming of source fields.
- Mass Ingestion Applications retains the masking configuration of source fields on the target. For example, if a masked field on the source contains the value 1234\*\*\*\*, the corresponding field on the target also contains the value 1234\*\*\*\*.
- Mass Ingestion Applications does not propagate fields of the following ServiceNow data types:
  - Audio
  - Basic Image

- Collection
- Encrypted Text
- FX Currency
- Image
- Journal List
- List
- Name-Value Pairs
- Video
- Wiki

## Guidelines for Zendesk sources

Consider the following guidelines when you use Zendesk sources:

- The first time an incremental load job runs to capture change data for a Zendesk source, the job retrieves and loads only the change records that are created after the date and time specified in the associated ingestion task. You can specify the date and time when you configure the source in the application ingestion task wizard. However, when you resume a stopped or aborted job, the job begins propagating change data from where it last left off.
- Incremental load jobs and combined initial and incremental load job configured for Zendesk sources do not detect and replicate the changes that are made to the source schema.
- If a source record contains multiple custom fields, application ingestion jobs store the data of all the custom fields as a JSON object in a single column of the target table.
- Application ingestion jobs can retrieve data from the custom fields that are present in the following Zendesk objects:
  - Organizations
  - Requests
  - Tickets
  - Users
- Initial load jobs propagate null values for the data stored in the following Zendesk fields:
  - dns results field in the Support Addresses object.
  - raw request and raw response fields in the Target Failures object.
  - reason code field in the Satisfactory Ratings object.
  - URL field in the Sharing Agreements object.
- Initial load jobs do not propagate the data stored in the following Zendesk objects:
  - Attachments
  - Channel Framework
  - Dynamic Content Item Variants
  - End users
  - Incremental Skill-based Routing
  - NPS<sup>®</sup> Invitations
  - NPS<sup>®</sup> Recipients

- NPS® Responses
- OAuth Tokens for Grant Types
- Push Notification Devices
- Search
- Side Conversations
- Side Conversation Attachment
- Side Conversation Events
- Skill-based Routing
- Ticket Comments
- Ticket Import
- User Identities
- User Passwords
- Mass Ingestion Applications can capture change data for the following Zendesk standard objects:
  - NPS Recipients
  - Organizations
  - Side Conversation Events
  - Tickets
  - Tickets Metric Events
  - Users
- Incremental load jobs can capture the deletion of records that are stored in the following objects:
  - Organizations
  - Tickets
  - Users
- Mass Ingestion Applications does not replicate the hierarchical structure of the source fields of the type Object. On the target table, all fields are at the same hierarchical level. When an application ingestion job replicates an Object field with multiple hierarchical levels, the job creates the corresponding columns at the same hierarchical level.

## Supported targets

The targets that Mass Ingestion Applications support depend on the sources specified for the application ingestion tasks.

The following table lists the targets that Mass Ingestion Applications support for each source type:

Source type	Supported target type
Adobe Analytics	Snowflake
Google Analytics	Amazon Redshift, Google BigQuery, Microsoft Azure Synapse Analytics, and Snowflake

Source type	Supported target type
Marketo	Amazon Redshift, Google BigQuery, Microsoft Azure Synapse Analytics, and Snowflake
Microsoft Dynamics 365	Amazon Redshift, Amazon S3, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, and Snowflake
NetSuite	Amazon Redshift, Amazon S3, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, and Snowflake
Oracle Cloud ERP	Snowflake
Oracle SCM	Snowflake
Salesforce	Amazon Redshift, Amazon S3, Apache Kafka, Databricks Delta, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, and Snowflake
SAP ECC	Amazon Redshift, Amazon S3, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, and Snowflake
SAP S4/HANA	Amazon Redshift, Amazon S3, Google BigQuery, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, and Snowflake
ServiceNow	Amazon Redshift, Google BigQuery, Microsoft Azure Synapse Analytics, and Snowflake
Workday	Google BigQuery, Microsoft Azure Synapse Analytics, and Snowflake
Zendesk	Amazon Redshift, Google BigQuery, Microsoft Azure Synapse Analytics, and Snowflake

To determine the connectors to use for the target types, see *Connectors and Connections > Mass Ingestion Applications connectors*.

## Guidelines for Amazon Redshift targets

Consider the following guidelines when you use Amazon Redshift targets:

- Before writing data to Amazon Redshift target tables, application ingestion jobs stage the data in an Amazon S3 bucket. You must specify the name of the bucket when you configure the application ingestion task. The ingestion jobs use the COPY command to load the data from the Amazon S3 bucket to the Amazon Redshift target tables. For more information about the COPY command, see the Amazon Web Services documentation.
- When you define a connection for an Amazon Redshift target, provide the access key and secret access key for the Amazon S3 bucket in which you want the application ingestion jobs to stage the data before loading it to the Amazon Redshift target tables.
- When you ingest data from a source to an Amazon Redshift target, the application ingestion job fails if the data source contains more than 32 data fields or columns that are defined as primary keys.
- Incremental load jobs and combined initial and incremental load jobs generate a recovery table named INFORMATICA\_CDC\_RECOVERY on the target to store internal service information. The data in the recovery table prevents the jobs that are restarted after a failure from propagating previously processed data again. The recovery table is generated in the schema of the target tables.

## Guidelines for Amazon S3, Google Cloud Storage, and Microsoft Azure Data Lake Storage Gen2 targets

Consider the following guidelines when you use Amazon S3, Google Cloud Storage, and Microsoft Azure Data Lake Storage Gen2 targets:

- When you configure an application ingestion task for an Amazon S3, Google Cloud Storage, or Microsoft Azure Data Lake Storage Gen2 target, you can select either CSV or AVRO as the format for the output files that contain the source data to be applied to the target.
- If you select **CSV** as the output file format, Mass Ingestion Applications creates the following files on the target for each source field:
  - schema.ini file that describes the schema of the field. The file also includes some settings for the output file on the target.
  - Output files that contain the data stored in the source field. Mass Ingestion Applications names the output files based on the name of the source field with an appended date and time.

The schema.ini file lists the sequence of columns for the rows in the corresponding output file. The following table describes the columns in the schema.ini file:

Column	Description
ColNameHeader	Indicates whether the source data files include column headers.
Format	Format of the output files. Mass Ingestion Applications uses a comma (,) to delimit column values.
CharacterSet	Character set that is used for the corresponding output file. By default, Mass Ingestion Applications generates the files in the UTF-8 character set.
COL<sequence_number>	Name and data type of the source field.

**Note:**

- You must not edit the schema.ini file.
- If you select the **Add Before Images** check box in the **Advanced** section of the **Target** page, the application ingestion job creates a *column\_name\_OLD* column to store the UNDO data and *column\_name\_NEW* column to store the REDO data for each source field.
- If you select **AVRO** as the output file format, you can specify the serialization format of the Avro output file and write the output data in uncompressed Parquet format. Additionally, you can specify the file compression type, Avro data compression type, and the directory where the Avro schema definitions that are generated for each source object are stored.
- For application ingestion tasks configured for Microsoft Azure Data Lake Storage Gen2 targets, Mass Ingestion Applications creates an empty directory on the target for each empty source field.
- For Amazon S3 targets, if you do not specify an access key and secret key in the connection properties, Mass Ingestion Applications tries to find the AWS credentials by using the default credential provider chain that is implemented by the DefaultAWSCredentialsProviderChain class. For more information, see the Amazon Web Services documentation.

- When an incremental load job that is configured for a target that uses the CSV output format propagates an Update operation that changed primary key values on the source, the job performs a Delete operation on the associated target row and then performs an Insert operation on the same row to replicate the change made to the source object. The Delete operation writes the before image to the target and the subsequent Insert image writes the after image to the target.

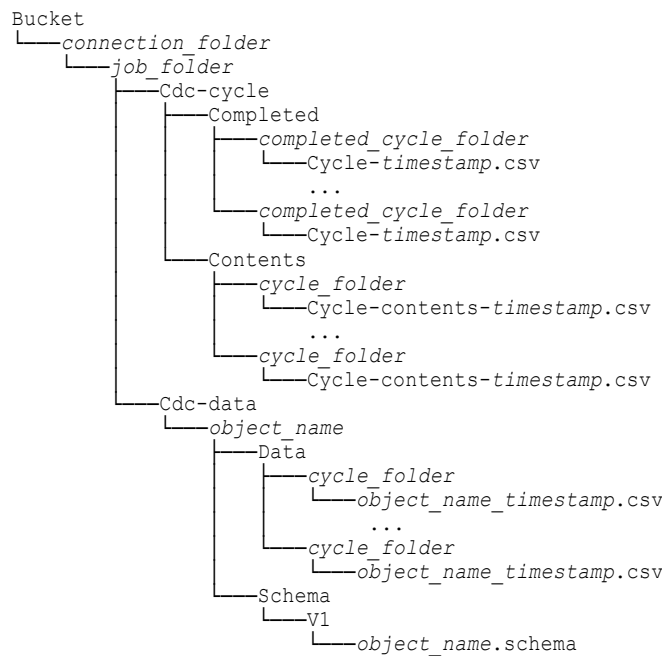
For Update operations that do not change primary key values, application ingestion jobs process each Update operation as a single operation and writes only the after image to the target.

**Note:** If a source object does not contain any primary key, Mass Ingestion Applications considers all fields of the object to be a part of the primary key. In such scenarios, Mass Ingestion Applications processes each Update operation performed on the source as a Delete operation followed by an Insert operation on the target.

## Default directory structure of CDC files on Amazon S3, Google Cloud Storage, and Microsoft Azure Data Lake Storage Gen2 targets

Application ingestion jobs create directories on Amazon S3, Google Cloud Storage, and Microsoft Azure Data Lake Storage Gen2 targets to store information about change data processing.

The following directory structure is created by default on the targets:



The following table describes the directories in the default structure:

Folder	Description
<i>connection_folder</i>	Contains the Mass Ingestion Applications objects. This folder is specified in the <b>Folder Path</b> field of the Amazon S3 connection properties or in the <b>Directory Path</b> field of the Microsoft Azure Data Lake Storage Gen2 connection properties. <b>Note:</b> This folder is not created for Google Cloud Storage targets.
<i>job_folder</i>	Contains job output files. This folder is specified in the <b>Directory</b> field on the <b>Target</b> page of the application ingestion task wizard.



Folder	Description
Cdc-cycle/Completed	Contains a subfolder for each completed CDC cycle. Each cycle subfolder contains a completed cycle file.
Cdc-cycle/Contents	Contains a subfolder for each CDC cycle. Each cycle subfolder contains a cycle contents file.
Cdc-data	Contains output data files and schema files for each object.
Cdc-data/object_name/Schema/V1	Contains a schema file. <b>Note:</b> Mass Ingestion Applications does not save a schema file in this folder if the output files use the Parquet format.
Cdc-data/object_name/Data	Contains a subfolder for each CDC cycle that produces output data files.

## Cycle directories

Mass Ingestion Applications uses the following pattern to name cycle directories:

```
[dt=]yyyy-mm-dd-hh-mm-ss
```

The "dt=" prefix is added to cycle folder names if you select the **Add Directory Tags** check box on the **Target** page of the application ingestion task wizard.

## Cycle contents files

Cycle contents files are located in Cdc-cycle/Contents/cycle\_folder subdirectories. Cycle contents files contain a record for each object that has had a DML event during the cycle. If no DML operations occurred on an object in the cycle, the object does not appear in the cycle contents file.

Mass Ingestion Applications uses the following pattern to name cycle content files:

```
Cycle-contents-timestamp.csv
```

A cycle contents csv file contains the following information:

- Object name
- Cycle name
- Path to the cycle folder for the object
- Start sequence for the object
- End sequence for the object
- Number of Insert operations
- Number of Update operations
- Number of Delete operations
- Schema version
- Path to the schema file for the schema version

**Note:** If the output data files use the Parquet format, Mass Ingestion Applications does not save a schema file at the path that is specified in the cycle contents file. Instead, use the schema file in the folder that is specified in the **Avro Schema Directory** field on the **Target** page of the application ingestion task wizard.

## Completed cycle files

Completed cycle files are located in Cdc-cycle/Completed/*completed\_cycle\_folder* subdirectories. An application ingestion job creates a cycle file in this subdirectory after a cycle completes. If this file is not present, the cycle has not completed yet.

Mass Ingestion Applications uses the following pattern to name completed cycle files:

```
Cycle-timestamp.csv
```

A completed cycle csv file contains the following information:

- Cycle name
- Cycle start time
- Cycle end time
- Current sequence number at the time the cycle ended
- Path to the cycle contents file
- Reason for the end of cycle  
Valid reason values are:
  - **NORMAL\_COMMIT**. A commit operation was encountered after the cycle had reached the DML limit or the end of the cycle interval. A cycle can end only on a commit boundary.
  - **NORMAL\_EXPIRY**. The cycle ended because the cycle interval expired. The last operation was a commit.

## Output data files

The data files contain records that include the following information:

- Operation type. Valid values are:
  - **I** for Insert operations
  - **U** for Update operations
  - **D** for Delete operations
- Sortable sequence number
- Data fields  
**Note:** Insert and Delete records contain only after images. Update records contain both before and after images.

## Custom directory structure of initial load output files on Amazon S3, Google Cloud Storage, and ADLS Gen2 targets

You can configure a custom directory structure for the output data files that initial load jobs write to Amazon S3, Google Cloud Storage, and Microsoft Azure Data Lake Storage (ADLS) Gen2 targets if you do not want to use the default structure.

By default, initial load jobs write output files to *tablename\_timestamp* subdirectories under the parent directory. For all targets except Google Cloud Storage, the parent directory is specified in the target connection properties if the **Connection Directory as Parent** check box is selected on the **Target** page of the task wizard. In an Amazon S3 connection, this parent directory is specified in the **Folder Path** field. In an ADLS Gen2 connection, the parent directory is specified in the **Directory Path** field. For Google Cloud Storage targets, the parent directory is the bucket container specified in the **Bucket** field on the **Target** page of the task wizard.

You can customize the directory structure to suit your needs. For example, you can write the output files under a root directory or directory path that is different from the parent directory specified in the connection

properties to better organize the files for your environment or to find them more easily. Or you can consolidate all output files for an object directly in a directory with the object name rather than write the files to separate timestamped subdirectories, for example, to facilitate automated processing of all of the files.

To configure a directory structure, you must use the **Data Directory** field on the **Target** page of the ingestion task wizard. The default value is `{TableName}_{Timestamp}`, which causes output files to be written to *tablename\_timestamp* subdirectories under the parent directory. You can configure a custom directory path by creating a directory pattern that consists of any combination of case-insensitive placeholders and directory names. The placeholders are:

- `{TableName}` for a target table name
- `{Timestamp}` for the date and time, in the format `yyyymmdd_hhmissms`, at which the initial load job started to transfer data to the target
- `{Schema}` for the target schema name
- `{YY}` for a two-digit year
- `{YYYY}` for a four-digit year
- `{MM}` for a two-digit month value
- `{DD}` for a two-digit day in the month

A pattern can also include the following functions:

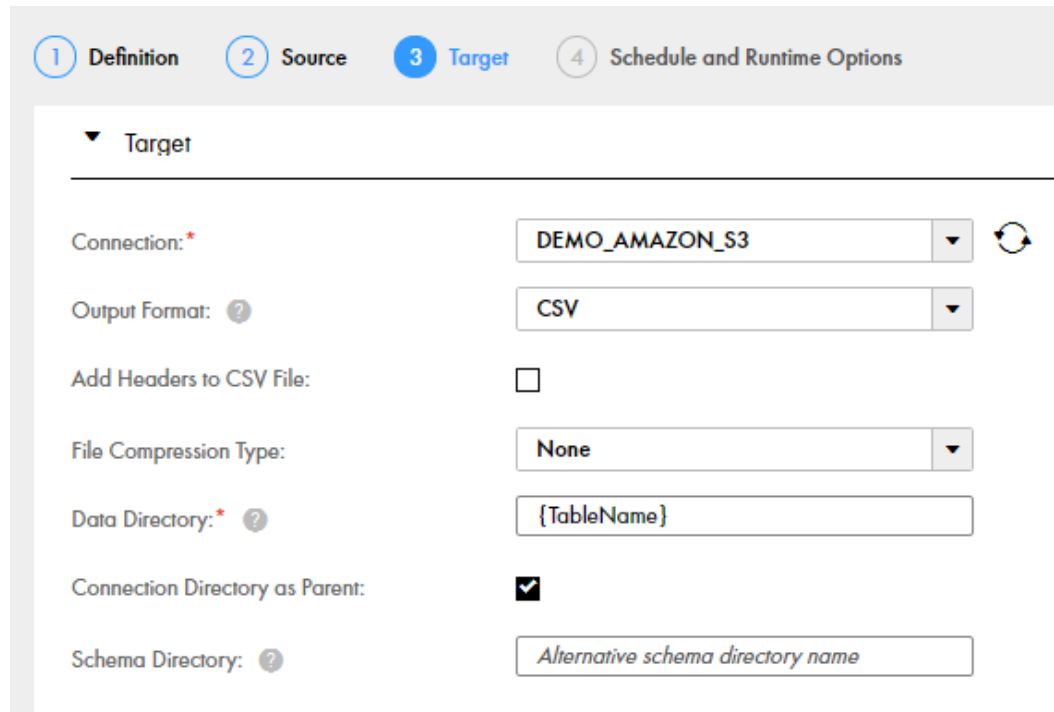
- `toLower()` to use lowercase for the values represented by the placeholder in parentheses
- `toUpper()` to use uppercase for the values represented by the placeholder in parentheses

By default, the target schema is also written to the data directory. If you want to use a different directory for the schema, you can define a directory pattern in the **Schema Directory** field.

#### Example 1


You are using an Amazon S3 target and want to write output files and the target schema to the same directory, which is under the parent directory specified in the **Folder Path** field of the connection properties. In this case, the parent directory is `idr-test/DEMO`. You want write all of the output files for an object to a directory that has a name matching the table name, without a timestamp. You must complete the **Data**

**Directory** field and select the **Connection Directory as Parent** check box. The following image shows this configuration on the **Target** page of the task wizard:



1 Definition 2 Source 3 Target 4 Schedule and Runtime Options

▼ Target

Connection: DEMO\_AMAZON\_S3 

Output Format: CSV

Add Headers to CSV File: ☐

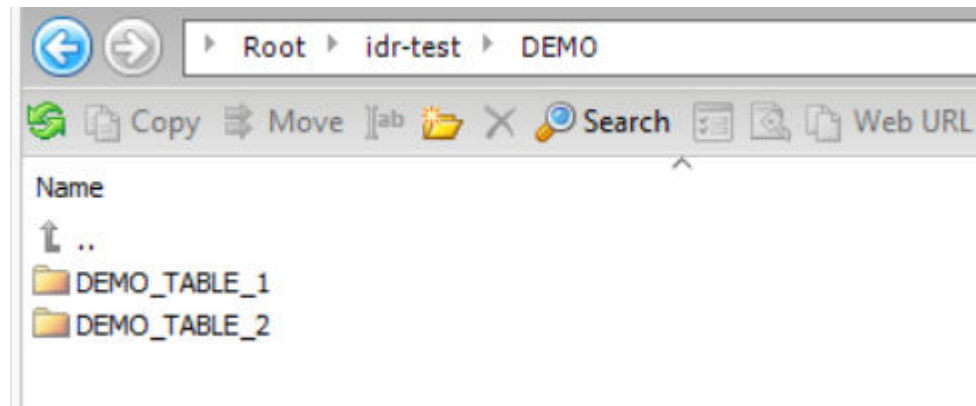
File Compression Type: None

Data Directory: {TableName}

Connection Directory as Parent: ☒

Schema Directory: Alternative schema directory name


Based on this configuration, the resulting directory structure is:



## Example 2

You are using an Amazon S3 target and want to write output data files to a custom directory path and write the target schema to a separate directory path. To use the directory specified in the **Folder Path** field in the Amazon S3 connection properties as the parent directory for the data directory and schema directory, select **Connection Directory as Parent**. In this case, the parent directory is `idr-test/DEMO`. In the **Data Directory** and **Schema Directory** fields, define directory patterns by using a specific directory name, such as `data_dir` and `schema_dir`, followed by the default `{TableName}_{Timestamp}` placeholder value. The placeholder creates

*tablename\_timestamp* destination directories. The following image shows this configuration on the **Target** page of the task wizard:

 DB2\_AMAZON\_S3\_Demo\_Timestamp

1 Definition

2 Source

3 Target

4 Schedule and Runtime Options

▼ Target

Connection: \*

S3\_Demo

?

Output Format: ?

CSV

?

Add Headers to CSV File:

☐

File Compression Type:

None

▼

Data Directory: \* ?

data\_dir/{TableName}\_{Timestamp}

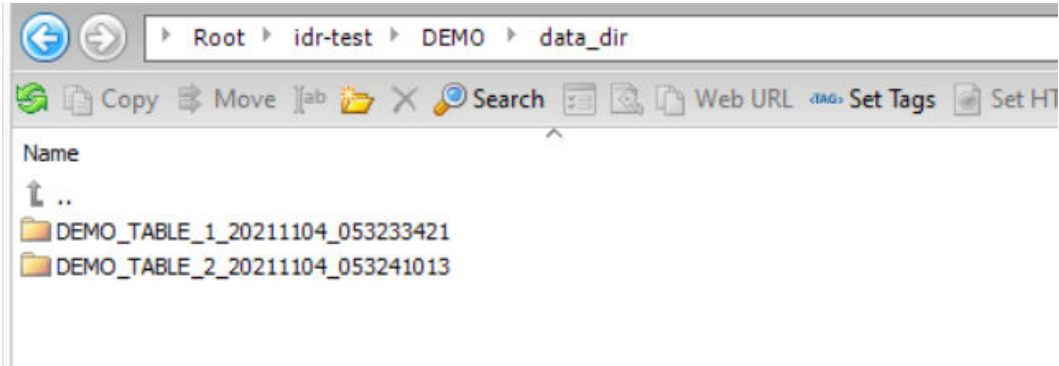
Connection Directory as Parent:

☒

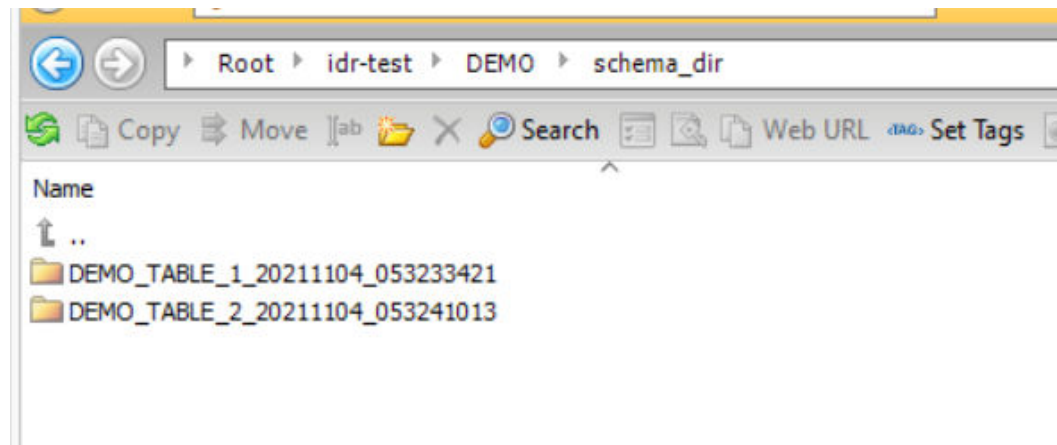
Schema Directory: ?

schema\_dir/{TableName}\_{Timestamp}

Based on this configuration, the resulting data directory structure is:



And the resulting schema directory structure is:



## Guidelines for Databricks Delta targets

Consider the following guidelines when you use Databricks Delta targets:

- When you use a Databricks Delta target for the first time, perform the following steps before you configure an application ingestion task for the target:
  1. Download the Databricks JDBC driver from the Databricks website.
  2. Copy the Databricks JDBC driver jar file, SparkJDBC42.jar, to the following directory:  
*Secure\_Agent\_installation\_directory/apps/Database\_Ingestion/ext/*
  3. On Windows, install Visual C++ Redistributable Packages for Visual Studio 2013 on the computer where the Secure Agent runs.
- For incremental load jobs, you must enable Change Data Capture (CDC) for all source fields.
- You can access Databricks Delta tables created on top of the following storage types:
  - Microsoft Azure Data Lake Storage (ADLS) Gen2
  - Amazon Web Services (AWS) S3

The Databricks Delta connection uses a JDBC URL to connect to the Databricks cluster. When you configure the target, specify the JDBC URL and credentials to use for connecting to the cluster. Also define the connection information that the target uses to connect to the staging location in Amazon S3 or ADLS Gen2.

- Before writing data to Databricks Delta target tables, application ingestion jobs stage the data in an Amazon S3 bucket or ADLS directory. You must specify the directory for the data when you configure the application ingestion task.

**Note:** Mass Ingestion Applications does not use the **ADLS Staging Filesystem Name** and **S3 Staging Bucket** properties in the Databricks Delta connection properties to determine the directory.
- Mass Ingestion Applications uses jobs that run once to load data from staging files on AWS S3 or ADLS Gen2 to external tables.

By default, Mass Ingestion Applications runs jobs on the cluster that is specified in the Databricks Delta connection properties. If you want to run the jobs on another cluster, set the `dbDeltaUseExistingCluster` custom property to false on the **Target** page in the application ingestion task wizard.
- By default, Mass Ingestion Applications uses the Databricks Delta COPY INTO feature to load data from the staging file to Databricks Delta target tables. You can disable it for all load types by setting the `writerDatabricksUseSqlLoad` custom property to false on the **Target** page in the application ingestion task wizard.

- If you use an AWS cluster, you must specify the **S3 Service Regional Endpoint** value in the Databricks Delta connection properties.

## Guidelines for Google BigQuery targets

Consider the following guidelines when you use Google BigQuery targets:

- When you use a Google BigQuery target for the first time, perform the following steps before you configure an application ingestion task for the target:
  1. Download the Google BigQuery JDBC driver from the [Google Cloud website](#).
  2. Copy the JDBC driver jar files to the following directory:
 

```
<Secure_Agent_installation_directory>/apps/Database_Ingestion/ext/
```
  3. Restart the Secure Agent.
- You must have a service account in your Google account to access Google BigQuery and Google Cloud Storage.
- Ensure that you have the client\_email, project\_id, private\_key, private\_key\_id, client\_id, and region ID values for the service account. You must enter these details when you create a Google BigQuery connection.

**Note:** Specify the private\_key\_id and client\_id values in the **Provide Optional Properties** field of the connection properties. Use the following format:

```
"private_key_id": "<private_key_id_value>", "client_id": "<client_id_value>"
```

- If you want to configure a timeout interval for a Google BigQuery connection, specify the timeout interval property in the **Provide Optional Properties** field of the connection properties. Use the following format:
 

```
"timeout": "<timeout_interval_in_seconds>"
```
- You must have read and write access to the following entities:
  - Google BigQuery datasets that contain the target tables.
  - Google Cloud Storage path where Mass Ingestion Applications creates the staging file.
- Application ingestion jobs configured for Google BigQuery targets do not replicate the modification and renaming of source fields on the target.
- You must have the following permissions to write data to a Google BigQuery table:
  - bigquery.datasets.get
  - bigquery.datasets.getIamPolicy
  - bigquery.models.\*
  - bigquery.routines.\*
  - bigquery.tables.create
  - bigquery.tables.delete
  - bigquery.tables.export
  - bigquery.tables.get
  - bigquery.tables.getData
  - bigquery.tables.list
  - bigquery.tables.update
  - bigquery.tables.updateData
  - bigquery.tables.updateTag
  - resourceManager.projects.get

- `resourcemanager.projects.list`
- `bigquery.jobs.create`

## Guidelines for Microsoft Azure Synapse Analytics targets

Consider the following guidelines when you use Microsoft Azure Synapse Analytics targets:

- To deploy and run an application ingestion task with a Microsoft Azure Synapse Analytics target, the target connection must specify a database user who has the CONTROL permission on the target database. To grant the CONTROL permission to the user, use the following SQL statements:

```
USE database_name;
GRANT CONTROL TO user_name;
```

The CONTROL permission is required for initial load, incremental load, and combined initial and incremental load jobs. The permission allows Mass Ingestion Applications to create target tables and database objects such as external data source, external file format, and database scoped credential objects if they do not exist in the database. The CONTROL permission is specifically required for creating external data source and database scoped credential objects.

**Note:** You must manually create the master key. To create the master key, you must have the CONTROL permission on the database.

- Application ingestion jobs first send data to a Microsoft Azure Data Lake Storage Gen2 staging file before writing the data to Microsoft Azure Synapse Analytics target tables. The staging file uses the hexadecimal x1d separator as the field delimiter. After the data is written to the target, the data stored in the table-specific directory that includes the staging files are deleted.
- If you use Microsoft Azure Data Lake Storage Gen2 with a Microsoft Azure Synapse Analytics connection, you must enable the **Hierarchical namespace** option in Microsoft Azure Data Lake Storage. With this setting, blob storage is not recommended.
- When you configure an application ingestion task for a Microsoft Azure Synapse Analytics target, ensure that each source object that you select for replication meets the following criteria:
  - The object must not contain more than 1024 fields and the size of each field must be less than 500 KB.
  - The object must not contain any record that is greater than 1 MB in size.
  - The object must not contain more than 32 primary keys.
  - The primary keys of the object must be of a data type that Microsoft Azure Synapse Analytics supports for primary keys.
- Incremental load jobs and combined initial and incremental load jobs generate a recovery table named `INFORMATICA_CDC_RECOVERY` on the target to store internal service information. The data in the recovery table prevents the jobs that are restarted after a failure from propagating previously processed data again. The recovery table is generated in the schema of the target tables.
- After an application ingestion job loads data to a Microsoft Azure Synapse Analytics target by using external tables, the job does not drop the log tables and external tables created on the target, even though these tables might be re-created when the job starts again.
- Application ingestion jobs configured for Microsoft Azure Synapse Analytics targets do not replicate the renaming of source fields on the target.

## Guidelines for Snowflake targets

Consider the following guidelines when you use Snowflake targets:

- Before writing data to Snowflake target tables, application ingestion jobs write the data to an internal staging area. You must specify the staging directory when you configure the application ingestion task.



- When you define a connection for a Snowflake target, you must set the **Additional JDBC URL Parameters** field to `database=target_database_name`. Otherwise, when you try to define the target in the application ingestion task wizard, an error message indicating that the list of schemas cannot be retrieved appears.
- Incremental load jobs generate a recovery table named `INFORMATICA_CDC_RECOVERY` on the target to store internal service information. The data in the recovery table prevents the jobs that are restarted after a failure from propagating previously processed data again. The recovery table is generated in the schema that contains the target tables.
- Snowflake targets cannot modify the scale of `NUMBER` fields. Additionally, Snowflake targets do not support changing the data type of an existing field to a different data type.

## Avro data types

Mass Ingestion Applications supports only some of the primitive and logical data types that Avro schemas provide.

A primitive data type is a type that allows you to represent a single data value. A logical type is an Avro primitive or complex type with extra attributes to represent a derived type. This topic applies to all target types that support Avro or Parquet output format.

The following table lists the primitive Avro data types that Mass Ingestion Applications supports:

Primitive data type	Description
INT	32-bit signed integer
LONG	64-bit signed integer
FLOAT	Single precision (32-bit) IEEE 754 floating-point number
DOUBLE	Double precision (64-bit) IEEE 754 floating-point number
BYTES	Sequence of 8-bit unsigned bytes
STRING	Unicode character sequence

The following table lists the logical Avro data types that Mass Ingestion Applications supports:

Logical data type	Description
DECIMAL	An arbitrary-precision signed decimal number of the form $\text{unscaled} \times 10^{-\text{scale}}$
DATE	A date, without reference to a time or time zone.
TIME	A time of day that has the precision of 1 millisecond or 1 microsecond, without reference to a time zone or date.
TIMESTAMP	A date and time value that has the precision of 1 millisecond or microsecond, without reference to a particular calendar or time zone.

# Handling source schema changes

You can configure Mass Ingestion Applications to automatically detect source schema changes, also called *schema drift*, and handle these changes on the target. This feature is available only for the incremental load and combined initial and incremental load tasks.

Mass Ingestion Applications can detect the following types of source schema changes:

- Add field
- Modify field
- Drop field
- Rename field

When you configure an application ingestion task, on the **Schedule and Runtime Options** page of the application ingestion task wizard, you can specify the types of source schema changes that Mass Ingestion Applications must propagate for the job associated with the task. You can also specify how the job must handle each type of source schema change. For example, you can configure the task to ignore the changes, replicate them, or stop the job when a particular type of schema change occurs on the source. For more information, see [“Configuring schedule and runtime options” on page 172](#).

Mass Ingestion Applications detects a schema change in a source object only after Data Manipulation Language (DML) operations occur on the altered source object. If multiple schema changes occur without intervening DML operations, Mass Ingestion Applications detects all the schema changes together when a DML operation occurs.

**Note:**

- Application ingestion jobs do not detect the schema changes that occur on the source after you deploy a job and before the first run of the job.
- Application ingestion jobs do not replicate source changes that add, remove, or modify primary key or unique key constraints. If these types of changes occur on the source, you must re-synchronize the target tables.
- Mass Ingestion Applications does not detect the addition and renaming of fields in the schema of Microsoft Dynamics 365 sources.
- The application ingestion jobs configured for Microsoft Azure Synapse Analytics targets do not replicate the renaming of source fields on the target.
- The application ingestion jobs configured for Google BigQuery targets do not replicate the modification and renaming of source fields on the target.

## Configuring application ingestion tasks

Use the application ingestion task wizard in Mass Ingestion to configure application ingestion tasks.

To configure an application ingestion task, perform the following tasks on the wizard:

1. [“Defining basic task information” on page 155](#), such as the task name, project location, runtime environment, and load type.
2. [Configure the source on page 156](#).
3. [Configure the target on page 160](#).
4. [Configure the task schedule and runtime options on page 172](#).

Click **Next** or **Back** to navigate from one page to another. At any point, you can click **Save** to save the information that you have entered until then.

After you complete all the wizard pages, save the information and then click **Deploy** to make the task available as an executable job to the Secure Agent.

## Before you begin

Before you configure an application ingestion task, complete the following prerequisite tasks in Administrator:

- Verify that the Secure Agent in your runtime environment is running and you can access the Mass Ingestion service.
- Define the source and target connections.

## Defining basic task information

To define an application ingestion task, you must first enter some basic information about the task, such as the task name, project or project folder location, and load operation type.

1. In Mass Ingestion, click **New > Application Ingestion Task**.  
The **Definition** page of the application ingestion task wizard appears.
2. Configure the following properties:

Property	Description
Name	<p>Name of the application ingestion task.</p> <p>The name of the application ingestion task must be unique within the organization. The name can contain alphanumeric characters, spaces, periods (.), commas (,), underscores (_), plus signs (+), and hyphens (-).</p> <p>Task names are not case sensitive. The maximum length is 50 characters.</p> <p><b>Note:</b> If you include spaces in the name of an application ingestion task, the spaces do not appear in the name of the job associated with the task.</p>
Location	<p>Project or folder in which you want to store the task.</p>
Runtime Environment	<p>Runtime environment in which you want to run the task.</p> <p>You can run an application ingestion task only on a Secure Agent. The runtime environment can include a Secure Agent Group with only one agent.</p> <p><b>Note:</b> You cannot run application ingestion tasks on a Hosted Agent or serverless runtime environment.</p>

Property	Description
Description	A brief description of the task. Maximum length is 4000 characters.
Load Type	Type of load operation that you want the application ingestion task to perform. You can select one of the following load types for the task: <ul style="list-style-type: none"> <li>- <b>Initial Load:</b> Loads data read at a specific point in time from the source application to the target in a batch operation. You can perform an initial load to materialize a target to which incremental change data will be sent.</li> <li>- <b>Incremental Load:</b> Propagates source data changes to a target continuously or until the job is stopped or ends. The job propagates the changes that have occurred since the last time the job ran or from a specific start point for the first job run.</li> <li>- <b>Initial and Incremental Load:</b> Performs an initial load of point-in-time data to the target and then automatically switches to propagating incremental data changes made to the same source objects on a continuous basis.</li> </ul>

3. Click **Next**.

## Configuring the source

You can configure the source on the **Source** page of the application ingestion task wizard.

Before you configure the source, ensure that the connection to the source is created in Administrator for the runtime environment that your organization uses.

1. From the **Connection** list, select the connection configured for the source application.

The list includes only the connections that are valid for the load type that you selected on the **Definition** page.

**Note:** After you deploy the ingestion task, you cannot change the connection without undeploying the associated ingestion job. After you change the connection, you must deploy the task again.

2. Select the source objects that you want to replicate on the target and specify the advanced and custom properties for the source.

The fields and options that appear on the **Source** tab of the task wizard vary based on the type of source connection that you select. For more information about the steps to complete the configuration of your source, see the following topics:

- [Configuring a Salesforce source.](#)

3. From the **Context** list, select the context containing the source objects that you want to replicate on the target.

**Note:** The **Context** drop-down list appears only for SAP ECC sources.

4. In the **Object Selection Rules** section, create object selection rules to select the source objects that you want to load to the target.

By default, an *Include* rule configured to select all objects on the source is defined in the task. If you do not want to load all the source objects to the target, you can define *Include* rules and *Exclude* rules to select the specific objects that you want to load.

Perform the following steps to create an object selection rule:

- a. Select **Object Selection** as the rule type.
- b. From the adjacent list, select **Include** or **Exclude** as the action that you want the rule to perform.

- c. In the condition field, enter an object name or an object-name mask to specify the source objects to include in or exclude from the list of selected objects.

**Notes:**

- A mask can contain the asterisk (\*) wildcard character to represent one or more characters, the question mark (?) wildcard character to represent a single character, or both types of wildcard characters. You can use a wildcard character multiple times in an object-name mask.
- Object names are case sensitive. When you define the object selection rules, you must specify the object names or masks in the case in which they are defined on the source.
- If an object name contains delimiters, such as quotation marks or brackets, do not include them when you specify the object name for the rule.
- If an object name includes special characters, such as backslash (\), asterisk(\*), dollar sign (\$), caret (^), or question mark (?), replace each special character in the name with a backslash (\) when you specify the object name for the rule.

- d. Click **Add Rule**.

The object selection rules are processed in the order in which they are listed in the **Object Rules** list. The rule at the top of the list is processed first. You can use the arrow icons to change the order in which the rules are listed. For an example of using multiple rules, see [“Example of rules for selecting source objects” on page 159](#).

After you create the rules, you can optionally click **Object Count** to display the number of source objects that match each rule in the **Object Affected** column and the total number of objects selected based on all the selection rules in the **Total Objects** field.

5. In the **Object Selection Rules** section, create data field action rules to perform actions on the data fields of the selected objects.

Perform the following steps to create a data field action rule:

- a. Select **Data Field Action** as the rule type.
- b. From the adjacent list, select one of the following action types:
  - **LTRIM**. Trims spaces to the left of character data field values.
  - **RTRIM**. Trims spaces to the right of character data field values.
  - **TRIM**. Trims spaces to the left of and to the right of character data field values.
- c. In the condition field, enter a data field name or a data field-name mask. The value that you enter is matched against data fields of the selected source objects to identify the data fields to which the action applies.
- d. Click **Add Rule**.

**Note:** The data field action rules are processed in the order in which they are listed in the **Data Field Actions** list. The rule at the top of the list is processed first. You can use the arrow icons to change the order in which the rules are listed.

6. Perform the following steps to download a list of source objects that match the selection rules:

- a. From the **List Objects** list, select the type of selection rule for which you want to download the list of source objects.
- b. If you want to include the fields in the list, select **Include Fields**.
- c. Click the Download icon.

The list of source objects that match the selection rules is downloaded to your local drive.

The information in the downloaded list is in the following format:

*status, object\_name, object\_type, field\_name, comment*

The following table describes the information in the downloaded file:

Field	Description
status	Indicates whether Mass Ingestion Applications includes or excludes the source object from processing. The possible values are: <ul style="list-style-type: none"><li>- <b>E</b>. The object is excluded from processing by an <i>Exclude</i> rule.</li><li>- <b>I</b>. The object is included for processing.</li><li>- <b>X</b>. The object is excluded from processing because its data type is not supported by Mass Ingestion Applications.</li></ul>
object_name	Name of the source object.
object_type	Type of the source object. The possible values are: <ul style="list-style-type: none"><li>- <b>O</b>: Object.</li><li>- <b>F</b>: Field.</li></ul>
field_name	Name of the source field. This information appears only if you selected <b>Include Fields</b> before downloading the list.
comment	Reason why a source object is excluded from processing even though it matches the selection rules.

7. In the **Fetch Size** field of the **Advanced** section, perform one of the following steps based on the source that you want to configure:

- For Salesforce sources, enter the number of records that the application ingestion job associated with the task must read at a time from the source. The default value for initial load operations is 50000 and the default value for incremental load operations is 2000.
- For SAP ECC sources, enter the size of data that the application ingestion job associated with the task must read at a time from the source. The value must be in megabytes (MB). Default value is 8.

**Note:** For Microsoft Dynamics 365 sources, application ingestion jobs read 5000 records at a time from the source irrespective of the value that you enter in the **Fetch Size** field.

8. For incremental load tasks and combined initial and incremental load tasks, perform the following steps in the **Advanced** section:

- a. In the **Restart Point for Incremental Load** field, specify the point in the source data stream from which the ingestion job associated with the application ingestion task starts extracting source change records.

For incremental load tasks with Microsoft Dynamics 365 or Salesforce sources, you must specify a date and time as the restart point. The ingestion job retrieves the changes that were made after the specified date and time, and then loads them to the target. For Microsoft Dynamics 365 sources, you must specify the date and time in Coordinated Universal Time (UTC) and for Salesforce sources, you must specify it in Greenwich Mean Time (GMT).

For SAP ECC sources and for combined initial and incremental load tasks with Microsoft Dynamics 365 or Salesforce sources, Mass Ingestion Applications retrieves the change records from the latest available position in the data stream.

- b. In the **CDC Interval** field, specify the time interval in which the application ingestion job must run to retrieve the change records for incremental load. Default value is 5 minutes.

For SAP ECC sources, the CDC interval must be less than the data retention period configured in the SAP system for the Operational Delta Queue (ODQ).

9. In the **Custom Properties** section, you can specify custom properties that Informatica provides for special cases. To add a property, click the **Add Property** icon, and then add the property name and value.

The custom properties are configured to address unique environments and special use cases.

**Note:** Specify the custom properties only at the direction of Informatica Global Customer Support.

10. Click **Next**.

## Example of rules for selecting source objects

When you define a source for an application ingestion task, you can define object selection rules to select the source objects that you want to load to the target. The following example demonstrates how you can use selection rules to select the required objects.

### Example

A source has 1,000 objects with different prefixes. You want to select the objects that have the prefix "2021\_SALES" and all objects with other prefixes except "2021\_".

Define the following rules in the order in which they are listed:

- Rule 1: **Rule Action=Include** and **Condition=\*** includes all objects on the source. When only the asterisk (\*) wildcard is specified, all objects on the source are selected.
- Rule 2: **Rule Action=Exclude** and **Condition=2021\_\*** excludes the source objects that have names with the prefix "2021\_".
- Rule 3: **Rule Action=Include** and **Condition=2021\_SALES\*** includes the source objects that have names with the prefix "2021\_SALES".

The following image shows the rules in the **Object Selection Rules** section of the **Source** page:

▼ Object Selection Rules

---

Create Rule: Object Selection ▼ Include ▼ Enter the condition Add Rule

Object Rules ⓘ Object Count | ^ v

Action	Condition	Objects Affected
Include	*	
Exclude	2021_*	
Include	2021_SALES*	

## Configuring the target

You can configure the target on the **Target** page of the application ingestion task wizard.

Before you configure the target, ensure that the connection to the target is created in Administrator for the runtime environment that your organization uses.

1. From the **Connection** list, select the connection configured for the source application.

The list includes only the connections that are valid for the load type that you selected on the **Definition** page.

**Note:** After you deploy the ingestion task, you cannot change the connection without undeploying the associated ingestion job. After you change the connection, you must deploy the task again.

2. Configure the target properties.

For descriptions of the target properties, see the following topics:

- [“Amazon Redshift target properties” on page 161](#)
- [Amazon S3 target properties on page 162](#)
- [Google BigQuery target properties on page 165](#)
- [Google Cloud Storage target properties on page 165](#)
- [Microsoft Azure Data Lake Storage Gen2 target properties on page 168](#)
- [Microsoft Azure Synapse Analytics target properties on page 171](#)
- [Snowflake target properties on page 172](#)

3. If you want to rename the target objects that are associated with the selected source objects, define table renaming rules.

For more information about the table renaming rules, see [“Rules for renaming tables on the target” on page 161](#).

4. If you want to override the default mappings of source data types to target data types, perform the following steps in the **Data Type Rules** section to define data type rules:

- a. In the **Create Rule** fields, enter the source data type for which you want to customize the mapping and then enter the target data type that you want to map to the source data type.

**Important:**

- Mass Ingestion Applications does not support the BYTE and CHAR semantics in data-type mappings rules.
- If a source data type has a default value, you must specify it in your rule.

- b. Click **Add Rule**.

The rule is created and appears in the rules list.

**Note:**

- You can define multiple data type mapping rules for a target. However, you can define only one rule for a source data type.
- After you deploy a task with custom mapping rules, you cannot edit the rules without undeploying the task.

5. In the **Custom Properties** section, you can specify custom properties that Informatica provides for special cases. To add a property, click the **Add Property** icon, and then add the property name and value.

The custom properties are configured to address unique environments and special use cases.

**Note:** Specify the custom properties only at the direction of Informatica Global Customer Support.

6. Click **Next**.



## Rules for renaming tables on the target

When you configure a target with an existing schema, you can optionally define rules for renaming the target tables that correspond to the selected source objects.

To create a rule for renaming tables, perform the following steps in the **Table Renaming Rules** section:

1. In the **Create Rule** fields, enter the name of the source object that you want to rename and then enter the name that you want to assign to the target table corresponding to the object.

**Notes:**

- You can enter only the asterisk (\*) wildcard character to select all source objects that match the selection criteria defined on the **Source** page. Alternatively, you can enter the name of a specific source object or an object-name pattern that includes the asterisk (\*) wildcard character.
- If an object or table name includes special characters such as a backslash (\), asterisk(\*), dot (.), or question mark (?), replace each special character in the name with a backslash (\) when you create the rule.
- If you want to use an table-name pattern with a wildcard character for the target table, you must also use the wildcard character in the name of the corresponding source object.

2. Click **Add Rule**.

The rule is created and appears in the rules list.

You can define multiple table renaming rules. Unless a table matches multiple rules, the order in which the rules are processed does not depend on the order in which they are listed in the **Table Renaming Rules** section. If a table matches multiple rules, the last matching rule determines the name of the table.

To delete a rule, click the Delete icon on the row that contains the rule.

### Example

You want to add the prefix "PROD\_" to the names of target tables that are associated with all selected source objects. In the **Table Renaming Rules** section, enter the following values in the **Create Rule** fields:

- For the source, enter the asterisk (\*) wildcard character to specify all the source objects that match the object selection rules defined on the **Source** page.
- For the target, enter PROD\_\* to add this prefix to the names of all target tables corresponding to the source objects.

## Amazon Redshift target properties

When you define an application ingestion task, you must specify the properties for your Amazon Redshift target on the **Target** page of the task wizard.

The following table describes the Amazon Redshift target properties that appear in **Target** section:

Property	Description
Target Creation	The only available option is <b>Create Target Tables</b> , which generates the target tables based on the source objects.
Schema	Select the target schema in which Mass Ingestion Applications creates the target tables.
Bucket	Specifies the name of the Amazon S3 bucket that stores, organizes, and controls access to the data objects that you load to Amazon Redshift.
Directory	Specifies the subdirectory where Mass Ingestion Applications stores output files for this job.

## Amazon S3 target properties

When you define an application ingestion task, you must specify the properties for your Amazon S3 target on the **Target** page of the task wizard.

The following table describes the Amazon S3 target properties that appear in **Target** section:

Property	Description
Output Format	Select the format of the output file. Options are: <ul style="list-style-type: none"><li>- <b>CSV</b></li><li>- <b>AVRO</b></li><li>- <b>PARQUET</b></li></ul> The default value is <b>CSV</b> . <b>Note:</b> Output files in CSV format use double-quotation marks (") as the delimiter for each field.
Add Headers to CSV File	If <b>CSV</b> is selected as the output format, select this check box to add a header with source column names to the output CSV file.
Parquet Compression Type	If the <b>PARQUET</b> output format is selected, you can select a compression type that is supported by Parquet. Options are: <ul style="list-style-type: none"><li>- <b>None</b></li><li>- <b>Gzip</b></li><li>- <b>Snappy</b></li></ul> The default value is <b>None</b> , which means no compression is used.
Avro Format	If you selected <b>AVRO</b> as the output format, select the format of the Avro schema that will be created for each source table. Options are: <ul style="list-style-type: none"><li>- <b>Avro-Flat</b>. This Avro schema format lists all Avro fields in one record.</li><li>- <b>Avro-Generic</b>. This Avro schema format lists all columns from a source table in a single array of Avro fields.</li><li>- <b>Avro-Nested</b>. This Avro schema format organizes each type of information in a separate record.</li></ul> The default value is <b>Avro-Flat</b> .
Avro Serialization Format	If <b>AVRO</b> is selected as the output format, select the serialization format of the Avro output file. Options are: <ul style="list-style-type: none"><li>- <b>None</b></li><li>- <b>Binary</b></li><li>- <b>JSON</b></li></ul> The default value is <b>Binary</b> .
Avro Schema Directory	If <b>AVRO</b> is selected as the output format, specify the local directory where Mass Ingestion Applications stores Avro schema definitions for each source table. Schema definition files have the following naming pattern: <i>schemaname_tablename.txt</i> <b>Note:</b> If this directory is not specified, no Avro schema definition file is produced.
File Compression Type	Select a file compression type for output files in CSV or AVRO output format. Options are: <ul style="list-style-type: none"><li>- <b>None</b></li><li>- <b>Deflate</b></li><li>- <b>Gzip</b></li><li>- <b>Snappy</b></li></ul> The default value is <b>None</b> , which means no compression is used.

Property	Description
Avro Compression Type	<p>If <b>AVRO</b> is selected as the output format, select an Avro compression type. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Bzip2</b></li> <li>- <b>Deflate</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Deflate Compression Level	<p>If <b>Deflate</b> is selected in the <b>Avro Compression Type</b> field, specify a compression level from 0 to 9. The default value is 0.</p>
Add Directory Tags	<p>For incremental load tasks, select this check box to add the "dt=" prefix to the names of apply cycle directories to be compatible with the naming convention for Hive partitioning. This check box is cleared by default.</p>
Data Directory	<p>For initial load tasks, define a directory structure for the directories where Mass Ingestion Applications stores output data files and optionally stores the schema. In this value, you can include the following types of entries to define directory patterns:</p> <ul style="list-style-type: none"> <li>- The placeholders {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The {Timestamp} values are in the format yyyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the initial load job starts to transfer data to the target.</li> <li>- Specific directory names.</li> <li>- The toUpper() and toLower() functions, which force the values for an associated (<i>placeholder</i>) to uppercase or lowercase.</li> </ul> <p>Examples:</p> <pre>myDir1/{SchemaName}/{TableName} myDir1/myDir2/{SchemaName}/{YYYY}/{MM}/{TableName}_{Timestamp} myDir1/{toLower(SchemaName)}/{TableName}_{Timestamp}</pre> <p>This value is not case sensitive.</p> <p>The default format is {TableName}_{Timestamp}.</p> <p>For Amazon S3 and Microsoft Azure Data Lake Storage Gen2 targets, Mass Ingestion Applications uses the directory specified in the target connection properties as the root for the data directory path. For Google Cloud Storage targets, Mass Ingestion Applications uses the <b>Bucket</b> name that you specify in the target properties for the ingestion task.</p>

Property	Description
Connection Directory as Parent	<p>For initial load and incremental load tasks, select this check box to use the directory value that is specified in the target connection properties as the parent directory for the custom directory paths specified in the task target properties. For initial load tasks, the parent directory is used in the <b>Data Directory</b> and <b>Schema Directory</b>. For incremental load tasks, the parent directory is used in the <b>Data Directory</b>, <b>Schema Directory</b>, <b>Cycle Completion Directory</b>, and <b>Cycle Contents Directory</b>.</p> <p>This check box is selected by default. If you clear it, for initial loads, define the full path to the output files in the <b>Data Directory</b> field. For incremental loads, optionally specify a root directory for the task in the <b>Task Target Directory</b>.</p>
Schema Directory	<p>For initial load and incremental load tasks, you can specify a custom directory in which to store the schema file if you want to store it in a directory other than the default directory. For initial loads, previously used values if available are shown in a drop-down list for your convenience. This field is optional.</p> <p>For initial loads, the schema is stored in the data directory by default. For incremental loads, the default directory for the schema file is <code>{TaskTargetDirectory}/cdc-data/{TableName}/schema</code></p> <p>You can use the same placeholders as for the <b>Data Directory</b> field. Ensure that you enclose placeholders with curly brackets <code>{ }</code>.</p> <p>If you include the <code>toUpper</code> or <code>toLower</code> function, put the placeholder name in parentheses and enclose the both the function and placeholder in curly brackets, for example:  <code>{toLower (SchemaName) }</code></p> <p><b>Note:</b> Schema is written only to output data files in CSV format. Data files in Parquet and Avro formats contain their own embedded schema.</p>

The following table describes the Amazon S3 advanced target properties that appear in **Advanced** section:

Field	Description
Add Operation Type	<p>Select this check box to add a metadata column that includes the source SQL operation type in the output that the job propagates to the target.</p> <p>For incremental loads, the job writes "I" for insert, "U" for update, or "D" for delete. For initial loads, the job always writes "I" for insert.</p> <p>By default, this check box is cleared.</p>
Add Operation Time	<p>Select this check box to add a metadata column that includes the source SQL operation time in the output that the job propagates to the target.</p> <p>For initial loads, the job always writes the current date and time.</p> <p>By default, this check box is cleared.</p>
Add Operation Owner	<p>Select this check box to add a metadata column that includes the owner of the source SQL operation in the output that the job propagates to the target.</p> <p>For initial loads, the job always writes "INFA" as the owner.</p> <p>By default, this check box is cleared.</p>

Field	Description
Add Operation Transaction Id	Select this check box to add a metadata column that includes the source transaction ID in the output that the job propagates to the target for SQL operations. For initial loads, the job always writes "1" as the ID. By default, this check box is cleared.
Add Before Images	Select this check box to include UNDO data in the output that an incremental load job writes to the target. For initial loads, the job writes nulls. By default, this check box is cleared.

## Google BigQuery target properties

When you define an application ingestion task, you must specify the properties for your Google BigQuery target on the **Target** page of the task wizard.

The following table describes the Google BigQuery target properties that appear in **Target** section:

Property	Description
Target Creation	The only available option is <b>Create Target Tables</b> , which generates the target tables based on the source objects.
Schema	The target schema in which Mass Ingestion Applications creates the target tables.
Bucket	Specifies the name of an existing bucket container that stores, organizes, and controls access to the data objects that you load to Google Cloud Storage.
Directory	Specifies the virtual directory for the Google Cloud Storage target objects that contain the data.

## Google Cloud Storage target properties

When you define an application ingestion task, you must specify the properties for your Google Cloud Storage target on the **Target** page of the task wizard.

The following table describes the Google Cloud Storage target properties that appear in **Target** section:

Property	Description
Output Format	Select the format of the output file. Options are: <ul style="list-style-type: none"> <li>- <b>CSV</b></li> <li>- <b>AVRO</b></li> <li>- <b>PARQUET</b></li> </ul> The default value is <b>CSV</b> . <b>Note:</b> Output files in CSV format use double-quotation marks (") as the delimiter for each field.
Add Headers to CSV File	If <b>CSV</b> is selected as the output format, select this check box to add a header with source column names to the output CSV file.

Property	Description
Parquet Compression Type	<p>If the <b>PARQUET</b> output format is selected, you can select a compression type that is supported by Parquet. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Gzip</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Avro Format	<p>If you selected <b>AVRO</b> as the output format, select the format of the Avro schema that will be created for each source table. Options are:</p> <ul style="list-style-type: none"> <li>- <b>Avro-Flat</b>. This Avro schema format lists all Avro fields in one record.</li> <li>- <b>Avro-Generic</b>. This Avro schema format lists all columns from a source table in a single array of Avro fields.</li> <li>- <b>Avro-Nested</b>. This Avro schema format organizes each type of information in a separate record.</li> </ul> <p>The default value is <b>Avro-Flat</b>.</p>
Avro Serialization Format	<p>If <b>AVRO</b> is selected as the output format, select the serialization format of the Avro output file. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Binary</b></li> <li>- <b>JSON</b></li> </ul> <p>The default value is <b>Binary</b>.</p>
Avro Schema Directory	<p>If <b>AVRO</b> is selected as the output format, specify the local directory where Mass Ingestion Applications stores Avro schema definitions for each source table. Schema definition files have the following naming pattern:</p> <p><i>schemaname_tablename.txt</i></p> <p><b>Note:</b> If this directory is not specified, no Avro schema definition file is produced.</p>
File Compression Type	<p>Select a file compression type for output files in CSV or AVRO output format. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Deflate</b></li> <li>- <b>Gzip</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Avro Compression Type	<p>If <b>AVRO</b> is selected as the output format, select an Avro compression type. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Bzip2</b></li> <li>- <b>Deflate</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Deflate Compression Level	<p>If <b>Deflate</b> is selected in the <b>Avro Compression Type</b> field, specify a compression level from 0 to 9. The default value is 0.</p>
Add Directory Tags	<p>For incremental load tasks, select this check box to add the "dt=" prefix to the names of apply cycle directories to be compatible with the naming convention for Hive partitioning. This check box is cleared by default.</p>
Bucket	<p>Specifies the name of an existing bucket container that stores, organizes, and controls access to the data objects that you load to Google Cloud Storage.</p>

Property	Description
Data Directory	<p>For initial load tasks, define a directory structure for the directories where Mass Ingestion Applications stores output data files and optionally stores the schema. In this value, you can include the following types of entries to define directory patterns:</p> <ul style="list-style-type: none"> <li>- The placeholders {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The {Timestamp} values are in the format yyyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the initial load job starts to transfer data to the target.</li> <li>- Specific directory names.</li> <li>- The toUpper() and toLower() functions, which force the values for an associated (<i>placeholder</i>) to uppercase or lowercase.</li> </ul> <p>Examples:</p> <pre>myDir1/{SchemaName}/{TableName} myDir1/myDir2/{SchemaName}/{YYYY}/{MM}/{TableName}_{Timestamp} myDir1/{toLower(SchemaName)}/{TableName}_{Timestamp}</pre> <p>This value is not case sensitive.</p> <p>The default format is {TableName}_{Timestamp}.</p> <p>For Amazon S3 and Microsoft Azure Data Lake Storage Gen2 targets, Mass Ingestion Applications uses the directory specified in the target connection properties as the root for the data directory path. For Google Cloud Storage targets, Mass Ingestion Applications uses the <b>Bucket</b> name that you specify in the target properties for the ingestion task.</p>
Schema Directory	<p>For initial load and incremental load tasks, you can specify a custom directory in which to store the schema file if you want to store it in a directory other than the default directory. For initial loads, previously used values if available are shown in a drop-down list for your convenience. This field is optional.</p> <p>For initial loads, the schema is stored in the data directory by default. For incremental loads, the default directory for the schema file is {TaskTargetDirectory}/cdc-data/{TableName}/schema</p> <p>You can use the same placeholders as for the <b>Data Directory</b> field. Ensure that you enclose placeholders with curly brackets { }.</p> <p>If you include the toUpper or toLower function, put the placeholder name in parentheses and enclose the both the function and placeholder in curly brackets, for example:</p> <pre>{toLower(SchemaName)}</pre> <p><b>Note:</b> Schema is written only to output data files in CSV format. Data files in Parquet and Avro formats contain their own embedded schema.</p>

The following table describes the Google Cloud Storage advanced target properties that appear in **Advanced** section:

Field	Description
Add Operation Type	<p>Select this check box to add a metadata column that includes the source SQL operation type in the output that the job propagates to the target.</p> <p>For incremental loads, the job writes "I" for insert, "U" for update, or "D" for delete. For initial loads, the job always writes "I" for insert.</p> <p>By default, this check box is cleared.</p>
Add Operation Time	<p>Select this check box to add a metadata column that includes the source SQL operation time in the output that the job propagates to the target.</p> <p>For initial loads, the job always writes the current date and time.</p> <p>By default, this check box is cleared.</p>

Field	Description
Add Operation Owner	Select this check box to add a metadata column that includes the owner of the source SQL operation in the output that the job propagates to the target. For initial loads, the job always writes "INFA" as the owner. By default, this check box is cleared.
Add Operation Transaction Id	Select this check box to add a metadata column that includes the source transaction ID in the output that the job propagates to the target for SQL operations. For initial loads, the job always writes "1" as the ID. By default, this check box is cleared.
Add Before Images	Select this check box to include UNDO data in the output that an incremental load job writes to the target. For initial loads, the job writes nulls. By default, this check box is cleared.

## Microsoft Azure Data Lake Storage Gen2 target properties

When you define an application ingestion task, you must specify the properties for your Microsoft Azure Data Lake Storage Gen2 target on the **Target** page of the task wizard.

The following table describes the Microsoft Azure Data Lake Storage Gen2 target properties that appear in **Target** section:

Property	Description
Output Format	Select the format of the output file. Options are: <ul style="list-style-type: none"> <li>- <b>CSV</b></li> <li>- <b>AVRO</b></li> <li>- <b>PARQUET</b></li> </ul> The default value is <b>CSV</b> . <b>Note:</b> Output files in CSV format use double-quotation marks (") as the delimiter for each field.
Add Headers to CSV File	If <b>CSV</b> is selected as the output format, select this check box to add a header with source column names to the output CSV file.
Parquet Compression Type	If the <b>PARQUET</b> output format is selected, you can select a compression type that is supported by Parquet. Options are: <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Gzip</b></li> <li>- <b>Snappy</b></li> </ul> The default value is <b>None</b> , which means no compression is used.
Avro Format	If you selected <b>AVRO</b> as the output format, select the format of the Avro schema that will be created for each source table. Options are: <ul style="list-style-type: none"> <li>- <b>Avro-Flat</b>. This Avro schema format lists all Avro fields in one record.</li> <li>- <b>Avro-Generic</b>. This Avro schema format lists all columns from a source table in a single array of Avro fields.</li> <li>- <b>Avro-Nested</b>. This Avro schema format organizes each type of information in a separate record.</li> </ul> The default value is <b>Avro-Flat</b> .



Property	Description
Avro Serialization Format	<p>If <b>AVRO</b> is selected as the output format, select the serialization format of the Avro output file. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Binary</b></li> <li>- <b>JSON</b></li> </ul> <p>The default value is <b>Binary</b>.</p>
Avro Schema Directory	<p>If <b>AVRO</b> is selected as the output format, specify the local directory where Mass Ingestion Applications stores Avro schema definitions for each source table. Schema definition files have the following naming pattern:</p> <p><i>schemaname_tablename.txt</i></p> <p><b>Note:</b> If this directory is not specified, no Avro schema definition file is produced.</p>
File Compression Type	<p>Select a file compression type for output files in CSV or AVRO output format. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Deflate</b></li> <li>- <b>Gzip</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Avro Compression Type	<p>If <b>AVRO</b> is selected as the output format, select an Avro compression type. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Bzip2</b></li> <li>- <b>Deflate</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Deflate Compression Level	<p>If <b>Deflate</b> is selected in the <b>Avro Compression Type</b> field, specify a compression level from 0 to 9. The default value is 0.</p>
Add Directory Tags	<p>For incremental load tasks, select this check box to add the "dt=" prefix to the names of apply cycle directories to be compatible with the naming convention for Hive partitioning. This check box is cleared by default.</p>

Property	Description
Data Directory	<p>For initial load tasks, define a directory structure for the directories where Mass Ingestion Applications stores output data files and optionally stores the schema. In this value, you can include the following types of entries to define directory patterns:</p> <ul style="list-style-type: none"> <li>- The placeholders {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The (Timestamp) values are in the format <code>yyyymmdd_hhmissms</code>. The generated dates and times in the directory paths indicate when the initial load job starts to transfer data to the target.</li> <li>- Specific directory names.</li> <li>- The <code>toUpper()</code> and <code>toLower()</code> functions, which force the values for an associated (<i>placeholder</i>) to uppercase or lowercase.</li> </ul> <p>Examples:</p> <pre>myDir1/{SchemaName}/{TableName} myDir1/myDir2/{SchemaName}/{YYYY}/{MM}/{TableName}_{Timestamp} myDir1/{toLower(SchemaName)}/{TableName}_{Timestamp}</pre> <p>This value is not case sensitive.</p> <p>The default format is <code>{TableName}_{Timestamp}</code>.</p> <p>For Amazon S3 and Microsoft Azure Data Lake Storage Gen2 targets, Mass Ingestion Applications uses the directory specified in the target connection properties as the root for the data directory path. For Google Cloud Storage targets, Mass Ingestion Applications uses the <b>Bucket</b> name that you specify in the target properties for the ingestion task.</p>
Connection Directory as Parent	<p>For initial load and incremental load tasks, select this check box to use the directory value that is specified in the target connection properties as the parent directory for the custom directory paths specified in the task target properties. For initial load tasks, the parent directory is used in the <b>Data Directory</b> and <b>Schema Directory</b>. For incremental load tasks, the parent directory is used in the <b>Data Directory</b>, <b>Schema Directory</b>, <b>Cycle Completion Directory</b>, and <b>Cycle Contents Directory</b>.</p> <p>This check box is selected by default. If you clear it, for initial loads, define the full path to the output files in the <b>Data Directory</b> field. For incremental loads, optionally specify a root directory for the task in the <b>Task Target Directory</b>.</p>
Schema Directory	<p>For initial load and incremental load tasks, you can specify a custom directory in which to store the schema file if you want to store it in a directory other than the default directory. For initial loads, previously used values if available are shown in a drop-down list for your convenience. This field is optional.</p> <p>For initial loads, the schema is stored in the data directory by default. For incremental loads, the default directory for the schema file is <code>{TaskTargetDirectory}/cdc-data/{TableName}/schema</code></p> <p>You can use the same placeholders as for the <b>Data Directory</b> field. Ensure that you enclose placeholders with curly brackets <code>{ }</code>.</p> <p>If you include the <code>toUpper</code> or <code>toLower</code> function, put the placeholder name in parentheses and enclose the both the function and placeholder in curly brackets, for example:</p> <pre>{toLower(SchemaName)}</pre> <p><b>Note:</b> Schema is written only to output data files in CSV format. Data files in Parquet and Avro formats contain their own embedded schema.</p>

The following table describes the Microsoft Azure Data Lake Storage Gen2 advanced target properties that appear in **Advanced** section:

Field	Description
Add Operation Type	Select this check box to add a metadata column that includes the source SQL operation type in the output that the job propagates to the target. For incremental loads, the job writes "I" for insert, "U" for update, or "D" for delete. For initial loads, the job always writes "I" for insert. By default, this check box is cleared.
Add Operation Time	Select this check box to add a metadata column that includes the source SQL operation time in the output that the job propagates to the target. For initial loads, the job always writes the current date and time. By default, this check box is cleared.
Add Operation Owner	Select this check box to add a metadata column that includes the owner of the source SQL operation in the output that the job propagates to the target. For initial loads, the job always writes "INFA" as the owner. By default, this check box is cleared.
Add Operation Transaction Id	Select this check box to add a metadata column that includes the source transaction ID in the output that the job propagates to the target for SQL operations. For initial loads, the job always writes "1" as the ID. By default, this check box is cleared.
Add Before Images	Select this check box to include UNDO data in the output that an incremental load job writes to the target. For initial loads, the job writes nulls. By default, this check box is cleared.

## Microsoft Azure Synapse Analytics target properties

When you define an application ingestion task, you must specify the properties for your Microsoft Azure Synapse Analytics target on the **Target** page of the task wizard.

The following table describes the Microsoft Azure Synapse Analytics target properties that appear in **Target** section:

Property	Description
Target Creation	The only available option is <b>Create Target Tables</b> , which generates the target tables based on the source objects.
Schema	Select the target schema in which Mass Ingestion Applications creates the target tables. The schema name that is specified in the connection properties is displayed by default. This field is case sensitive. Therefore, ensure that you entered the schema name in the connection properties in the correct case.

## Snowflake target properties

When you define an application ingestion task, you must specify the properties for your Snowflake target on the **Target** page of the task wizard.

The following table describes the Snowflake target properties that appear in **Target** section:

Property	Description
Target Creation	The only available option is <b>Create Target Tables</b> , which generates the target tables based on the source objects.
Schema	The target schema in which Mass Ingestion Applications creates the target tables.
Stage	The name of internal staging area that holds the data read from the source before the data is written to the target tables. The name must not include spaces. If the staging area does not exist, it will be automatically created.
Apply Mode	<p>For incremental load and combined initial and incremental load jobs with Snowflake targets, indicates how source DML changes, including inserts, updates, and deletes, are applied to the target. Options are:</p> <ul style="list-style-type: none"><li>- <b>Standard.</b> Accumulate the changes in a single apply cycle and intelligently merge them into fewer SQL statements before applying them to the target. For example, if an update followed by a delete occurs on the source row, no row is applied to the target. If multiple updates occur on the same column or field, only the last update is applied to the target. If multiple updates occur on different columns or fields, the updates are merged into a single update record before being applied to the target.</li><li>- <b>Soft Delete.</b> Apply source delete operations to the target as soft deletes. A soft delete marks the deleted row as deleted without actually removing it from the database. For example, a delete on the source results in a change record on the target with "D" displayed in the INFA_OPERATION_TYPE column. If an update followed by a delete occurs on the source, two records are written to the target both with "D" displayed in the INFA_OPERATION_TYPE column.</li></ul> <p>Consider using soft deletes if you have a long-running business process that needs the soft-deleted data to finish processing, to restore data after an accidental delete operation, or to track deleted values for audit purposes.</p> <ul style="list-style-type: none"><li>- <b>Audit.</b> For Snowflake targets only, ingest change data into an audit table on the target system by using insert operations, instead of merging and applying the changes to the target database. Consider using audit tables if you want to perform computations or other downstream processing on the data before applying it to the target database or if you want to examine the changes. You can add metadata columns for SQL change operations to the audit table by setting options under the <b>Advanced</b> section.</li></ul> <p>Default is Standard.</p>

## Configuring schedule and runtime options

On the **Schedule and Runtime Options** page in the application ingestion task wizard, you can specify a schedule for running the initial load jobs and configure the runtime options for jobs of all load types.

1. In the **Schema Drift Options** section, specify the schema drift option to use for each type of Data Definition Language (DDL) operation.

**Note:** The **Schema Drift Options** section appears only for incremental load and combined initial and incremental load tasks.

Mass Ingestion Applications supports the following types of DDL operations:

- Add Field

- Modify Field
- Drop Field
- Rename Field

The following table describes the schema drift options that you can specify for the DDL operations:

Option	Description
Ignore	Does not replicate DDL changes that occur on the source schema to the target. For Amazon Redshift, Microsoft Azure Synapse Analytics, and Snowflake targets, this option is the default option for the Drop Field and Rename Field operation types. For Google BigQuery targets, this option is the default option for all the operation types.
Replicate	Allows the application ingestion job to replicate the DDL changes to the target. For Amazon Redshift, Microsoft Azure Synapse Analytics, and Snowflake targets, this option is the default option for the Add Field and Modify Field operation types. For Amazon S3, Google Cloud Storage, and Microsoft Azure Data Lake Storage Gen2, targets, this option is the default option for all types of DDL operations. For Google Big Query targets, this option is the default option for Add Field and Drop Field operation types.  <b>Note:</b> <ul style="list-style-type: none"> <li>- If you try to replicate an unsupported schema change type on the target, the application ingestion jobs associated with the task will fail with an error.</li> <li>- Add Field operations that add a primary-key field are not supported and might cause unpredictable results.</li> <li>- Modify Field operations that change the NULL or NOT NULL constraint of a field are not replicated to the target.</li> </ul>
Stop Job	Stops the application ingestion job.
Stop Object	Stops processing the source object on which the DDL change occurred.  <b>Note:</b> When one or more objects are excluded from replication because of the Stop Object schema drift option, the status of the job changes to <b>Running with Warning</b> . The application ingestion job cannot retrieve the data changes that occurred on the source object after the job stops processing the changes. This action leads to data loss on the target. To avoid data loss, you must re-synchronize the source and target objects that the job stopped processing before you resume the application ingestion job.

- Optionally, in the **Advanced** section, modify the value in the **Number of Rows in Output File** value to specify the maximum number of rows that the application ingestion task writes to an output file on the target.

Valid values are 1 through 100000000 and the default value is 100000 rows.

**Note:** For incremental load and combined initial and incremental load operations, change data is flushed to the target either when the specified number of rows is reached or when the flush latency period expires and the job is not in the middle of processing a transaction. The flush latency period is the time that the job waits for more change data before flushing data to the target. The latency period is set to 10 seconds and cannot be changed.

- If you want the application ingestion job associated with the task to run in specific intervals based on a schedule, select **Run this task based on a schedule** in the **Schedule** section, and then select a predefined schedule for the job.

By default, **Do not run this task based on a schedule** is selected, which configures the job to run only when it is manually triggered.

**Note:** This field is available only for initial load tasks.

You can view and edit the job schedule options in Administrator. If you edit the schedule, the changes are automatically applied to all the jobs that are configured to run based on the schedule. If you change the schedule for a task that is already deployed, the updated schedule is automatically applied to the application ingestion job associated with the task.

If a job is about to be triggered based on its schedule when its previous run is still in progress, Mass Ingestion Applications does not run the job and allows the job run that is already in progress to complete.

4. In the **Custom Properties** section, you can specify custom properties that Informatica provides for special cases. To add a property, click the **Add Property** icon, and then add the property name and value.

The custom properties are configured to address unique environments and special use cases.

**Note:** Specify the custom properties only at the direction of Informatica Global Customer Support.

## Deploying an application ingestion task

After you define an application ingestion task, deploy the task to create an executable job instance on the on-premises system that contains the Secure Agent and the Database Ingestion agent service. You can run an application ingestion job only after you deploy the associated task. When you deploy the task, Mass Ingestion Applications also validates the task definition.

If you undeploy a job and then want to run the job again, you must deploy the task again to create a new job instance. The new job instance name ends with an incremented number in the format *taskname-job\_instance\_number*. The job instance number is incremented each time you deploy the ingestion task by adding 1 to the maximum instance number across all ingestion jobs.

Before you deploy a task that is configured for a Snowflake target, you must drop the existing target tables that do not match the structure of source objects. The existing target tables might not match the structure of source objects because of newly added source fields, dropped source or target fields, or modified field null constraints or data types. When you deploy the task after dropping the existing target tables, a new set of target tables are generated based on the source object selection rules and target table renaming rules.

- To deploy a task, in the application ingestion task wizard, save the task and then click **Deploy**.

If you included spaces in the name of the application ingestion task, the spaces are omitted from the name of the corresponding application ingestion job.

After you successfully deploy a task, an application ingestion job is created and the status of the job is Deployed. You can run the job from the **My Jobs** page in Mass Ingestion or from the **All Jobs** tab on the **Mass Ingestion** page in Monitor.

If the deployment fails, the status of the corresponding application ingestion job is Failed. To diagnose the error, you can download the error log from the **My Jobs** page in Mass Ingestion or from the **All Jobs** tab on the **Mass Ingestion** page in Monitor. To download the error log, on the **Actions** menu for the job, click **Error Log**. After you resolve the issue, deploy the task again from the application ingestion task wizard or from the **Actions** menu for the job.

**Note:**

- If the Secure Agent is restarted while the task is deploying, the job status switches to Failed. Avoid restarting the Secure Agent while tasks are being deployed.
- If a task appears to be hung in the Deploying state, restart the Secure Agent. The associated job instance acquires the status of Failed. You can then deploy it again.

# Running an application ingestion job

You can run the application ingestion jobs that are deployed and are in any state other than Undeployed.

You can run an application ingestion job from the **My Jobs** page in the Mass Ingestion service or from the **All Jobs** tab on the **Mass Ingestion** page in Monitor.

For initial load jobs, you can specify a schedule for running the job when you configure the corresponding application ingestion task.

- On the **Actions** menu for the job that you want to run, click **Run**.

A subtask is started for each source object.

## Notes:

- If the initial load job fails to load data to a target table, the application ingestion job retries the subtask for the object up to three times. The minimum interval between the retries is 60 seconds. If all the initial load retries fail, Mass Ingestion Applications excludes the object from replication.
- If an initial load job detects inconsistencies between field definitions in the source and target objects, the job drops the target table and then re-creates it to be consistent with the source object before loading the source data to the target.
- Initial load jobs may take a long time to complete the ingestion if the source objects contain many records.

# Stopping an application ingestion job

You can stop an application ingestion job of any load type that is in the Up and Running, Running with Warning, or On Hold status.

You can stop the job from the **My Jobs** page in the Mass Ingestion service or from the **All Jobs** tab on the **Mass Ingestion** page in Monitor.

When you stop an incremental load job, Mass Ingestion Applications records an identifier for the position in the change stream where it has stopped the incremental processing. The identifier is stored in a recovery table named INFORMATICA\_CDC\_RECOVERY on the target. If you restart the job, Mass Ingestion Applications uses this identifier to identify the last change record that was loaded to the target and starts loading the changes that were made after that point in the change stream.

For initial load jobs, the job stops only after its subtasks that are already running complete their operation. The subtasks that are not running remain in their current states.

- On the **Actions** menu for the job that you want to stop, select **Stop**.

The status of the job changes to Stopping and then changes to Stopped.

**Tip:** If the job takes too long to stop, you can abort the job.

## Aborting an application ingestion job

You can abort an application ingestion job of any load type that is in the Up and Running, Running with Warning, On Hold, or Stopping status.

You can abort an application ingestion job from the **My Jobs** page in the Mass Ingestion service or from the **All Jobs** tab on the **Mass Ingestion** page in Monitor.

When you abort an incremental load job, Mass Ingestion Applications records an identifier for the position in the change stream where it has stopped the incremental processing. The identifier is stored in a recovery table named `INFORMATICA_CDC_RECOVERY` on the target. If you restart the job, Mass Ingestion Applications uses this identifier to identify the last change record that was loaded to the target and starts loading the changes that were made after that point in the change stream.

For initial load jobs, the subtasks that are already running stop immediately, and then the job stops. The subtasks that are not running remain in their current states.

- ▶ On the **Actions** menu for the job that you want to abort, select **Abort**.

The status of the job changes to Aborting and then changes to Aborted.

For initial load jobs, the status of the subtasks that were running change to Aborted. For incremental load jobs, the status of the subtasks change to Stopped.

## Resuming an application ingestion job

You can resume an application ingestion job that is in the Stopped, Aborted, or Failed status.

You can resume an application ingestion job from the **My Jobs** page in the Mass Ingestion service or from the **All Jobs** tab on the **Mass Ingestion** page in Monitor.

When you resume an initial load job that has multiple subtasks, Mass Ingestion Applications starts only the subtasks that are in the Failed, Stopped, Aborted, or Queued status.

When you resume an incremental load job, Mass Ingestion Applications resumes propagating source data from where it last left off.

- ▶ On the **Actions** menu for the job that you want to run, click **Resume**.

A subtask is started for each source object.

**Note:** The **Resume** option is not available if the job is in Failed state because the task deployment failed.

## Restart and recovery for incremental load jobs

Mass Ingestion Applications can restart the incremental load jobs that stopped because of an error and the jobs that were stopped or aborted by users without any loss of change data.

After the first job run, Mass Ingestion Applications continually records an identifier for the processing position in the change stream as changes are applied to the target. The identifier is stored in a recovery table named `INFORMATICA_CDC_RECOVERY` on the target.

When you resume an incremental load job, the job uses the last position recorded in the recovery table to identify the change records that it must load to the target. This process ensures that all changes are ingested to the target.



# Redeploying an application ingestion job

Some fields in application ingestion tasks are editable even when you do not undeploy the associated application ingestion jobs. If you edit any available field in an application ingestion task without undeploying the associated job, you must redeploy the job so that the changes can take effect.

If the fields that you want to edit are uneditable, you must undeploy the associated job and then edit them. For more information about undeploying a job, see [“Undeploying an application ingestion job” on page 177](#).

When you redeploy a job, Mass Ingestion Applications stops all subtasks for the source objects. After the subtasks are stopped, Mass Ingestion Applications deploys the updated ingestion task, and then starts the subtasks. The subtasks that are started includes the subtasks that were previously stopped and the subtasks that are newly created due to the configuration changes in the task.

**Note:** For incremental load jobs, the redeployment does not change the source objects that were selected for ingestion during the previous deployment. To update the list of objects, you must edit the object selection rules in the associated task, and then redeploy the job.

1. On the **My Jobs** page, navigate to the row for the job that you want to redeploy.
2. On the **Actions** menu for the row, click **Redeploy**.

The job instance starts running with the updated configurations.

# Undeploying an application ingestion job

You can undeploy an application ingestion job that you no longer want to run. You can also undeploy an application ingestion job if you want to edit the associated ingestion task to update a connection or property that cannot be updated without undeploying the job.

After you undeploy a job, you cannot run it again or redeploy it. If you want to run a job that is undeployed, you must deploy the associated task again from the application ingestion task wizard to create a new job instance. For example, if you want to change the target connection for a job, you must undeploy the job, edit the ingestion task to change the connection, deploy the task again, and then run the new job instance.

1. Ensure that the job is not running.
2. On the **My Jobs** page in Mass Ingestion or on the **All Jobs** tab on the **Mass Ingestion** page in Monitor, navigate to the row for the job that you want to undeploy.
3. On the **Actions** menu for the row, click **Undeploy**.

The Undeploy option is not available for jobs that are in Failed status and for jobs that have not been previously deployed.

If Mass Ingestion Applications fails to undeploy the job, the status of the job changes to Failed irrespective of its current status.

## CHAPTER 5

# Mass Ingestion Databases

Mass Ingestion Databases can ingest data at scale from common relational databases and propagate the data to multiple types of targets, including cloud-based targets and targets that can handle big data. Mass Ingestion Databases is a separately licensed feature of the Informatica Intelligent Cloud Services Mass Ingestion service.

The Mass Ingestion service provides an easy-to-use interface for configuring and deploying database ingestion tasks and for running and monitoring ingestion jobs. A job is an executable instance of an ingestion task.

Mass Ingestion Databases can perform the following types of load operations:

- *Initial load.* Loads source data read at a single point in time to a target. After the data is loaded, the ingestion job ends. You can use this load type to materialize a target to which incremental changes will be sent later, to migrate from an on-premises database system to a cloud-based system, or to add data to a data lake or data warehouse.
- *Incremental load.* Propagates data changes continuously or until the job is stopped or ends. The job propagates the changes that have occurred since the last time it ran or from a specific start point. You can use this load type to keep data in separate reporting and analytics systems up to date so that you can make informed decisions for your business or organization based on the latest data. You can also use this load type to feed the latest changes to data warehouses and cloud data lakes for big data processing.
- *Initial and incremental load.* Performs an initial load of point-in-time data to the target and then automatically switches to propagating incremental data changes made to the same source tables on a continuous basis.

For more information about the sources and targets supported for each load type, see "Mass Ingestion Databases source and target types."

A database ingestion task automatically maps source tables and fields to target tables and fields based on name matching. You can define rules to customize the target table names.

## Use cases

Mass Ingestion Databases can be used to solve multiple business problems.

You can use Mass Ingestion Databases in the following scenarios:

- **Offline reporting.** Move user reporting activity from a mission-critical production database system to a separate reporting system to avoid degrading database performance.
- **Data warehousing.** Help build out data warehouses by transferring data from multiple databases, including on-premises databases, to the data warehouse system. After an initial batch load of data to the

data warehouse, Mass Ingestion Databases can propagate data changes continuously from a source database to keep the data in the data warehouse up to date.

- **Real-time fraud detection.** Run real-time fraud detection analytics against a replica database that Mass Ingestion Databases keeps up to date by providing change data continuously. Fraud detection processes can then run against the latest data without degrading the source system.
- **Coordination with big data applications.** Keep data lakes synchronized with on-premises sources in a database management system (DBMS) or provide data to Data Integration for at scale processing.
- **Migration to cloud-based systems.** Migrate data from on-premises database systems to cloud-based systems.

## Supported source and target types

The source and target types that Mass Ingestion Databases supports depend on the load type that the database ingestion tasks use: an initial load point-in-time operation, an incremental load of only the data changes, or an initial load followed by an incremental load.

### Source types

The following table shows the source types that are supported (S) for each load type:

Source Type	Initial Load	Incremental Load	Initial and Incremental Loads
Db2 for i	S	S	S
Db2 for Linux, UNIX, and Windows (LUW)	S	-	-
Db2 for z/OS	S	S	-
Microsoft Azure SQL Database	S	-	-
Microsoft SQL Server, including RDS for SQL Server	S	S	S
MongoDB	S	S	-
MySQL, including RDS for MySQL	S	-	-
Netezza	S	-	-
Oracle, including RDS for Oracle	S	S	S
PostgreSQL, including Amazon Aurora PostgreSQL and RDS for PostgreSQL	S	S	-
SAP HANA	S	S	-
Teradata	S	-	-

To determine the connectors to use for these source types, see *Connectors and Connections > Mass Ingestion Databases connectors*.

## Target types

The following table shows the target types that are supported (S) for each load type:

Target Type	Initial Load	Incremental Load	Initial and Incremental Loads
Amazon Redshift	S	S	S
Amazon S3	S	S	S
Apache Kafka, Confluent Kafka, Amazon Managed Streaming for Apache Kafka (MSK)	-	S	-
Azure Event Hubs enabled for use with Kafka clients	-	S	-
Databricks Delta	S	S	S
Flat file	S	-	-
Google BigQuery	S	S	S
Google Cloud Storage	S	S	S
Microsoft Azure Data Lake Storage Gen2	S	S	S
Microsoft Azure Synapse Analytics	S	S	S
Oracle, including RDS for Oracle	S	S	S
Snowflake	S	S	S

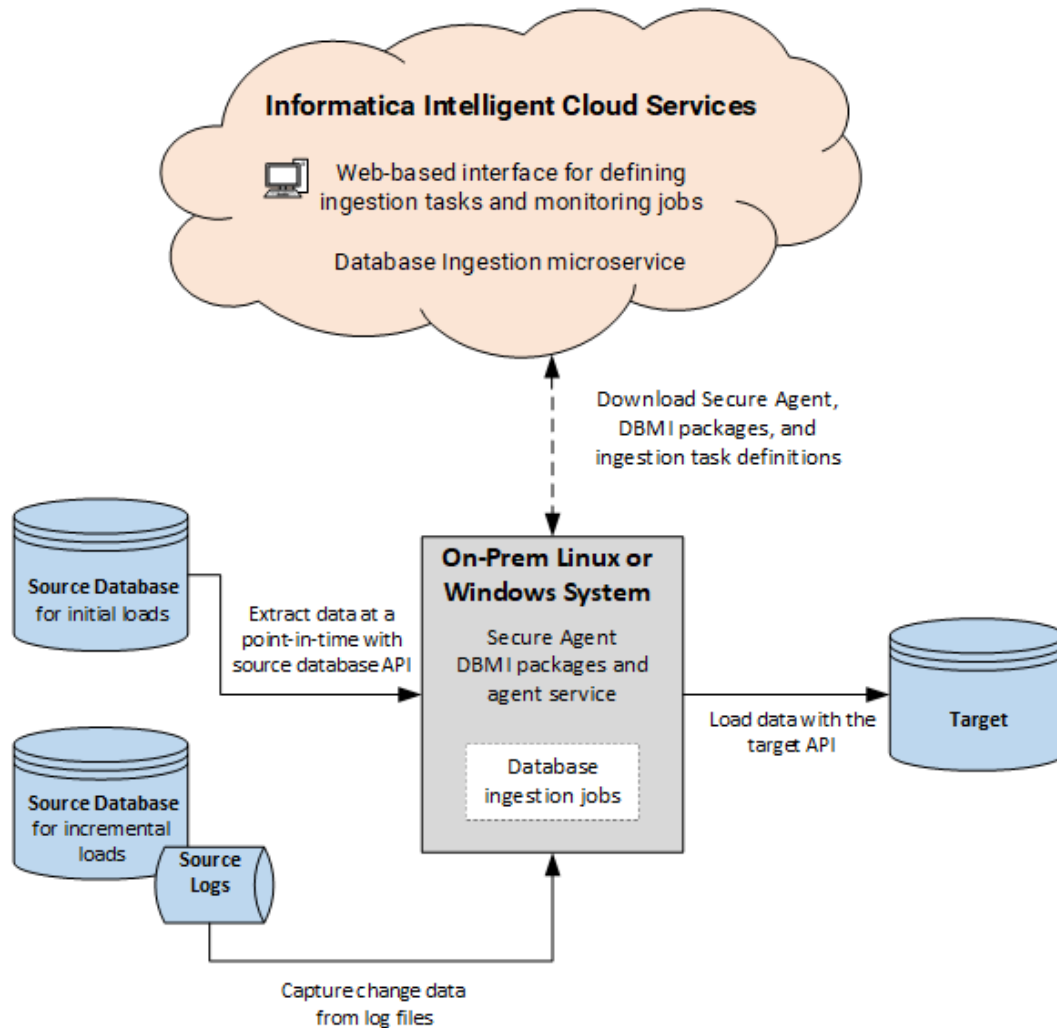
To determine the connectors to use for these target types, see *Connectors and Connections > Mass Ingestion Databases connectors*.

## Mass Ingestion Databases architecture

On each system from which you plan to use Mass Ingestion Databases, you must install the Secure Agent.

After you start the Secure Agent the first time, the Mass Ingestion Databases agent and packages are installed locally. You then can use the Mass Ingestion to configure database ingestion tasks and to run and monitor database ingestion jobs.

The following image shows the general architecture of Mass Ingestion Databases:



From the Web-based interface in Informatica Intelligent Cloud Services, you can create and manage ingestion tasks and run and monitor ingestion jobs.

The following interactions occur:

1. When you download the Secure Agent to your on-premises system, the Database Ingestion DBMI packages are also downloaded, provided that you have a license for Database Ingestion. You can then configure the DBMI agent service.
2. From the Informatica Intelligent Cloud Services Web-based interface, you define database ingestion tasks.
3. When you deploy a data ingestion task, a corresponding executable job is created on the Secure Agent system.
4. When you run a database ingestion job, the ingestion task metadata is pushed down to the Secure Agent. The ingestion job uses this information to process data.
  - For an initial load operation, the ingestion job extracts data at a specific point-in-time from the source tables and fields. The job uses the database API of the relational source to retrieve the data.

- For an incremental load operation, the ingestion job captures changes, such as inserts, updates, and deletes, for the source tables and fields from the source database logs. Change data capture runs continuously or until the job is stopped or ends.

The data is loaded to the target using the appropriate target API.

## Mass Ingestion Databases system requirements

The following table lists the Mass Ingestion Databases minimum system requirements for the Secure Agent:

Component	Minimum requirement
Cores per CPU	8 minimum, 16 recommended if you need to process a large number of source tables in an initial load
Memory	32 GB
Disk space	5 GB per job, based on a row size of 2 KB

## General limitations and guidelines

Before you configure database ingestion tasks for initial, incremental, or combined initial and incremental load operations, review the following limitations and guidelines:

- Mass Ingestion Databases does not support the Hosted Agent.
- Mass Ingestion Databases supports multiple agents in a Secure Agent group.  
In initial load jobs, each subtask of the job is assigned to the agent that has the maximum spare capacity. If the job stops, you can restart it. After the restart, the job will be allocated again to agents with the most spare capacity at the time.  
In incremental load jobs and combined initial and incremental load jobs, only one agent is assigned. The same agent will be used if the job is restarted.
- If a Secure Agent in a runtime environment that is assigned to a database ingestion task terminates, you cannot undeploy the associated ingestion job, update the task to specify another runtime environment, and deploy the task again. In this situation, perform one of the following actions:
  - Assign a different Secure Agent to the runtime environment. Ensure that the new Secure Agent is running. Then restart the associated ingestion job.
  - Copy the task. In the task copy, specify another runtime environment that has an active Secure Agent. Then deploy the task and run the associated ingestion job.
- If you run a Secure Agent service on Windows and plan to use Flat File connections, ensure that the logon account for the Secure Agent is an Administrator account. Otherwise, an error occurs when you try to configure a Flat File connection.
- Mass Ingestion Databases uses UTF-8 as the code page. If you select another code page type when defining a connection, Mass Ingestion Databases ignores it.

- Mass ingestion jobs unload binary data in hexadecimal format when the data will be sent to an Amazon S3, Flat File, or Microsoft Azure Data Lake Storage target. Each hexadecimal column value has the "0x" prefix. If you want to use output files to load the data to a target, you might need to edit the files to remove the "0x" prefixes.
- If a source column has a numeric data type that is not compatible with any numeric data type on the target, Mass Ingestion Databases maps the source column to a target varchar column.
- Mass Ingestion Databases does not display error messages that are longer than 1024 characters in the user interface. Instead, Mass Ingestion Databases prompts you to see a log file with the error, which is automatically downloaded.
- Informatica recommends that source tables have primary keys. If a source table does not have a primary key, Mass Ingestion Databases treats all columns as part of the primary key. In this case, each Update operation is processed as a Delete followed by an Insert on any target. Also, if you change the source primary key and have an Amazon S3, Flat File, Google Cloud Storage, or Microsoft Azure Data Lake Storage target, Mass Ingestion Databases processes each Update operation as a Delete followed by an Insert on the target.

## Mass Ingestion Databases sources - preparation and usage considerations

Before you configure database ingestion tasks for initial load, incremental load, or combined initial and incremental operations, prepare the source database and review any usage considerations for your sources to avoid unexpected results.

### Db2 for i sources

To use Db2 for i sources in database ingestion tasks, first prepare the source database and review the usage considerations.

#### Source preparation

- For incremental load jobs, journaling must be active on each database physical file that corresponds to a selected source table. Also, each journal must be configured with the IMAGES(\*BOTH) option to store both before and after images of change data.

If journaling is not active on a physical file for a source table, when you define a database ingestion task, you can generate a CDC script that activates it. The script issues the following command, which activates journaling and sets the IMAGES option to BOTH:

```
CALL QSYS2.QCMDEXC('STRJRNPF FILE(library/physical-file) JRN(library/journal-name)
IMAGES(*BOTH)')
```

If journaling is already active for a physical file for a source table, the CDC script output contains the following comment:

```
Table 'table_name' is skipped because journaling is already enabled.
```

- Mass Ingestion Databases uses the DataDirect JDBC for IBM Db2 driver to connect to the Db2 for i database. Informatica recommends that the first user who creates and tests a Db2 for i Database Ingestion connection to the source database has DBA authority on the database. This authority is needed for the driver to create and upload the packages that it uses for Db2 access and to grant the EXECUTE privilege on the packages to PUBLIC. If a DBA user does not perform the first connection test, you must grant \*USE authority on the CRTSQLPKG command for creating the packages and grant \*CHANGE authority on the library in which the packages are created.

### Usage considerations

- When you define a database ingestion task, on the **Source** page, specify a journal name that is associated with the source tables that are enabled for journaling.

**Important:** Ensure that the case and spelling of the name in the **Journal Name** field and table selection rules match the journal and table name values in the Db2 source catalog.

- Mass Ingestion Databases does not support the following Db2 for i data types:

- BLOB
- CLOB
- DATALINK
- DBCLOB
- GRAPHIC
- LONG VARGRAPHIC
- VARGRAPHIC
- XML

Database ingestion jobs propagate nulls for columns that have these data types.

For information about the default mappings of supported data types, see [“Default Data Type Mappings” on page 276](#).

## Db2 for LUW sources

To use Db2 for Linux, UNIX, and Windows (LUW) sources in database ingestion tasks, first prepare the source database and review the usage considerations.

### Usage considerations

- Mass Ingestion Databases does not support the following Db2 for LUW data types:

- LONG VARCHAR
- LONG VARGRAPHIC
- LONG VARCHAR FOR BIT DATA
- BLOB
- CLOB
- DBCLOB
- NCLOB
- XML

Database ingestion jobs propagate nulls for columns that have these data types.

For information about the default mappings of supported data types, see [“Default Data Type Mappings” on page 276](#).



## Db2 for z/OS sources

To use Db2 for z/OS sources in database ingestion tasks, first prepare the source database and review the usage considerations.

### Source preparation

- Database ingestion incremental load jobs with a Db2 for z/OS source use a stored procedure to call the Db2 Instrumentation Facility Interface (IFI) to read change data from Db2 logs on the z/OS source system. Mass Ingestion Databases delivers the stored procedure libraries and JCL in the Db2 for z/OS Database Ingestion connector package ZIP file. You must receive the stored procedure libraries into an APF-authorized library and then customize the JCL for your environment. The stored procedure runs in a Workload Manager (WLM) address space on the Db2 source system. For more information about z/OS system requirements, stored procedure setup, and required permissions, see [“Installing and configuring the stored procedure for Db2 for z/OS CDC” on page 186](#).

- When you define a Db2 source in a database ingestion incremental load task, you must select the option **Enable CDC for all columns** in the **CDC Script** field. Mass Ingestion Databases generates a script for enabling CDC on the source tables. You can either execute the CDC script from the user interface if you have been granted the following privilege or ask a Db2 DBA who has SYSDBA authority to run the script:

```
GRANT ALTER TABLE schema.table DATA CAPTURE CHANGES TO user  <--for each source table
COMMIT;
```

- For incremental load operations, ensure that the following privileges and authorities are granted:

**Note:** If you do not have the authority required to issue these grants, ask a Db2 administrator who has SYSDBA authority or a higher authority level to issue them.

Grant the following privileges to the user specified in the Db2 for z/OS Database Ingestion connection properties:

- To make the change data in the Db2 logs available to the stored procedure:

```
GRANT ALTER TABLE schema.table DATA CAPTURE CHANGES TO user  <--for each source
table
COMMIT;
```

- To enable the user to obtain required information from the Db2 Instrumentation Facility Interface (IFI):

```
GRANT MONITOR2 TO user;
COMMIT;
```

- To grant authorities on the global temporary table to the user:

```
GRANT READ ON !SCHEMA!.!STRPRC!_RS_TBL to user;
GRANT DELETE ON !SCHEMA!.!STRPRC!_RS_TBL to user;
GRANT EXECUTE !SCHEMA!.!STRPRC! to user;
COMMIT;
```

Where !SCHEMA! and !STRPRC! are variables in the job JCL, which represent the stored procedure schema name and procedure name respectively.

The first two privileges allow the user to read and delete the contents of the Global Temporary Table to which the stored procedure is writing the log data. The third privilege allows the user to run the stored procedure.

Grant the following privileges to PUBLIC to enable the stored procedure to bind its Db2 plan:

```
GRANT BIND, EXECUTE ON PLAN !STRPRC! TO PUBLIC;
COMMIT;
```

### Usage considerations

- Mass Ingestion Databases does not support schema drift options for Db2 for z/OS sources in incremental load jobs.

- Mass Ingestion Databases does not support the following Db2 for z/OS data types:

- BLOB
- CLOB
- DBCLOB
- XML

Database ingestion jobs propagate nulls for columns that have these data types.

For information about the default mappings of supported data types, see [“Default Data Type Mappings” on page 276](#).

## Installing and configuring the stored procedure for Db2 for z/OS CDC

To perform Db2 for z/OS CDC processing for incremental load jobs, Mass Ingestion Databases provides a stored procedure that runs on the z/OS source system. The stored procedure calls the Db2 Instrumentation Facility Interface (IFI) to collect change data from the Db2 logs.

### z/OS system requirements

Before you begin, verify that the Db2 for z/OS source system meets the following requirements:

- The z/OS system has the recommended operating system version of 2.3 or later.
- The Db2 for z/OS source uses Db2 version 11 or 12.
- The Db2 subsystem has Application Connectivity to Db2 and the following features enabled:
  - Distributed data facility (DDF) access to Db2
  - TCP/IP protocol

If Application Connectivity is not enabled, see the following IBM documentation for more information:

- For Db2 11:
  - [https://www.ibm.com/support/knowledgecenter/SSEPEK\\_11.0.0/java/src/tpc/imjcc\\_jccinstallwithoutdb2z.html](https://www.ibm.com/support/knowledgecenter/SSEPEK_11.0.0/java/src/tpc/imjcc_jccinstallwithoutdb2z.html)
- For Db2 12:
  - [https://www.ibm.com/support/knowledgecenter/SSEPEK\\_12.0.0/java/src/tpc/imjcc\\_jccinstallwithoutdb2z.html](https://www.ibm.com/support/knowledgecenter/SSEPEK_12.0.0/java/src/tpc/imjcc_jccinstallwithoutdb2z.html)
- Ensure that a Workload Manager (WLM) address space exists for executing the Mass Ingestion Databases stored procedure.

If you have not set up a Db2 WLM address space, see the following IBM documentation for more information:

- For Db2 11:
  - [https://www.ibm.com/support/knowledgecenter/en/SSEPEK\\_11.0.0/inst/src/tpc/db2z\\_setupwlmenvironment.html](https://www.ibm.com/support/knowledgecenter/en/SSEPEK_11.0.0/inst/src/tpc/db2z_setupwlmenvironment.html)
- For Db2 12:
  - [https://www.ibm.com/support/knowledgecenter/en/SSEPEK\\_12.0.0/inst/src/tpc/db2z\\_setupwlmenvironment.html](https://www.ibm.com/support/knowledgecenter/en/SSEPEK_12.0.0/inst/src/tpc/db2z_setupwlmenvironment.html)
- Ensure that the library where the stored procedure for Db2 for z/OS CDC will be received exists and is APF-authorized on the z/OS system.

## Install the Stored Procedure Libraries and Customize the JCL

On your client PC, perform the following steps:

1. Verify that the Db2 for z/OS Database Ingestion connector is available.  
The Db2 for z/OS Database Ingestion connector package contains the Db2 for z/OS stored procedure libraries. When the connector is enabled for the Secure Agent Group (runtime environment), the connector package .zip file is downloaded to the downloads folder in the installation location. The package name has the format package-DB2ZMI.*nnnnn*, where *nnnnn* is an incremented package version number. If multiple package versions exist, use the latest one.
2. Unzip the package-DB2ZMI.*nnnnn*.zip file. The stored procedure files are added to the Db2WLMStoredProcedure folder under the package name.
3. Use FTP to transfer the #STPINST file in the Db2WLMStoredProcedure folder to a sequential file, PDS, or PDSE on the z/OS system.

Notes:

- Transfer the file without binary mode set.
  - Add a High Level Qualifier (HLQ) if needed to meet system requirements.
4. Use FTP or another file transfer method to transfer the DBMI.ZOS.DBRMLIB.XMI file to a data set on the z/OS system.

Notes:

- Add or edit the High Level Qualifier (HLQ) if needed to meet system requirements.
  - Transfer the file in binary mode.
  - Ensure that the data set has the DCB attributes of LRECL=80, BLKSIZE=3120, and RECFM=FB.  
You might need to pre-allocate the data set to specify the required attribute values.
5. Use FTP or another file transfer method to transfer the DBMI.ZOS.LOADLIB.XMI file to the z/OS system.

Notes:

- Add or edit the High Level Qualifier (HLQ) if needed to meet system requirements.
- Transfer the file in binary mode.
- Ensure that the data set has the DCB attributes of LRECL=80, BLKSIZE=3120, and RECFM=FB.  
You might need to pre-allocate the data set to specify the required attribute values.

On the z/OS system, use TSO to receive the transmit (XMI) files into an APF-authorized library and then edit the stored procedure JCL member:

1. Receive the DBRMLIB transmit data set:

```
RECEIVE INDATASET(DBMI.ZOS.DBRMLIB.XMI)
```

Notes:

- If you specified a HLQ when transferring the data set to z/OS, include the HLQ.
- When you see message INMR906A Enter restore parameters or 'DELETE' or 'END' +, enter your APF-authorized library:

```
DA(your.library_name) UNIT(unit) VOLUME(volume)
```

The UNIT() and VOLUME() operands are optional. Include them if your installation does not put RECEIVE files on a work unit or volume by default.

2. Receive the LOADLIB transmit data set:

```
RECEIVE INDATASET(DBMI.ZOS.LOADLIB.XMI)
```

Use the same guidelines as in Step 1.

3. Customize the #STPINST file that contains the stored procedure JCL for your environment.

The JCL creates the stored procedure and a global temporary table that holds the results of the requests to the Db2 IFI for data. It also binds the stored procedure package.

Notes:

- Based on the comments at the top of the JCL file, replace variables in the JCL with the values appropriate for your z/OS environment, including the Db2 subsystem name (!DSN!), stored procedure schema name (!SCHEMA!), stored procedure name (!STRPRC!), WLM environment name (!WLMENV!), and the name of the DBRMLIB transmit data set (!DBRMLIB! received in Step 1.
- If you used a HLQ for the received data sets, include the HLQ in the JCL.
- The WLM environment name is specified in the procedure APPLENV parameter or in the EXEC PARM of the WLM address space.  
In the procedure parameter:

```
//STARTING EXEC DSNBWLGM, DB2SSN=DSNB, APPLENV='DSNBWLM_GENERAL'
```

In the EXEC PARM:

```
PARM='DSNB, 40, DSNBWLM_GENERAL'
```

- You can use the received LOADLIB library after it is APF-authorized, or copy the contents of the library to your own APF-authorized library.
- The STEPLIB concatenation in the WLM address space must contain only APF-authorized libraries for the Db2 IFI to run.

## Permissions Required for Executing the Stored Procedure on z/OS

Ensure that the following Db2 permissions are granted before you run the JCL job:

- Ensure that the user who executes the stored procedure job has SYSADM authority, or ask the Db2 for z/OS DBA to run it.
- For the stored procedure to run, you must grant the following Db2 permissions to the procedure schema name specified in the #STPINST JCL file:

- SELECT authority on Db2 catalog tables:

```
GRANT SELECT ON SYSIBM.* TO schema;
```

- EXECUTE authority on the package name specified in the JCL.

```
GRANT EXECUTE ON PACKAGE package_name TO schema;
```

Additionally, grant INSERT and DELETE authority on the schema and stored procedure name that are specified in the JCL for the global temporary table:

```
GRANT INSERT, DELETE ON schema.stored_procedure_name_RS_TBL TO user
```

## Microsoft SQL Server and Azure SQL Database sources

To use Microsoft SQL Server or Microsoft Azure SQL Database sources in database ingestion tasks, first prepare the source database and review the usage considerations.

### Source preparation

- For SQL Server sources, verify that you use Microsoft SQL Server Enterprise Edition, Standard Edition, or Developer Edition. Other SQL Server editions, such as SQL Server Express, are not supported.
- To deploy and run a database ingestion task that has a SQL Server or Azure SQL Database source, the source connection must specify a database user who has the privileges required to perform an initial or incremental load operation.

For initial load operations with a SQL Server or Azure SQL Database source, use the following SQL statements to grant VIEW ANY DEFINITION and SELECT privileges to the user:

```
use [master]
CREATE USER user_name FOR LOGIN login_name;
GRANT VIEW ANY DEFINITION TO [user_name];
GRANT SELECT TO user_name;
GO
```

For incremental load operations with SQL Server sources, the database user that you specify in the SQL Server source connection must have the sysadmin role, or have the db\_owner role and the SELECT permission on the master.sys.fn\_dblog function. Use the following SQL statements if you want the user to have the db\_owner role:

```
use [master]
GRANT SELECT ON master.sys.fn_dblog TO user_name;
GRANT VIEW SERVER STATE TO user_name;
GRANT VIEW ANY DEFINITION TO user_name;
GO
use [source_database]
EXEC sp_addrolemember 'db_owner', 'user_name';
GO
```

**Note:** For SQL Server on-premises instances, the user must have the sysadmin role.

- For database ingestion incremental load and combined initial and incremental load jobs that have SQL Server sources, you must enable SQL Server Change Data Capture (CDC) on the source database by running the sys.sp\_cdc\_enable\_db stored procedure in the database context. To run the procedure, you must have the sysadmin role. When SQL Server CDC is enabled, SQL Server writes additional information to the transaction log and CDC tables, which Mass Ingestion Databases uses during incremental load processing.

Also ensure that SQL Server CDC is enabled on the source tables.

**Restriction:** Mass Ingestion Databases cannot enable CDC for tables that contain more than 1019 columns.

When you create a database ingestion task, you have the option of generating a script that enables CDC on the database and on all columns in the selected source tables.

## Usage considerations

- Mass Ingestion Databases supports Amazon Relational Database Service (RDS) for SQL Server sources for initial, incremental, and combined initial and incremental load jobs.
- Mass Ingestion Databases supports cloud Microsoft Azure SQL Database sources for initial loads. For CDC, you must run the Azure SQL Database source instance on a cloud Azure Managed Instance (AMI).
- Mass Ingestion Databases reads change data for incremental load jobs from the SQL Server transaction log or from the enabled SQL Server CDC tables. The change data is read from the transaction log if the required restart point (LSN) is available there. If the restart point pre-dates the active transaction log, Mass Ingestion Databases automatically transitions to reading the change data from the CDC tables instead. After reading data from the CDC tables, Mass Ingestion Databases switches back to the transaction log in a transparent manner. The same behavior can occur if you resume an incremental load job that stopped for any reason, or if the transaction log is truncated while capture processing is down.
- When you enable CDC on the SQL Server database, SQL Server automatically creates a capture job and a cleanup job that will be executed by the SQL Server Agent. The capture job is responsible for populating the SQL Server CDC tables. The cleanup job is responsible for cleaning up records from the CDC tables. The default value for data retention in the CDC table is 72 hours, or 3 days. You can check the current retention period by running the sys.sp\_cdc\_help\_jobs stored procedure and checking the retention value in the results. If you expect a downtime greater than 3 days, you can adjust the retention in the sys.sp\_cdc\_change\_job stored procedure or in the SQL Server Agent cleanup job. You can also suspend the cleanup job.

- Mass Ingestion Databases supports SQL Server page compression and row compression of source data.
- Informatica recommends that each source table have a primary key. Mass Ingestion Databases does not honor unique indexes in place of a primary key. If no primary key is specified, Mass Ingestion Databases treats all columns as if they are part of the primary key.
- Mass Ingestion Databases requires read-write access on the source database. If you use SQL Server Always On availability groups, this requirement means that Mass Ingestion Databases can capture change data from the read-write primary replica but not from the read-only secondary replica.
- If a Microsoft SQL Server source uses the Always Encrypted method to encrypt column data, the CDC script that is generated from the **CDC Script** field on the **Source** page in the database ingestion task fails to run. This problem is caused by a SQL Server limitation. This problem does not occur with Transparent Data Encryption (TDE).
- Mass Ingestion Databases supports schema drift options for Microsoft SQL Server sources in database ingestion incremental load jobs. The following limitations apply:
  - Microsoft SQL Server does not support renaming tables and columns for which Change Data Capture (CDC) is enabled.
  - Microsoft SQL Server does not support changing primary keys for CDC tables.
- If source table partition changes cause rowset IDs to change, Mass Ingestion Databases can process the changes to enable database ingestion jobs to continue capturing DML changes from the tables.
- If the Mass Ingestion Databases log reader encounters a row for an Insert, Delete, or Update operation on the source that is greater than 8000 bytes in size, change data loss occurs with a warning message.
- Database ingestion initial, incremental, and combined initial and incremental load jobs can replicate data from Microsoft SQL Server large-object columns to Snowflake targets. If an IMAGE or VARBINARY(MAX) column contains more than 8 MB of data, the data is truncated to 8388608 bytes before being written to a BINARY column on the target. If a VARCHAR(MAX), NVARCHAR(MAX), TEXT, NTEXT, or XML column contains more than 16 MB of data, the data is truncated to 16777216 bytes before being written to a VARCHAR column on the target. To replicate data from these large-object columns, you must select **Include LOBs** under **Advanced** on the **Source** page when you configure the task.
- Mass Ingestion Databases does not replicate data from SQL Server computed columns.
- Mass Ingestion Databases does not support the following SQL Server data types:
  - GEOGRAPHY
  - GEOMETRY
  - HIERARCHYID
  - IMAGE

Database ingestion jobs propagate nulls for columns that have these data types.

For information about the default mappings of supported data types, see [“Default Data Type Mappings” on page 276](#).

## MongoDB sources

To use MongoDB sources in database ingestion tasks, review the following considerations.

### Usage considerations

- Mass Ingestion Databases supports MongoDB sources for initial load and incremental load jobs.
- Mass Ingestion Databases supports the following targets for MongoDB sources: Amazon S3, Google Cloud Storage, and Microsoft Azure Data Lake Storage Gen2.

- The database ingestion task moves the MongoDB data to the target as key-value pairs, where the key is the ObjectID and the value is the JSON string, which is a BSON document.
- For MongoDB sources, data type mappings do not occur. All data is persisted on the target as string data.
- In incremental load operations, the change in the data at source is tracked via the unique key (ObjectID) and the same changed JSON string is applied at the target side.
- Mass Ingestion Databases uses MongoDB change streams to access real-time data changes on a single collection, a database, or an entire deployment.
- To open a change stream on a single database, a custom role with privileges that grant `find` and `changeStream` actions is required. Use the following statement to grant the actions on all non-system collections in a database:
 

```
{ resource: { db: <dbname>, collection: "" }, actions: [ "find", "changeStream" ] }
```
- Mass Ingestion Databases does not support time series collections in incremental load jobs that have MongoDB sources.
- In incremental load operations, Mass Ingestion Databases retrieves the change records from the date and time specified as the restart point. For MongoDB sources, the default value for the restart point is the current time. You can change this value and specify a different date and time. You must specify the time in Greenwich Mean Time (GMT).
- If schema drift occurs on the MongoDB source, the data in BSON documents that are sent to the target reflect the schema changes. However, Mass Ingestion Databases does not specifically detect and report the schema changes.

## MySQL sources

To use MySQL sources in database ingestion tasks, first prepare the source database and review the usage considerations.

### Source preparation

- To deploy and run a database ingestion task that includes a MySQL source, the source connection must specify a database user who has the privileges that are required to perform an initial load operation. Use the following SQL statements to grant these privileges to the user:

```
GRANT SELECT ON database_name.* TO 'user_name'@'%';
GRANT SELECT TABLES ON database_name.* TO 'user_name'@'%';
```

- If you use Amazon Relational Database Service (RDS) for MySQL source, you must download the MySQL Native Driver and copy it to the following directory:

```
Secure_Agent_installation_directory/ext/connectors/thirdparty/com.mysql
```

Database ingestion jobs can then use the driver to connect to the RDS for MySQL source.

### Usage considerations

- Mass Ingestion Databases supports RDS for MySQL sources for initial load jobs.
- Mass Ingestion Databases does not support the following MySQL data types:
  - BLOB
  - GEOMETRY
  - GEOMETRYCOLLECTION
  - JSON
  - LINestring
  - LONGBLOB

- LONGTEXT
- MEDIUMBLOB
- MEDIUMTEXT
- MULTILINESTRING
- MULTIPOINT
- MULTIPOLYGON
- POINT
- POLYGON
- TEXT
- TINYBLOB
- TINYTEXT

Database ingestion jobs propagate nulls for columns that have these data types.

For information about the default mappings of supported data types, see [“Default Data Type Mappings” on page 276](#).

## Netezza sources

To use Netezza sources in database ingestion tasks, first prepare the source database and review the usage considerations.

### Source preparation

1. Download and install the Netezza JDBC driver.
  - a. Download the Netezza JDBC driver from the IBM website.
  - b. Copy the Netezza JDBC driver jar file, nzjdbc.jar, to the following directory:
 

```
<Secure Agent installation directory>/apps/Database_Ingestion/ext/
```
  - c. Restart the Secure Agent.
2. To deploy and run a database ingestion task that includes a Netezza source, the source connection must specify a database user who has the privileges that are required to perform an initial load operation. Configure SELECT permissions for the Netezza user account on the following system views:
  - \_V\_JDBC\_SCHEMA1
  - \_V\_JDBC\_SCHEMA3
  - \_V\_ODBC\_TABLES3
  - \_V\_ODBC\_COLUMNS3
  - \_V\_ODBC\_PRIMARYKEYS3

### Usage considerations

- Mass Ingestion Databases does not support the following Netezza data type:
  - ST\_GEOMETRY

Database ingestion jobs either fail to deploy or propagate nulls for columns that have this data type.

For information about the default mappings of supported data types, see [“Default Data Type Mappings” on page 276](#).



## Oracle sources

To use Oracle sources in database ingestion tasks, first prepare the source database and review the usage considerations.

### Source preparation

- Define the ORACLE\_HOME environment variable on the Linux or Windows system where the Secure Agent runs for Mass Ingestion Databases to use the Oracle Call Interface (OCI) to communicate with the Oracle source database.
- Make sure the Mass Ingestion Databases user has the Oracle privileges that are required for the database ingestion load type to be performed. For more information, see [“Oracle privileges” on page 197](#).
- Database ingestion jobs require read access to Oracle online and archive redo logs to read incremental change data. If the redo logs are remote from the on-premises system where the Secure Agent runs, make sure that read access to the logs is provided, for example, by using Oracle Automatic Storage Management (ASM), mounting the logs to a network file system (NFS), or configuring BFILE access to logs that are on the Oracle file system.
- For incremental load or combined initial and incremental load operations, perform the following prerequisite tasks in Oracle:

- Enable ARCHIVELOG mode for the Oracle database. If the database is not in an Amazon RDS environment, issue the following SQL statements:

```
SHUTDOWN IMMEDIATE;  
STARTUP MOUNT;  
ALTER DATABASE ARCHIVELOG;  
ALTER DATABASE OPEN;  
SHUTDOWN IMMEDIATE;  
STARTUP;
```

For an Amazon RDS for Oracle databases, set the backup retention period to place the database in ARCHIVELOG mode and enable automated backups.

- Define an archive log destination.
- Enable Oracle minimal global supplemental logging on the source database.
- If your Oracle source tables have primary keys, ensure that supplemental logging is enabled for all primary key columns. For source tables that do not have primary keys, ensure that supplemental logging is enabled for all columns from which change data will be captured.

**Note:** When you create a database ingestion task, you have the option of generating a script that implements supplemental logging for all columns or only primary key columns for the selected source tables.

- Ensure that the Oracle MAX\_STRING\_SIZE initialization parameter is *not* set to EXTENDED. If it is set to EXTENDED, Mass Ingestion Databases will not be able to replicate inserts and updates for tables containing columns defined with large (extended size) VARCHAR2, NVARCHAR2, or RAW columns.

If you do not have the authority to perform these tasks, ask your Oracle database administrator to perform them. For more information, see the Oracle documentation.

- Ensure that the Oracle Database Client or Instant Client is installed and configured on the Secure Agent server for the Secure Agent to communicate with Oracle. If you do not already have an Oracle client installed, you can download a client and access installation information from the Oracle web site, or ask your Oracle DBA to download and configure an Oracle client.

### Amazon Relational Database Service (RDS) for Oracle source preparation:

1. Create the ONLINELOG\_DIR and ARCHIVELOG\_DIR directories that will hold the online and archive redo logs, respectively, on the RDS file system. Use the following exec statements:

```
exec rdsadmin.rdsadmin_master_util.create_archivelog_dir;  
exec rdsadmin.rdsadmin_master_util.create_onlinelog_dir;
```

2. Grant the Oracle privileges that are required for the Amazon RDS for Oracle source type to the Mass Ingestion Databases user.  
For more information about the privileges required for an Amazon RDS for Oracle source, see [“Oracle log access methods for CDC” on page 201](#).

3. Define an appropriate retention time for the archived redo logs. Use the following exec statement:

```
exec rdsadmin.rdsadmin_util.set_configuration('archivelog retention  
days', number_of_days);
```

4. In the Amazon RDS console, set the backup retention period for the source database to a value greater than zero to enable automated backups of the database instance.

**Note:** This step enables ARCHIVELOG mode for the database.

5. Ensure that supplemental logging is enabled at the database level. Use the following statement:

```
exec rdsadmin.rdsadmin_util.alter_supplemental_logging('ADD');
```

When you create a database ingestion task, you can generate a script to enable supplemental logging for the selected source tables.

6. Optionally, in the Amazon RDS console, you can create a parameter group and define the cache sizes of the default buffer pool. The default buffer pool holds buffers that use the primary database block size. Use the following DB\_CACHE\_SIZE parameter values:

- DB\_2K\_CACHE\_SIZE
- DB\_4K\_CACHE\_SIZE
- DB\_16K\_CACHE\_SIZE
- DB\_32K\_CACHE\_SIZE

Then select the parameter group for the source database.

### Usage considerations

- Informatica recommends that each source table have a primary key. Mass Ingestion Databases does not honor unique indexes in place of a primary key. If no primary key is specified, Mass Ingestion Databases treats all columns as if they are part of the primary key.
- If Oracle source CHAR or VARCHAR columns contain nulls, the database ingestion job does not delimit the null values with double-quotation (") marks or any other delimiter when writing data to a Amazon S3, Flat File, Microsoft Azure Data Lake, or Microsoft Azure Synapse Analytics target.
- Mass Ingestion Databases supports Oracle Data Guard logical and physical standby databases as sources. For implementation details, contact Informatica Global Customer Support.
- Alternative strategies for accessing the Oracle redo logs are available. For more information, see [Oracle log access methods for CDC on page 201](#).
- If a database ingestion incremental load or combined initial and incremental load task contains an Oracle source table name or one or more column names that are longer than 30 characters, Oracle suppresses supplemental logging for the entire table, including primary keys and foreign keys. As a result, most operations on the table fail. This problem is caused by an Oracle restriction. In this situation, exclude the table from capture processing or rename the long table and column names to names of 30 characters or less.
- Database ingestion initial load jobs can replicate data from Oracle BLOB, CLOB, and NCLOB columns to Snowflake targets. If a BLOB column contains more than 8 MB of data, the data is truncated to 8388608

bytes before being written to a BINARY column on the target. If a CLOB or NCLOB column contains more than 16 MB of data, the data is truncated to 16777216 bytes before being written to a VARCHAR column on the target. To replicate data from BLOB, CLOB, or NCLOB columns, you must select **Include LOBs** under **Advanced** on the **Source** page when you configure the task.

- Mass Ingestion Databases does not support the following Oracle source data types with any target type or any load type:
  - INTERVAL
  - LOBs, except for BLOBs, CLOBs, and NCLOBs which are supported for initial load jobs with Snowflake targets
  - LONG
  - LONG RAW
  - TIMESTAMP WITH LOCAL TIME ZONE
  - TIMESTAMP WITH TIME ZONE
  - UROWID
  - XMLTYPE

Source columns that have unsupported data types are excluded from the target definition.

For information about the default mappings of supported data types, see [“Default Data Type Mappings” on page 276](#).

- Mass Ingestion Databases does not support invisible columns in Oracle source columns, regardless of the target type. For these columns, database ingestion incremental load jobs and combined initial and incremental load jobs propagate nulls to the corresponding target columns.
- For Oracle sources that use the multitenant architecture, the source tables must reside in a single pluggable database (PDB) within a multitenant container database (CDB).

## Gathering Information About the Mass Ingestion Databases environment

Before you start creating database ingestion tasks, gather the following information:

### General Information

**What Oracle version do you use?**

Answer: \_\_\_\_\_

**Do you run Oracle on premises or in a cloud-based Amazon RDS for Oracle environment?**

Answer: \_\_\_\_\_

**What is the target type?**

Answer: \_\_\_\_\_

**What type of load operation do you plan to perform: an initial load (point-in-time bulk load), incremental load (only the changes), or combined initial and incremental load (initial load followed by incremental load)?**

Answer: \_\_\_\_\_

**What are the number of cores, memory, and disk space on the system where the Secure Agent will run?**

Answer: \_\_\_\_\_

### Oracle environment

**What are the host name and port number of the Oracle source database server?**

Answer: \_\_\_\_\_

**What is the Oracle system identifier (SID) for the database?**

Answer: \_\_\_\_\_

**What are the Oracle database user name and password to use for connecting to the database?**

Answer: \_\_\_\_\_

**Is the Oracle Database Client or Instant Client installed on the system where the Secure Agent will run?**

Answer: \_\_\_\_\_

**Does the database run in an Oracle Real Application Cluster (RAC)? What's the maximum number of RAC members, including inactive nodes?**

Answer: \_\_\_\_\_

**Do you need to capture change data from an Oracle Data Guard logical or physical standby database?**

Answer: \_\_\_\_\_

**Do you need to capture change data from tables in a pluggable database (PDB) in an Oracle multitenant environment?**

Answer: \_\_\_\_\_

**Do you need to capture change data from tablespaces that use Oracle Transparent Data Encryption (TDE)? If yes, what are the TDE wallet directory and password?**

Answer: \_\_\_\_\_

**What is the typical size of units of work (UOWs) for the source tables?**

Answer: \_\_\_\_\_

### Oracle redo logs

**Are the redo logs in an Oracle Automatic Storage Management (ASM) environment? If you plan to connect to an ASM instance to read redo logs, are you allowed to create a login user ID for ASM that has SYSDBA or SYSASM authority?**

Answer: \_\_\_\_\_

**Are ARCHIVELOG mode and minimal global supplemental logging enabled for the Oracle source database? If not, can they be enabled?**

Answer: \_\_\_\_\_

**What are the primary and secondary archive log destinations for the archived redo logs from which you want to read change data?**

Answer: \_\_\_\_\_

**What is the average amount of archived redo log that is created per hour during peak and non-peak periods for the Oracle database?**

Answer: \_\_\_\_\_

**Do you have read access to the redo logs in your environment?**

Answer: \_\_\_\_\_

**If you do not have the authority to read the redo logs directly, can the archived redo log files be copied to shared disk or to a file system from which you can access them?**

Answer: \_\_\_\_\_

**Do you want Mass Ingestion Databases to read change data from the online log as well as the archived logs?**

Answer: \_\_\_\_\_

Can you make your archived redo logs available for diagnostic use by Informatica Global Customer Support, if necessary, to diagnose an error or anomaly during CDC processing?

Answer: \_\_\_\_\_

### More details for configuring database ingestion

What is the schema name for the source tables from which to replicate data?

Answer: \_\_\_\_\_

Do you want to replicate data from all tables in the schema or a subset of those tables? If a subset, create a list of them.

Answer: \_\_\_\_\_

Do the source tables have primary keys? Can supplemental logging be enabled for all of the primary keys?

Answer: \_\_\_\_\_

Do you have any unkeyed source tables?

Answer: \_\_\_\_\_

Do the source tables contain columns that have unsupported data types? To determine which data types are not supported for your source types, see the source-specific topics under "Mass Ingestion Databases source considerations" in the Mass Ingestion help.

Answer: \_\_\_\_\_

Is the default code page of UTF-8 acceptable? If not, which code page do you want to use?

Answer: \_\_\_\_\_

Do you want to use SSL to encrypt data exchanged between the Secure Agent and database server? Which encryption SSL or TLS protocol do you use?

Answer: \_\_\_\_\_

Are you allowed to create a new Oracle user and assign the privileges that Mass Ingestion Databases requires to that user? Determine the user name to use.

Answer: \_\_\_\_\_

Do the source tables contain any Oracle data types that Mass Ingestion Databases does not support?

Answer: \_\_\_\_\_

Do you want to capture schema drift changes on the source, including add, drop, modify, and rename column operations?

Answer: \_\_\_\_\_

## Oracle privileges

To deploy and run a database ingestion task that has an Oracle source, the source connection must specify a Mass Ingestion Databases user who has the privileges required for the ingestion load type.

### Privileges for incremental load processing

For a database ingestion task that performs an incremental load or combined initial and incremental load, ensure that the Mass Ingestion Databases user (*cmid\_user*) has been granted the following privileges:

**Note:** If the Oracle logs are managed by ASM, the user must have SYSASM or SYSDBA authority.

```
GRANT CREATE SESSION TO <cmid_user>;
```

```
GRANT SELECT ON table TO <cmid_user>;           -- For each source table created by user
GRANT EXECUTE ON DBMS_FLASHBACK TO <cmid_user>;
```

```
-- In the following, do not use ANY TABLE unless your security policy allows it.
```

```

GRANT FLASHBACK ON table|ANY TABLE TO <cmid_user>;

-- Include the following grant only if you want to Execute the CDC script for enabling
supplemental logging from the
-- user interface. If you manually enable supplemental logging, this grant is not needed.
GRANT ALTER table|ANY TABLE TO <cmid_user>;

GRANT SELECT ON DBA_CONSTRAINTS TO <cmid_user>;
GRANT SELECT ON DBA_CONS_COLUMNS TO <cmid_user>;
GRANT SELECT ON DBA_INDEXES TO <cmid_user>;
GRANT SELECT ON DBA_LOG_GROUPS TO <cmid_user>;
GRANT SELECT ON DBA_LOG_GROUP_COLUMNS TO <cmid_user>;
GRANT SELECT ON DBA_OBJECTS TO <cmid_user>;
GRANT SELECT ON DBA_OBJECT_TABLES TO <cmid_user>;
GRANT SELECT ON DBA_TABLES TO <cmid_user>;
GRANT SELECT ON DBA_TABLESPACES TO <cmid_user>;
GRANT SELECT ON DBA_USERS TO <cmid_user>;

GRANT SELECT ON "PUBLIC".V$ARCHIVED LOG TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$CONTAINERS TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$DATABASE TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$DATABASE_INCARNATION TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$ENCRYPTION_WALLET TO <cmid_user>;      -- For Oracle TDE access
GRANT SELECT ON "PUBLIC".V$LOG TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$LOGFILE TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$PARAMETER TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$PDBS TO <cmid_user>;      -- For Oracle multitenant environments
GRANT SELECT ON "PUBLIC".V$SPPARAMETER TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$STANDBY LOG TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$THREAD TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$TRANSACTION TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$TRANSPORTABLE_PLATFORM TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$VERSION TO <cmid_user>;

GRANT SELECT ON SYS.ATTRCOL$ TO <cmid_user>;
GRANT SELECT ON SYS.CCOL$ TO <cmid_user>;
GRANT SELECT ON SYS.CDEF$ TO <cmid_user>;
GRANT SELECT ON SYS.COL$ TO <cmid_user>;
GRANT SELECT ON SYS.COLTYPE$ TO <cmid_user>;
GRANT SELECT ON SYS.IDNSEQ$ TO <cmid_user>;
GRANT SELECT ON SYS.IND$ TO <cmid_user>;
GRANT SELECT ON SYS.INDPART$ TO <cmid_user>;
GRANT SELECT ON SYS.OBJ$ TO <cmid_user>;
GRANT SELECT ON SYS.PARTOBJ$ TO <cmid_user>;
GRANT SELECT ON SYS.RECYCLEBIN$ TO <cmid_user>;
GRANT SELECT ON SYS.TAB$ TO <cmid_user>;
GRANT SELECT ON SYS.TABCOMPART$ TO <cmid_user>;
GRANT SELECT ON SYS.TABPART$ TO <cmid_user>;
GRANT SELECT ON SYS.TABSUBPART$ TO <cmid_user>;

-- Also ensure that you have access to the following ALL_* views:
ALL_CONSTRAINTS
ALL_CONS_COLUMNS
ALL_ENCRYPTED_COLUMNS
ALL_INDEXES
ALL_IND_COLUMNS
ALL_OBJECTS
ALL_TABLES
ALL_TAB_COLS
ALL_TAB_PARTITIONS
ALL_USERS

```

### Privileges for initial load processing

For a database ingestion task that performs an initial load, ensure that the user has the following privileges at minimum:

```

GRANT CREATE SESSION TO <cmid_user>;

GRANT SELECT ON DBA_INDEXES TO <cmid_user>;

```

```

GRANT SELECT ON DBA_OBJECT TABLES TO <cmid_user>;
GRANT SELECT ON DBA_OBJECTS TO <cmid_user>;
GRANT SELECT ON DBA_TABLES TO <cmid_user>;
GRANT SELECT ON DBA_USERS TO <cmid_user>;
GRANT SELECT ON DBA_VIEWS TO <cmid_user>; -- Only if you unload data from views

GRANT SELECT ANY TABLE TO <cmid_user>;
-or-
GRANT SELECT ON table TO <cmid_user>; -- For each source table created by user
GRANT SELECT ON ALL_CONSTRAINTS TO <cmid_user>;
GRANT SELECT ON ALL_CONS_COLUMNS TO <cmid_user>;
GRANT SELECT ON ALL_ENCRYPTED_COLUMNS TO <cmid_user>;
GRANT SELECT ON ALL_IND_COLUMNS TO <cmid_user>;
GRANT SELECT ON ALL_INDEXES TO <cmid_user>;
GRANT SELECT ON ALL_OBJECTS TO <cmid_user>;
GRANT SELECT ON ALL_TAB_COLS TO <cmid_user>;
GRANT SELECT ON ALL_USERS TO <cmid_user>;

GRANT SELECT ON "PUBLIC"."V$DATABASE" TO <cmid_user>;
GRANT SELECT ON "PUBLIC"."V$CONTAINERS" TO <cmid_user>;
GRANT SELECT ON SYS.ATRCOL$ TO <cmid_user>;
GRANT SELECT ON SYS.CCOL$ TO <cmid_user>;
GRANT SELECT ON SYS.CDEF$ TO <cmid_user>;
GRANT SELECT ON SYS.COL$ TO <cmid_user>;
GRANT SELECT ON SYS.COLTYPE$ TO <cmid_user>;
GRANT SELECT ON SYS.IND$ TO <cmid_user>;
GRANT SELECT ON SYS.IDNSEQ$ TO <cmid_user>;
GRANT SELECT ON SYS.OBJ$ TO <cmid_user>;
GRANT SELECT ON SYS.RECYCLEBIN$ TO <cmid_user>;
GRANT SELECT ON SYS.TAB$ TO <cmid_user>;

```

## Oracle privileges for Amazon RDS for Oracle sources

If you have an Amazon RDS for Oracle source, you must grant certain privileges to the Mass Ingestion Databases user.

**Important:** You must log in to Amazon RDS under the master username to run GRANT statements and procedures.

To grant the SELECT privilege, at minimum, on objects and system tables that are required for CDC processing, execute the following GRANT statements to the Mass Ingestion Databases user (*cmid\_user*):

```

GRANT SELECT ON "PUBLIC"."V$ARCHIVED LOG" TO "cmid_user";
GRANT SELECT ON "PUBLIC"."V$DATABASE" TO "cmid_user";
GRANT SELECT ON "PUBLIC"."V$LOG" TO "cmid_user";
GRANT SELECT ON "PUBLIC"."V$LOGFILE" TO "cmid_user";
GRANT SELECT ON "PUBLIC"."V$TRANSPORTABLE_PLATFORM" TO "cmid_user";
GRANT SELECT ON "PUBLIC"."V$THREAD" TO "cmid_user";
GRANT SELECT ON "PUBLIC"."V$DATABASE_INCARNATION" TO "cmid_user";

GRANT SELECT ON "SYS"."DBA_LOG_GROUPS" TO "cmid_user";
GRANT SELECT ON "SYS"."DBA_LOG_GROUP_COLUMNS" TO "cmid_user";
GRANT SELECT ON "SYS"."DBA_TABLESPACES" TO "cmid_user";

GRANT SELECT ON "SYS"."OBJ$" TO "cmid_user";
GRANT SELECT ON "SYS"."TAB$" TO "cmid_user";
GRANT SELECT ON "SYS"."IND$" TO "cmid_user";
GRANT SELECT ON "SYS"."COL$" TO "cmid_user";

GRANT SELECT ON "SYS"."PARTOBJ$" TO "cmid_user";
GRANT SELECT ON "SYS"."TABPART$" TO "cmid_user";
GRANT SELECT ON "SYS"."TABCOMPART$" TO "cmid_user";
GRANT SELECT ON "SYS"."TABSUBPART$" TO "cmid_user";
COMMIT;

```

To grant the SELECT privilege on some additional objects, run the following Amazon RDS procedures:

```

begin
rdsadmin.rdsadmin_util.grant_sys_object(

```

```

p_obj_name => 'DBA_USERS',
p_grantee => 'cmid_user',
p_privilege => 'SELECT',
p_grant_option => false);
end;
--
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'ALL TABLES',
p_grantee => 'cmid_user',
p_privilege => 'SELECT',
p_grant_option => false);
end;
--
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'ALL TAB PARTITIONS',
p_grantee => 'cmid_user',
p_privilege => 'SELECT',
p_grant_option => false);
end;
--
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'V_$PARAMETER',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
--
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'V_$SPPARAMETER',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
--
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'V_$STANDBY_LOG',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
--
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'V_$VERSION',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
--
begin
rdsadmin.rdsadmin_util.grant_sys_object(
p_obj_name => 'INDPART$',
p_grantee => 'cmid_user',
p_privilege => 'SELECT');
end;
--
end;

```

To provide read access to the Amazon RDS online and archived redo logs, execute the following GRANT statements:

```

GRANT READ ON DIRECTORY ONLINELOG_DIR to "cmid_user";
GRANT READ ON DIRECTORY ARCHIVELOG_DIR to "cmid_user";

```



## Oracle log access methods for CDC

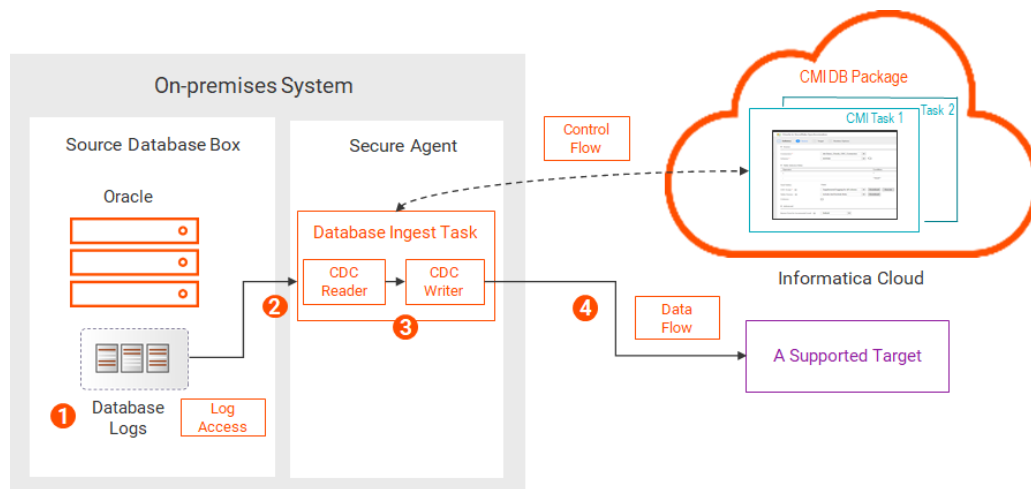
Database ingestion incremental load and combined initial and incremental load jobs can access the Oracle redo logs for CDC processing in alternative ways, depending on your environment and requirements.

### Direct log access

Database ingestion jobs can directly access the physical Oracle redo logs on the on-premises source system to read change data.

**Note:** If you store the logs on solid-state disk (SSD), this method can provide the best performance.

The following image shows the data flow:

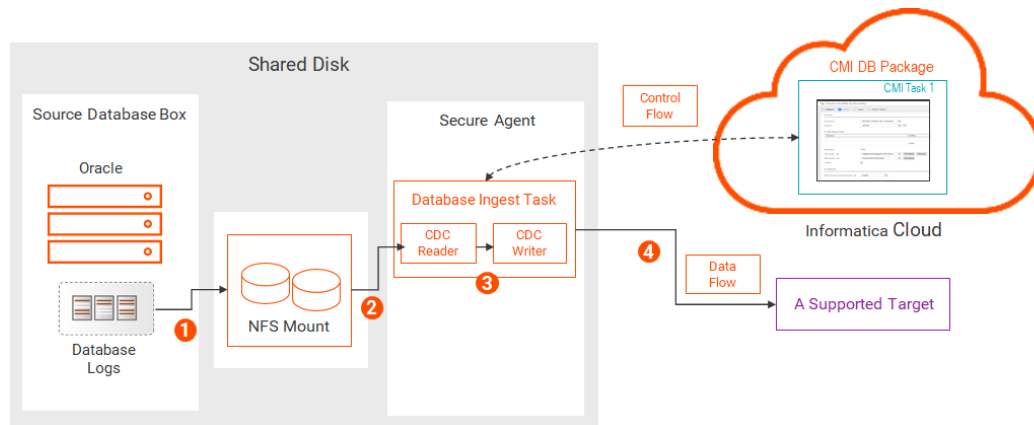


1. The Oracle database writes change records to the database log files on disk.
2. The Mass Ingestion Databases CDC Reader reads the physical log files and extracts change records from the log files for the source tables of CDC interest.
3. The Mass Ingestion Databases CDC Writer reads the change records.
4. The CDC Writer applies the change records to the target.

### NFS-mounted logs

Database ingestion jobs can access to Oracle database logs from shared disk by using a Network File Sharing (NFS) mount or another method such as Network Attached Storage (NAS) or clustered storage.

The following image shows the data flow:

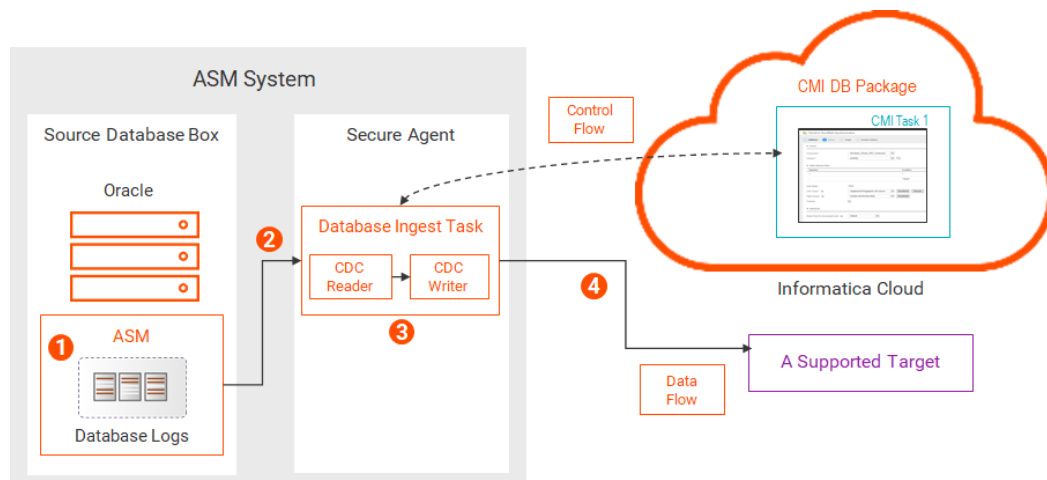


1. The Oracle database writes change records to database log files. The log files are written to shared disk. The shared disk can be on any system that allows the files to appear as local to both the database and Secure Agent hosts. This sharing can be achieved by using NFS, as shown above, or by using Network Attached Storage (NAS) or clustered storage.
2. The Mass Ingestion Databases CDC Reader reads the log files from the NFS server over the network and extracts the change records for the source tables of CDC interest.
3. The Mass Ingestion Databases CDC Writer reads the change records.
4. The CDC Writer applies the change records to the target.

### ASM-managed logs

Database ingestion jobs can access Oracle redo logs that are stored in an Oracle Automatic Storage Management (ASM) system. To read change data from the ASM-managed redo logs, the ASM user must have SYSASM or SYSDBA authority on the ASM instance.

The following image shows the data flow:



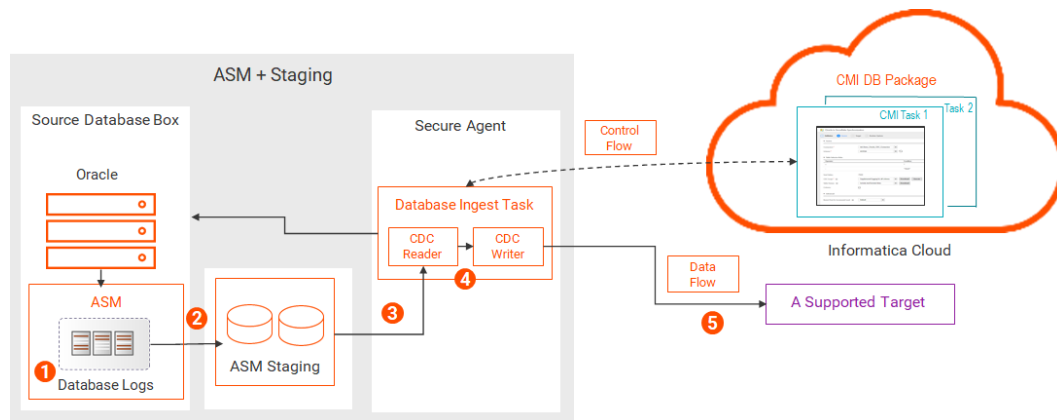
1. The Oracle database writes change records to the ASM-managed database log files.
2. The Mass Ingestion Databases CDC Reader reads the ASM-managed log files and extracts the change records for the source tables of CDC interest.
3. The Mass Ingestion Databases CDC Writer reads the change records.

4. The CDC Writer applies the change records to the target.

### ASM-managed logs with a staging directory

Database ingestion jobs can access ASM-managed redo logs from a staging directory in the ASM environment. In comparison to using ASM only, this method can provide faster access to the log files and reduce I/O on the ASM system. To read change data from the ASM-managed logs, the ASM user must have SYSASM or SYSDBA authority on the ASM instance.

The following image shows the data flow:

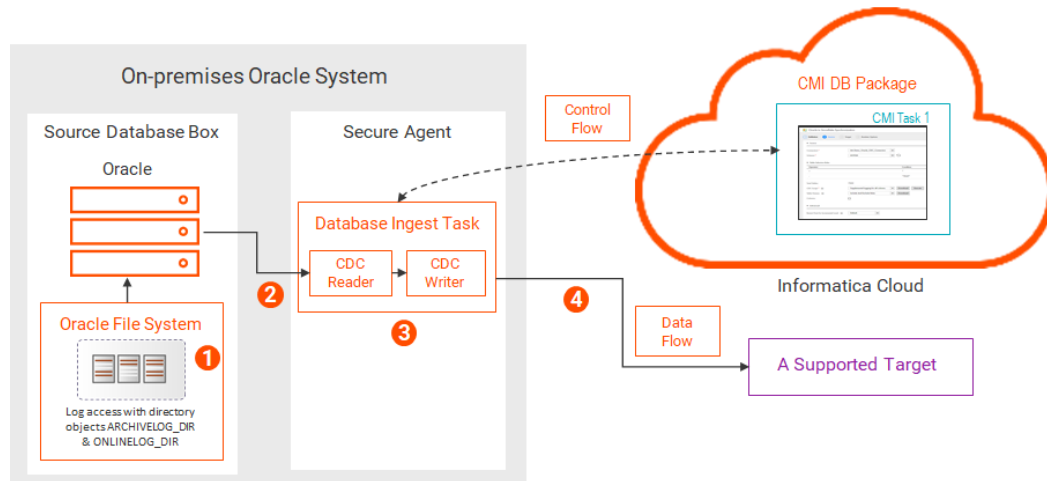


1. The Oracle database writes change records to the ASM-managed log files.
2. ASM copies the logs to a staging directory.  
The staging directory must be on shared disk, such as an NFS mount, so that ASM can write data to it and database ingestion jobs can read data from it.
3. The Mass Ingestion Databases CDC Reader reads the log files in the staging directory and extracts the change records for the source tables of CDC interest.
4. The Mass Ingestion Databases CDC Writer reads the change records.
5. The CDC Writer applies the change records to the target.

### BFILE access to logs in the Oracle server file system by using directory objects

On an on-premises Oracle source system, you can configure Mass Ingestion Databases to read online and archived redo logs from the local Oracle server file system by using Oracle directory objects with BFILE locators. You must create Oracle directory objects named ARCHIVELOG\_DIR and ONLINELOG\_DIR that point to the locations of the Oracle redo log files. For information about configuring BFILE access, see ["Configuring BFILE access to Oracle redo logs in the Oracle file system" on page 204](#).

The following image shows the data flow:



1. The Oracle database writes change records to the redo log files in the local Oracle server file system. When a database ingestion task needs to read log files, it connects to Oracle and issues a select request that references the ARCHIVELOG\_DIR or ONLINELOG\_DIR directory object to access the logs.
2. The Mass Ingestion Databases CDC Reader reads the log file and extracts the change records for the source tables of CDC interest.
3. The Mass Ingestion Databases CDC Writer reads the change records.
4. The CDC Writer applies the change records to the target.

## Configuring BFILE access to Oracle redo logs in the Oracle file system

If you store redo logs in the local Oracle server file system and want to access the logs by using Oracle directory objects with BFILES, perform the following configuration tasks:

Perform the following usual Oracle source preparation tasks, which are not specific to BFILE access:

- Define the ORACLE\_HOME environment variable on the Linux or Windows system where the Secure Agent runs for Mass Ingestion Databases to use the Oracle Call Interface (OCI) to communicate with the Oracle source database.
- Make sure the Mass Ingestion Databases user has the Oracle privileges that are required for the database ingestion incremental load processing. For more information, see [“Oracle privileges” on page 197](#).
- Enable ARCHIVELOG mode for the Oracle database.
- Define an archive log destination.
- Enable Oracle minimal global supplemental logging on the source database.
- If your Oracle source tables have primary keys, ensure that supplemental logging is enabled for all primary key columns. For source tables that do not have primary keys, ensure that supplemental logging is enabled for all columns from which change data will be captured.

**Note:** When you create a database ingestion task, you have the option of generating a script that implements supplemental logging for all columns or only primary key columns for the selected source tables.

- Ensure that the Oracle MAX\_STRING\_SIZE initialization parameter is *not* set to EXTENDED. If it is set to EXTENDED, Mass Ingestion Databases will not be able to replicate inserts and updates for tables containing columns defined with large (extended size) VARCHAR2, NVARCHAR2, or RAW columns.

Additionally, for BFILE access, perform the following steps:

1. Query the Oracle database for the online and archived redo log locations in the Oracle server file system. You can use the following example queries:

To get location of the online redo logs:

```
select * from v$logfile;
```

To get the archive log destination:

```
select dest_id, dest_name, destination, status from V$ARCHIVE_DEST;
```

2. Create the ONLINELOG\_DIR and ARCHIVELOG\_DIR directory objects that point to the locations of log files from step 1. An Oracle directory object specifies a logical alias name for a physical directory on the Oracle server file system under which the log files to be accessed are located. For example:

```
CREATE DIRECTORY ONLINELOG_DIR AS '/u01/oracle/data';
CREATE DIRECTORY ARCHIVELOG_DIR AS '/u01/oracle/archivedata';
```

**Note:** The Oracle database does not verify that the directories you specify exist. Make sure you specify valid directories that exist in the Oracle file system.

3. To verify that the directory objects were created with the correct file system paths for the redo logs, issue a select statement such as:

```
select * from all_directories;
OWNER      DIRECTORY_NAME      DIRECTORY_PATH
-----
SYS        ARCHIVELOG_DIR          /u01/oracle/data/JO112DTL
SYS        ONLINELOG_DIR          /u01/oracle/data/JO112DTL
```

4. Grant read and write access on the ONLINELOG\_DIR and ARCHIVELOG\_DIR directory objects to the Mass Ingestion Databases user who is specified in the Oracle Database Ingestion connection properties. For example:

```
grant read, write on directory "ARCHIVELOG_DIR" to "cmid_user";
grant read, write on directory "ONLINELOG_DIR" to "cmid_user";
```

5. In the Oracle Database Ingestion connection properties, select the **BFILE Access** check box.

## PostgreSQL sources

To use PostgreSQL sources in database ingestion tasks, first prepare the source database and review the usage considerations.

### PostgreSQL source preparation

To use PostgreSQL sources in database ingestion tasks, first prepare the source database and review the usage considerations.

1. On Windows, install the latest version of the 64-bit PostgreSQL ODBC driver.
  - a. Download and install the PostgreSQL ODBC driver.

**Note:** If the source database contains objects with multibyte-character names, such as table names, column names, and publication names, you must use either a PostgreSQL Unicode ODBC driver or the DataDirect ODBC for PostgreSQL driver. This requirement applies to all PostgreSQL source types, including Amazon Aurora PostgreSQL and RDS for PostgreSQL. If you do not use a Unicode-compatible ODBC driver, your incremental load jobs will fail when encountering a multibyte-character name.
  - b. Set the PGSQL\_ODBC\_DRIVER environment variable to the driver name that is displayed by ODBC Data Source Administrator (64-bit).

On Linux and UNIX, the DataDirect ODBC driver for PostgreSQL is delivered as part of the Linux installation. You can optionally install the unixODBC driver manager or PostgreSQL ODBC driver.

- a. Optional. Install the unixODBC or iODBC driver manager.
- b. Optional. Install the PostgreSQL ODBC driver if you do not want to use the DataDirect ODBC for PostgreSQL driver that is provided in the Linux installation.

**Note:** If the source database contains objects with multibyte-character names, such as table names, column names, and publication names, you must use either a PostgreSQL Unicode ODBC driver or the DataDirect ODBC for PostgreSQL driver. This requirement applies to all PostgreSQL source types, including Amazon Aurora PostgreSQL and RDS for PostgreSQL. If you do not use a Unicode-compatible ODBC driver, your incremental load jobs will fail when encountering a multibyte-character name.

- c. Add a PostgreSQL entry to `odbcinst.ini`.

```
[PGSQL]
Description = ODBC for PostgreSQL
Driver =
Setup =
Driver64 = /usr/pgsql-9.6/lib/psqlodbc.so
Setup64 = /usr/lib64/libodbcpsqlS.so
FileUsage = 1
```

- d. Optional. Set the following environment variables:

- Set the `ODBCSYSINI` variable to the directory where `odbcinst.ini` is located. If `odbcinst.ini` is located in the default `/etc` directory, you do not need to set the `ODBCSYSINI` variable.
- Add the directory where the PostgreSQL ODBC driver is installed to the `LD_LIBRARY_PATH` variable. If the driver is installed in the default directory of `/usr/lib64`, you do not need to add the path to the `LD_LIBRARY_PATH` variable.
- Set the `PGSQL_ODBC_DRIVER` parameter to the driver name that you specified in `odbcinst.ini`.

For example:

```
export ODBCSYSINI=/root/infraagent
export LD_LIBRARY_PATH=/usr/pgsql-9.6/lib
export PGSQL_ODBC_DRIVER=PGSQL
```

2. For incremental load jobs, ensure that the PostgreSQL `postgresql.conf` configuration file specifies the `wal_level=logical` parameter.

This parameter determines how much information PostgreSQL writes to the Write-Ahead Log (WAL). The setting of logical adds information that is required to support logical decoding.

To set `wal_level` to logical on Amazon Aurora PostgreSQL or Amazon Relational Database Service (RDS) for PostgreSQL sources, set the `rds.logical_replication` parameter to 1 in the cluster parameter group.

3. If you use the DataDirect ODBC for PostgreSQL driver, ensure that the database does not use the SCRAM-SHA-256 authentication method. Use another authentication method, such as MD5.

**Note:** The PostgreSQL ODBC driver supports the SCRAM-SHA-256 authentication method. In PostgreSQL 13, this authentication method became the default method.

4. To deploy and run a database ingestion task that includes a PostgreSQL source, the source connection must specify a database user who has the required privileges. Create the user and grant privileges to that user in the following ways:

- For initial load jobs, use the following SQL statements:

```
CREATE USER dbmi_user WITH PASSWORD password;
GRANT SELECT ON ALL TABLES IN SCHEMA schema TO dbmi_user;
```

- For incremental load jobs, use the following SQL statement:

```
CREATE USER dbmi_user WITH PASSWORD password REPLICATION;
```

Also, if you use the pgoutput plugin, use the following SQL statement to grant ownership of the tables in the database that you want to add to the pgoutput publication to the *dbmi\_user* that you created:

```
GRANT CREATE ON DATABASE database TO dbmi_user;
```

5. If you plan to use the wal2json plugin for logical decoding output for incremental load jobs, install the plugin.
6. If you plan to use the pgoutput plugin for incremental load jobs, use the following SQL statement to create publications for database ingestion jobs:

```
CREATE PUBLICATION publication_name [FOR TABLE [ONLY] table_name [*] [,...] | FOR  
ALL TABLES ]
```

Ensure that the publication includes all tables that you want to replicate to the target.

7. For incremental load jobs with PostgreSQL 9.6 sources, ensure that the *max\_replication\_slots* parameter in the *postgresql.conf* configuration file has a value greater than or equal to the number of concurrent database ingestion jobs that you plan to use.

**Important:** All replication slots must be unique across all concurrent jobs.

8. For incremental load jobs, ensure that the PostgreSQL sources use the UTF-8 encoding.

## Usage considerations

- Mass Ingestion Databases initial or incremental load jobs support RDS for PostgreSQL sources.
- Mass Ingestion Databases incremental load jobs support Amazon Aurora PostgreSQL sources.
- Mass Ingestion Databases supports schema drift options for PostgreSQL sources in database ingestion incremental load jobs. The following limitations apply:
  - PostgreSQL does not support renaming tables and columns for which Change Data Capture (CDC) is enabled.
  - PostgreSQL does not support changing primary keys for CDC tables.
  - Database ingestion jobs cannot capture DML changes from source tables if the table partition IDs are changed.
- For PostgreSQL 9.6, the pgoutput plugin is not available.
- For initial load jobs, Mass Ingestion Databases does not support the following PostgreSQL data types:
  - ABSTIME
  - Array types
  - JSON
  - NAME
  - Object identifier types
  - PG\_LSN
  - RELTIME
  - Text search types:
    - TSQUERY
    - TSVECTOR
  - User-defined types

For incremental load jobs, Mass Ingestion Databases does not support the following PostgreSQL data types, in addition to those not supported for initial load jobs:

- BYTEA
- MONEY

- Spatial types
  - Box
  - Circle
  - Line
  - LSeg
  - Path
  - Point
  - Polygon
- TEXT
- Unbounded varying types
- XML

Database ingestion jobs either fail to deploy or propagate nulls for columns that have these data types.

For information about the default mappings of supported data types, see [“Default Data Type Mappings” on page 276](#).

## SAP HANA sources

To use SAP HANA sources in database ingestion tasks, first prepare the source database and review the usage considerations.

### Source preparation

- The SAP HANA Database Ingestion connector uses JDBC to connect to the SAP HANA database to read data and metadata and to test connection properties. You must download the SAP HANA JDBC driver file, `ngdbc.jar`, and copy it to a specific subdirectory of the Secure Agent installation directory on the machine where the Secure Agent runs.
  1. Download the SAP HANA JDBC driver jar file, `ngdbc.jar`, to the Linux or Windows machine where the Secure Agent runs.  
Verify that you download the most recent version of the file. If you encounter any issues with downloading the file, contact SAP Customer Support.
  2. Copy the `ngdbc.jar` file to the following directory:  
`<Secure Agent installation directory>/ext/connectors/thirdparty/informatica.hanami`
  3. Restart the Secure Agent.
- To deploy and run a database ingestion task that includes a SAP HANA source, the source connection must specify a database user who has the privileges to read the following monitoring and system views:
  - SYS.M\_DATABASE
  - SYS.M\_CS\_PARTITIONS
  - SYS.SCHEMAS
  - SYS.TABLES
  - SYS.TABLE\_COLUMNS
  - SYS.INDEXES
  - SYS.INDEX\_COLUMNS

For incremental load tasks, grant the following privileges:



To enable the Mass Ingestion Databases user to write information about captured changes to the PKLOG table and to write change data to the shadow \_CDC tables, execute the following grant statement:

```
GRANT INSERT ON SCHEMA schema_name TO [user_id|user_role];
```

This statement grants the INSERT permission on the schema to the user that runs insert, update, and delete operations on the base source tables.

```
GRANT INSERT ON SCHEMA schema_name TO [schema_user];
```

This statement grants the INSERT privilege on the schema where the PKLOG, PROCESSED, and shadow \_CDC tables exist to the schema (as the user) where the triggers exist. This permission enables triggers to run with the permissions held by the schema where the triggers exist.

If you want to capture data from all or most tables in a database, execute the following statement to grant access to all objects in the source database:

```
GRANT SELECT ON SCHEMA schema_name TO [user_id|user_role];  
GRANT TRIGGER ON SCHEMA schema_name TO [user_id|user_role];
```

If you want to capture data from just a few tables, you can limit access to only those tables by executing the following statement for each selected source table:

```
GRANT SELECT ON database.table_name TO [user_id|user_role];  
GRANT TRIGGER ON database.table_name TO [user_id|user_role];
```

## Usage considerations

- Mass Ingestion Databases supports SAP HANA sources on Red Hat Linux or SUSE Linux for initial load and incremental load jobs but not for combined initial and incremental load jobs.
- Initial load and incremental load jobs with an SAP HANA source can have any target type except for Apache Kafka or Azure Event Hubs.
- Mass Ingestion Databases does not require primary keys on the source tables for initial load or incremental load jobs.
- Mass Ingestion Databases does not support the following source data types, even though they're mapped to default column data types on the target:
  - ARRAY
  - BINTTEXT
  - BLOB
  - CLOB
  - NCLOB
  - ST\_GEOMETRY
  - ST\_POINT
  - TEXT

Mass Ingestion Databases jobs propagate nulls for columns that have these data types.

For information about the default mappings of supported data types, see [“Default Data Type Mappings” on page 276](#).

- For incremental load jobs, Mass Ingestion Databases requires the following tables in the source database:
  - PKLOG log table. Contains metadata about captured DML changes, such as the change type and timestamp, transaction ID, schema name, and table name.
  - PROCESSED log table. Contains the maximum sequence number (SCN) for the most recent change data capture cycle.
  - Shadow *<schema>.<tablename>\_CDC* tables. Contains before images of updates and after images of inserts, updates, and deletes captured from the source tables, with metadata such as the transaction ID and timestamp. A shadow table must exist for each source table from which changes are captured.

Also, because SAP HANA does not provide direct access to its log files, Mass Ingestion Databases uses AFTER DELETE, AFTER INSERT, and AFTER UPDATE triggers on the source tables to get before images and after images of the DML changes for each source table and to write entries for the changes to the PKLOG table and shadow \_CDC tables. Mass Ingestion Databases also writes SAP HANA sequence values to each shadow \_CDC table and to the PKLOG table for each insert, update, and delete row processed. The sequence values link the rows of the shadow \_CDC table to the rows of the PKLOG table during CDC processing.

From the **Source** page in the task wizard, you can download or execute a CDC script that creates these tables, triggers, and sequences. If you specified a **Trigger Prefix** value in the SAP HANA Database Ingestion connection properties, the names of the generated triggers begin with *prefix\_*.

When you deploy a task, Mass Ingestion Databases validates the existence of the PKLOG, PROCESSED, and shadow \_CDC tables and the triggers and sequences. The deploy operation fails if these items do not exist.

- Database ingestion incremental load jobs support SAP HANA table names up to 120 characters in length.
- Schema drift options are not supported for incremental load jobs with SAP HANA sources.

## Teradata sources

To use Teradata sources in database ingestion tasks, first prepare the source database and review the usage considerations.

### Source preparation

- To deploy and run a database ingestion task that includes a Teradata source, the source connection must specify a database user who has the privileges that are required to perform an initial load operation. Use the following SQL statements to grant these privileges to the user:

```
GRANT SELECT ON database_name TO user_name/user_role;
```

### Usage considerations

- Mass Ingestion Databases does not support the following Teradata data types:
  - ARRAY
  - BLOB
  - CLOB
  - JSON
  - ST\_GEOMETRY
  - XML

For information about the default mappings of supported data types, see [“Default Data Type Mappings” on page 276](#).

# Mass Ingestion Databases targets - usage considerations

Before you configure database ingestion tasks for initial load, incremental load, or combined initial and incremental load operations, review the following guidelines for your target types to avoid unexpected results:

## Amazon Redshift targets

The following list identifies considerations for preparing and using Amazon Redshift targets:

- Before writing data to Amazon Redshift target tables, database ingestion jobs stage the data in an Amazon S3 bucket. You must specify the name of the bucket when you configure the database ingestion task. The ingestion jobs use the COPY command to load the data from the Amazon S3 bucket to the Amazon Redshift target tables. For more information about the COPY command, see the Amazon Web Services documentation.
- When you define a connection for an Amazon Redshift target, provide the access key and secret access key for the Amazon S3 bucket in which you want the database ingestion jobs to stage the data before loading it to the Amazon Redshift target tables.
- Incremental load jobs and combined initial and incremental load jobs generate a recovery table named INFORMATICA\_CDC\_RECOVERY on the target to store internal service information. The data in the recovery table prevents jobs that are restarted after a failure from propagating previously processed data again. The recovery table is generated in the same schema as the target tables.

## Amazon S3, Flat File, Google Cloud Storage, and Microsoft Azure Data Lake Storage targets

The following list identifies considerations for using Amazon S3, Flat File, Google Cloud Storage, and Microsoft Azure Data Lake Storage targets:

- When you define a database ingestion task that has an Amazon S3, Flat File, Google Cloud Storage, or Microsoft Azure Data Lake Storage target, you can select either CSV or Avro format for the generated output files that contain the source data to be applied to the target.
- If you select the **CSV** output format, Mass Ingestion Databases creates the following files on the target for each source table:
  - A schema.ini file that describes the schema and includes some settings for the output file on the target.
  - One or multiple output files for each source table, which contain the source data. Mass Ingestion Databases names these text files based on the name of the source table with an appended date and time.

The schema.ini file lists a sequence of columns for the rows in the corresponding output file. The following table describes the columns in the schema.ini file:

Column	Description
ColNameHeader	Indicates whether the source data files include column headers.
Format	Describes the format of the output files. Mass Ingestion Databases uses a comma (,) to delimit column values.

Column	Description
CharacterSet	Specifies the character set that is used for output files. Mass Ingestion Databases generates the files in the UTF-8 character set.
COL<sequence_number>	<p>The name and data type of the column.</p> <p><b>Notes:</b></p> <ul style="list-style-type: none"> <li>- If you selected any of the <b>Add Operation...</b> properties under <b>Advanced</b> on the <b>Target</b> page of the task wizard, the list of columns includes metadata columns for the operation type, time, owner, or transaction ID.</li> <li>- If you selected the <b>Add Before Images</b> check box, for each source column, the job creates a <i>column_name_OLD</i> column for UNDO data and <i>column_name_NEW</i> column for REDO data.</li> </ul>

**Important:** You should not edit the schema.ini file.

- If you select the **Avro** output format, you can select an Avro format type, a file compression type, an Avro data compression type, and the directory that stores the Avro schema definitions generated for each source table. The schema definition files have the following naming pattern: *schemaname\_tablename.txt*.
- On Flat File and Microsoft Azure Data Lake Storage targets, Mass Ingestion Databases creates an empty directory for each empty source table. Mass Ingestion Databases does not create empty directories on Amazon S3 and Google Cloud Storage targets.
- If you do not specify an access key and secret key in the Amazon S3 connection properties, Mass Ingestion Databases tries to find AWS credentials by using the default credential provider chain that is implemented by the DefaultAWSCredentialsProviderChain class. For more information, see the *Amazon Web Services* documentation.
- If database ingestion incremental load jobs replicate Update operations that change primary key values on the source to any of these targets that use the CSV output format, the job processes each Update record as two records on the target: a Delete followed by an Insert. The Delete contains the before image. The Insert contains the after image for the same row.

For Update operations that do not change primary key values, database ingestion jobs process each Update as one operation and writes only the after image to the target.

**Note:** If source tables do not have primary keys, Mass Ingestion Databases treats the tables as if all columns were part of the primary key. In this case, each Update operation is processed as a Delete followed by an Insert.

## Databricks Delta targets

To use Databricks Delta targets in database ingestion tasks, first prepare the target and review the usage considerations.

### Target preparation:

1. Download the Databricks JDBC driver from the Databricks website.
2. Copy the Databricks JDBC driver jar file, SparkJDBC42.jar, to the following directory:  
`Secure_Agent_installation_directory/apps/Database_Ingestion/ext/`
3. On Windows, install Visual C++ Redistributable Packages for Visual Studio 2013 on the computer where the Secure Agent runs.

### Usage considerations:

- For incremental load jobs, you must enable Change Data Capture (CDC) for all source columns.

- You can access Databricks Delta tables created on top of the following storage types:

- Microsoft Azure Data Lake Storage (ADLS) Gen2
- Amazon Web Services (AWS) S3

The Databricks Delta connection uses a JDBC URL to connect to the Databricks cluster. When you configure the target, specify the JDBC URL and credentials to use for connecting to the cluster. Also define the connection information that the target uses to connect to the staging location in Amazon S3 or ADLS Gen2.

- Before writing data to Databricks Delta target tables, database ingestion jobs stage the data in an Amazon S3 bucket or ADLS directory. You must specify the directory for the data when you configure the database ingestion task.

**Note:** Mass Ingestion Databases does not use the **ADLS Staging Filesystem Name** and **S3 Staging Bucket** properties in the Databricks Delta connection properties to determine the directory.

- Mass Ingestion Databases uses jobs that run once to load data from staging files on AWS S3 or ADLS Gen2 to external tables.  
By default, Mass Ingestion Databases runs jobs on the cluster that is specified in the Databricks Delta connection properties. If you want to run the jobs on another cluster, set the `dbDeltaUseExistingCluster` custom property to false on the **Target** page in the database ingestion task wizard.
- By default, Mass Ingestion Databases uses the Databricks Delta COPY INTO feature to load data from the staging file to Databricks Delta target tables. You can disable it for all load types by setting the `writerDatabricksUseSqlLoad` custom property to false on the **Target** page in the database ingestion task wizard.
- If you use an AWS cluster, you must specify the **S3 Service Regional Endpoint** value in the Databricks Delta connection properties.
- If you use Databricks Delta SQL endpoint to load data, you must specify the JDBC URL in the **SQL Endpoint JDBC URL** field in the Databricks Delta connection properties.

## Google BigQuery targets

The following list identifies considerations for preparing and using Google BigQuery targets:

- Download and install the Google BigQuery JDBC driver.
  1. Download the Google BigQuery JDBC driver zip from the [Google Cloud website](#).
  2. Copy all of the jar files in the installation zip to the following directory:  
`Secure Agent installation directory/apps/Database_Ingestion/ext/`
  3. Restart the Secure Agent.
- Ensure that you have a service account in your Google account to access Google BigQuery and Google Cloud Storage.
- Ensure that you have the `client_email`, `project_id`, `private_key`, `private_key_id`, `client_id`, and region ID values for the service account. You must enter these details when you create a Google BigQuery connection.

**Note:** Specify the `private_key_id` and `client_id` values in the **Provide Optional Properties** field of the connection properties. Use the following format:

```
"private_key_id": "<private_key_id_value>", "client_id": "<client_id_value>"
```

- If you want to configure a timeout interval for a Google BigQuery connection, specify the timeout interval property in the **Provide Optional Properties** field of the connection properties. Use the following format:

```
"timeout": "<timeout_interval_in_seconds>"
```

- Verify that you have read and write access to the following entities:
  - Google BigQuery datasets that contains the target tables
  - Google Cloud Storage path where Mass Ingestion Databases creates the staging file
- To write data to a Google BigQuery table, you must have the following permissions:
  - bigquery.datasets.get
  - bigquery.datasets.getIamPolicy
  - bigquery.models.\*
  - bigquery.routines.\*
  - bigquery.tables.create
  - bigquery.tables.delete
  - bigquery.tables.export
  - bigquery.tables.get
  - bigquery.tables.getData
  - bigquery.tables.list
  - bigquery.tables.update
  - bigquery.tables.updateData
  - bigquery.tables.updateTag
  - resourceManager.projects.get
  - resourceManager.projects.list
  - bigquery.jobs.create
- Mass Ingestion Databases loads source data in bulk mode to Google BigQuery targets.
- For database ingestion incremental load tasks, you must enable source database Change Data Capture (CDC) on all source columns.

## Kafka targets and Kafka-enabled Azure Event Hubs targets

The following list identifies considerations for using Kafka targets:

- Mass Ingestion Databases supports Apache Kafka, Confluent Kafka, Amazon Managed Streaming for Apache Kafka (MSK), and Kafka-enabled Azure Event Hubs as targets for incremental load jobs. All of these Kafka target types use the Kafka connection type.
 

To indicate the Kafka target type, you must specify Kafka producer properties in the task definition or Kafka connection properties. To specify these properties for a task, enter a comma-separated list of *key:value* pairs in the **Producer Configuration Properties** field on the **Target** page of the task wizard. To specify the producer properties for all tasks that use a Kafka connection, enter the list of properties in the **Additional Connection Properties** field in the connection properties. You can override the connection-level properties for specific tasks by also defining producer properties at the task level. For more information about producer properties, see the Apache Kafka, Confluent Kafka, Amazon MSK, or Azure Event Hubs for Kafka documentation.
- If you select **AVRO** as the output format for a Kafka target, Mass Ingestion Databases generates a schema definition file for each table with a name in the following format:

*schemaname\_tablename.txt*

If a source schema change is expected to alter the target in an incremental load job, Mass Ingestion Databases regenerates the Avro schema definition file with a unique name that includes a timestamp:

```
schemaname_tablename_YYYYMMDDhhmmss.txt
```

This unique naming pattern preserves older schema definition files for audit purposes.

- If you have a Confluent Kafka target that uses Confluent Schema Registry to store schemas, you must configure the following settings on the **Target** page of the task wizard:
  - In the **Output Format** field, select **AVRO**.
  - In the **Avro Serialization Format** field, select **None**.
- You can specify Kafka producer properties in either the **Producer Configuration Properties** field on the **Target** page of the task wizard or in the **Additional Connection Properties** field in the Kafka connection properties. Enter property=value pairs that meet your business needs and are supported by your Kafka vendor.

For example, if you use Confluent Kafka, you can use the following **Additional Connection Properties** entry to specify the Schema Registry URL:

```
schema.registry.url=http://abcxqa01:8081
key.serializer=org.apache.kafka.common.serialization.StringSerializer
value.serializer=io.confluent.kafka.serializers.KafkaAvroSerializer
```

If you use Amazon MSK, you can use the following **Additional Connection Properties** entry to enable IAM role authentication for access to Amazon MSK targets:

```
security.protocol=SASL_SSL,sasl.mechanism=AWS_MSK_IAM,sasl.jaas.config=software.amazon
.msk.auth.iam.IAMLoginModule
required;;sasl.client.callback.handler.class=software.amazon.msk.auth.iam.IAMClientCal
lbackHandler
```

Ensure that you enable IAM role authentication on the Amazon EC2 instance where the Secure Agent is installed.

For more information about Kafka properties, see the documentation of your Kafka vendor.

- Database ingestion incremental load jobs can replicate change data to Kafka targets that support SASL\_SSL secured access, including Confluent Kafka, Amazon MSK, and Azure Event Hubs targets. In Administrator, you must configure a Kafka connection that includes the appropriate properties in the **Additional Connection Properties** field. For example, for Azure Event Hubs, you could use the following **Additional Connection Properties** entry to enable SASL\_SSL:

```
bootstrap.servers=NAMESPACENAME.servicebus.windows.net:9093
security.protocol=SASL_SSL
sasl.mechanism=PLAIN
sasl.jaas.config=org.apache.kafka.common.security.plain.PlainLoginModule required
username="$ConnectionString" password="{YOUR.EVENTHUBS.CONNECTION.STRING}";
```

## Microsoft Azure Synapse Analytics targets

The following list identifies considerations for preparing and using Microsoft Azure Synapse Analytics targets:

- To deploy and run a database ingestion task with a Microsoft Azure Synapse Analytics target, the target connection must specify a database user who has the CONTROL permission on the target database. To grant the CONTROL permission to a user, use the following SQL statements:

```
USE database_name;
GRANT CONTROL TO user_name;
```

This permission is required for initial load, incremental load, and combined initial and incremental load jobs. This permission allows Mass Ingestion Databases to create target tables and database objects, such as external data source, external file format, and database scoped credential objects, if they do not

exist in the database. This permission is specifically required for creating external data source and database scoped credential objects.

**Note:** You must manually create the master key. To create the master key, you must have the CONTROL permission on the database.

- Database ingestion jobs first send data to a Microsoft Azure Data Lake Storage Gen2 staging file before writing the data to Microsoft Azure Synapse Analytics target tables. The staging file uses the hexadecimal x1d separator as the field delimiter. After the data is written to the target, the entire contents of the table-specific directory that includes the staging files are deleted.
- If you use Microsoft Azure Data Lake Storage Gen2 with a Synapse Analytics connection, you must enable the **Hierarchical namespace** option in Microsoft Azure Data Lake Storage. With this setting, use of blob storage is not recommended.
- The number of columns in a source table that a database ingestion job can propagate to a Microsoft Azure Synapse Analytics target must not exceed 508 columns.
- Database ingestion incremental load jobs and combined initial and incremental load jobs generate a recovery table named INFORMATICA\_CDC\_RECOVERY on the target to store internal service information that prevents jobs restarted after a failure from propagating previously processed data again. This recovery table is generated in the same schema as the target tables
- After a database ingestion job loads data to a Microsoft Azure Synapse Analytics target by using external tables, the job does not drop the log tables and external tables created on the target, even though these tables might be re-created when the job starts again.

## Oracle targets

The following list identifies considerations for preparing and using Oracle targets:

- Mass Ingestion Databases supports Oracle as a target in jobs that replicate data from Db2 for i, Microsoft SQL Server, and Oracle sources. For Db2 for i sources, only initial load jobs are supported. For Microsoft SQL Server and Oracle sources, initial load, incremental load, and initial and incremental load jobs are supported.
- Mass Ingestion Databases requires users to have certain privileges to load data to Oracle target databases. Grant the following user privileges to the Mass Ingestion Databases user (*cmid\_user*) to connect to the Oracle target:

```
GRANT CREATE SESSION TO <cmid_user>;

GRANT SELECT ON "PUBLIC".V$DATABASE TO <cmid_user>;
GRANT SELECT ON "PUBLIC".V$CONTAINERS TO <cmid_user>;

GRANT SELECT ON DBA_USERS TO <cmid_user>;
GRANT SELECT ON DBA_TABLES TO <cmid_user>;
GRANT SELECT ON DBA_OBJECT_TABLES TO <cmid_user>;
GRANT SELECT ON DBA_INDEXES TO <cmid_user>;
GRANT SELECT ON DBA_OBJECTS TO <cmid_user>;

GRANT CREATE TABLE TO <cmid_user>;
GRANT SELECT ON ALL_CONSTRAINTS TO <cmid_user>;
GRANT SELECT ON ALL_OBJECTS TO <cmid_user>;

GRANT SELECT ON SYS.TAB$ TO <cmid_user>;
GRANT SELECT ON SYS.RECYCLEBIN$ TO <cmid_user>;
GRANT SELECT ON SYS.CCOL$ TO <cmid_user>;
GRANT SELECT ON SYS.CDEF$ TO <cmid_user>;
GRANT SELECT ON SYS.OBJ$ TO <cmid_user>;
GRANT SELECT ON SYS.COLTYPE$ TO <cmid_user>;
GRANT SELECT ON SYS.ATTRCOL$ TO <cmid_user>;
GRANT SELECT ON SYS.IDNSEQ$ TO <cmid_user>;
GRANT SELECT ON SYS.ATTRCOL$ TO <cmid_user>;
```



```
GRANT SELECT ON SYS.IDNSEQ$ TO <cmid_user>;
GRANT SELECT ON SYS.IND$ TO <cmid_user>;
```

## Snowflake targets

The following list identifies considerations for preparing and using Snowflake targets:

- Before writing data to Snowflake target tables, database ingestion jobs first write the data to an internal staging area that has the stage name you specified in the associated database ingestion task.
- When you define a connection for a Snowflake target, you must set the **Additional JDBC URL Parameters** field to `database=target_database_name`. Otherwise, when you try to define the target in the database ingestion task wizard, an error message reports that the list of schemas cannot be retrieved.
- Database ingestion incremental load jobs and combined initial and incremental load jobs generate a recovery table named `INFORMATICA_CDC_RECOVERY` on the target to store internal service information that prevents jobs restarted after a failure from propagating previously processed data again. This recovery table is generated in the same schema as the target tables.

## Ingest incremental change operations into audit tables on the target

For database ingestion incremental load and combined initial and incremental load tasks that have Snowflake targets, you can configure the tasks to ingest data directly to audit tables on the target system, instead of merging and applying changes to a target database.

Consider using an audit table when you want to perform downstream computations or other processing on the data before writing it to the target database or when you want to examine the changes.

**Note:** Ensure that no constraints other than indexes exist on the audit tables.

When you define the task, you must select **Audit** in the **Apply Mode** field on the **Target** page to enable the use of audit tables. This field is available for new or undeployed tasks. The audit tables are created, one for each source table, the first time the job runs.

Under **Advanced** on the **Target** page, you can optionally select check boxes for adding metadata columns to the audit table:

- Select the **Add Operation <type>** check boxes to add columns that contain metadata for DML change operations, such as the operation type, time, transaction ID, and owner, and a generated ascending sequence number. The columns are populated when data is loaded to the target audit tables.
- Select the **Add Before Images** check box to add `_OLD` columns that contain before-image data. You can then compare the old and new column values in the audit tables.
- Select the **Audit Columns Prefix** check box to add a prefix to the names of added metadata columns to differentiate them from other table columns. Default is `INFA_`.

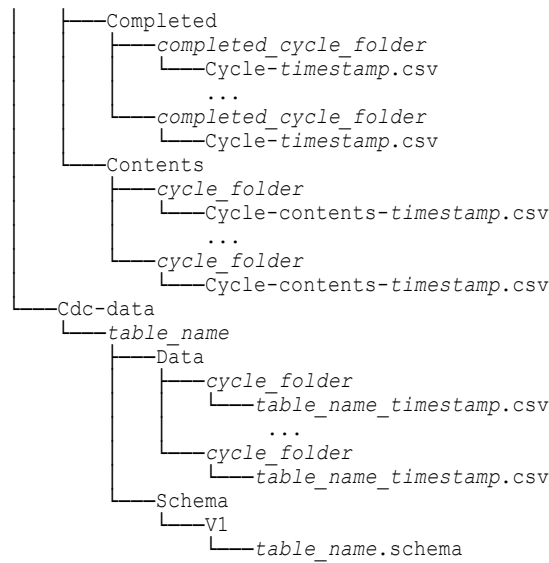
These fields are optional. Only the **Add Operation Type** check box is selected by default.

## Default directory structure for CDC files on Amazon S3, Google Cloud Storage, and Azure Data Lake Storage Gen2 targets

Database ingestion jobs create directories on Amazon S3, Google Cloud Storage, and Microsoft Azure Data Lake Storage Gen2 targets to store information about change data processing.

The following directory structure is created by default on the targets:

```
Bucket
├── connection_folder
│   └── job_folder
│       └── Cdc-cycle
```



The following table describes the directories in the default structure:

Folder	Description
<i>connection_folder</i>	Contains the Mass Ingestion Databases objects. This folder is specified in the <b>Folder Path</b> field of the Amazon S3 connection properties or in the <b>Directory Path</b> field of the Microsoft Azure Data Lake Storage Gen2 connection properties. <b>Note:</b> This folder is not created for Google Cloud Storage targets.
<i>job_folder</i>	Contains job output files. This folder is specified in the <b>Directory</b> field on the <b>Target</b> page of the database ingestion task wizard.
Cdc-cycle/Completed	Contains a subfolder for each completed CDC cycle. Each cycle subfolder contains a completed cycle file.
Cdc-cycle/Contents	Contains a subfolder for each CDC cycle. Each cycle subfolder contains a cycle contents file.
Cdc-data	Contains output data files and schema files for each table.
Cdc-data/table_name/Schema/V1	Contains a schema file. <b>Note:</b> Mass Ingestion Databases does not save a schema file in this folder if the output files use the Parquet format.
Cdc-data/table_name/Data	Contains a subfolder for each CDC cycle that produces output data files.

## Cycle directories

Mass Ingestion Databases uses the following pattern to name cycle directories:

[dt=]yyyy-mm-dd-hh-mm-ss

The "dt=" prefix is added to cycle folder names if you select the **Add Directory Tags** check box on the **Target** page of the database ingestion task wizard.

## Cycle contents files

Cycle contents files are located in `Cdc-cycle/Contents/cycle_folder` subdirectories. Cycle contents files contain a record for each table that has had a DML event during the cycle. If no DML operations occurred on a table in the cycle, the table does not appear in the cycle contents file.

Mass Ingestion Databases uses the following pattern to name cycle content files:

```
Cycle-contents-timestamp.csv
```

A cycle contents csv file contains the following information:

- Table name
- Cycle name
- Path to the cycle folder for the table
- Start sequence for the table
- End sequence for the table
- Number of Insert operations
- Number of Update operations
- Number of Delete operations
- Schema version
- Path to the schema file for the schema version

**Note:** If the output data files use the Parquet format, Mass Ingestion Databases does not save a schema file at the path that is specified in the cycle contents file. Instead, use the schema file in the folder that is specified in the **Avro Schema Directory** field on the **Target** page of the database ingestion task wizard.

## Completed cycle files

Completed cycle files are located in `Cdc-cycle/Completed/completed_cycle_folder` subdirectories. A database ingestion job creates a cycle file in this subdirectory after a cycle completes. If this file is not present, the cycle has not completed yet.

Mass Ingestion Databases uses the following pattern to name completed cycle files:

```
Cycle-timestamp.csv
```

A completed cycle csv file contains the following information:

- Cycle name
- Cycle start time
- Cycle end time
- Current sequence number at the time the cycle ended
- Path to the cycle contents file
- Reason for the end of cycle

Valid reason values are:

- **NORMAL\_COMMIT**. A commit operation was encountered after the cycle had reached the DML limit or the end of the cycle interval. A cycle can end only on a commit boundary.
- **NORMAL\_EXPIRY**. The cycle ended because the cycle interval expired. The last operation was a commit.

## Output data files

The data files contain records that include the following information:

- Operation type. Valid values are:
  - **I** for Insert operations
  - **U** for Update operations
  - **D** for Delete operations
- Sortable sequence number
- Data columns

**Note:** Insert and Delete records contain only after images. Update records contain both before and after images.

## Custom directory structure for output files on Amazon S3, Google Cloud Storage, Flat File, and ADLS Gen2 targets

You can configure a custom directory structure for the output files that *initial load* or *incremental load* jobs write to Amazon S3, Google Cloud Storage, Flat File, or Microsoft Azure Data Lake Storage (ADLS) Gen2 targets if you do not want to use the default structure.

### Initial loads

By default, initial load jobs write output files to *tablename\_timestamp* subdirectories under the parent directory. For Amazon S3, Flat File, and ADLS Gen2 targets, the parent directory is specified in the target connection properties if the **Connection Directory as Parent** check box is selected on the **Target** page of the task wizard.

- In an Amazon S3 connection, this parent directory is specified in the **Folder Path** field.
- In a Flat File connection, the parent directory is specified in the **Directory** field.
- In an ADLS Gen2 connection, the parent directory is specified in the **Directory Path** field.

For Google Cloud Storage targets, the parent directory is the bucket container specified in the **Bucket** field on the **Target** page of the task wizard.

You can customize the directory structure to suit your needs. For example, for initial loads, you can write the output files under a root directory or directory path that is different from the parent directory specified in the connection properties to better organize the files for your environment or to find them more easily. Or you can consolidate all output files for a table directly in a directory with the table name rather than write the files to separate timestamped subdirectories, for example, to facilitate automated processing of all of the files.

To configure a directory structure, you must use the **Data Directory** field on the **Target** page of the ingestion task wizard. The default value is `{TableName}_{Timestamp}`, which causes output files to be written to *tablename\_timestamp* subdirectories under the parent directory. You can configure a custom directory path by creating a directory pattern that consists of any combination of case-insensitive placeholders and directory names. The placeholders are:

- `{TableName}` for a target table name
- `{Timestamp}` for the date and time, in the format `yyyymmdd_hhmissms`, at which the initial load job started to transfer data to the target
- `{Schema}` for the target schema name
- `{YY}` for a two-digit year
- `{YYYY}` for a four-digit year

- {MM} for a two-digit month value
- {DD} for a two-digit day in the month


A pattern can also include the following functions:

- toLower() to use lowercase for the values represented by the placeholder in parentheses
- toUpper() to use uppercase for the values represented by the placeholder in parentheses

By default, the target schema is also written to the data directory. If you want to use a different directory for the schema, you can define a directory pattern in the **Schema Directory** field.

### Example 1

You are using an Amazon S3 target and want to write output files and the target schema to the same directory, which is under the parent directory specified in the **Folder Path** field of the connection properties. In this case, the parent directory is `idr-test/DEMO`. You want write all of the output files for a table to a directory that has a name matching the table name, without a timestamp. You must complete the **Data Directory** field and select the **Connection Directory as Parent** check box. The following image shows this configuration on the **Target** page of the task wizard:

 DB2\_AMAZON\_S3\_Demo

1 Definition
2 Source
3 Target
4 Schedule and Runtime Options

▼ Target

Connection: \*

S3\_Demo

↺

Output Format: ?

CSV

▼

Add Headers to CSV File:

☐

File Compression Type:

None

▼

Data Directory: \* ?

{TableName}

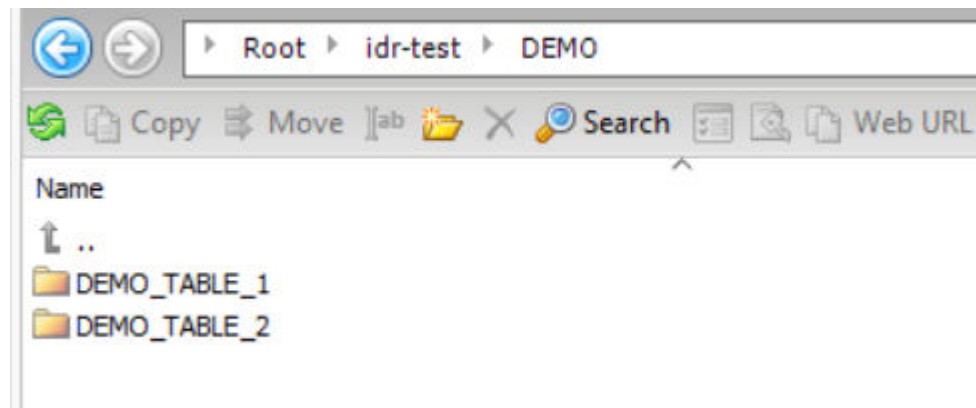
Connection Directory as Parent:

☒

Schema Directory: ?

Alternative schema directory name

Based on this configuration, the resulting directory structure is:



### Example 2

You are using an Amazon S3 target and want to write output data files to a custom directory path and write the target schema to a separate directory path. To use the directory specified in the **Folder Path** field in the Amazon S3 connection properties as the parent directory for the data directory and schema directory, select **Connection Directory as Parent**. In this case, the parent directory is `idr-test/DEMO`. In the **Data Directory** and **Schema Directory** fields, define directory patterns by using a specific directory name, such as `data_dir` and `schema_dir`, followed by the default `{TableName}_{Timestamp}` placeholder value. The placeholder creates `tablename_timestamp` destination directories. The following image shows this configuration on the **Target** page of the task wizard:

DB2\_AMAZON\_S3\_Demo\_Timestamp

1 Definition

2 Source

3 Target

4 Schedule and Runtime Options

▼ Target

Connection: \*

S3\_Demo

↺

Output Format: ?

CSV

▼

Add Headers to CSV File:

☐

File Compression Type:

None

▼

Data Directory: \* ?

data\_dir/{TableName}\_{Timestamp}

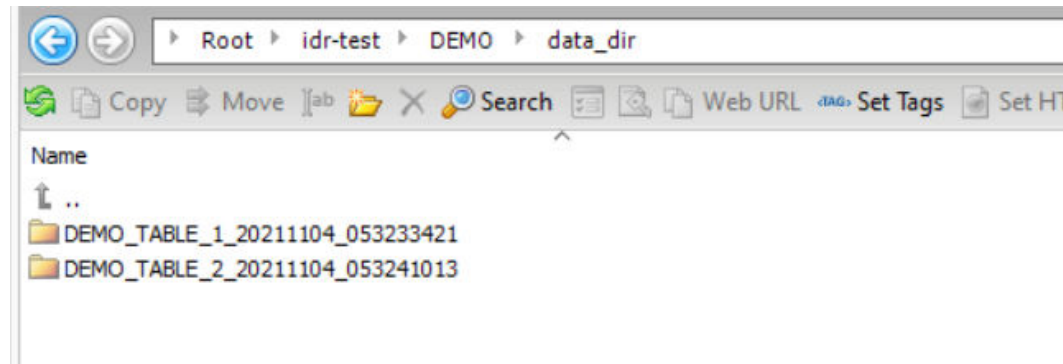
Connection Directory as Parent:

☒

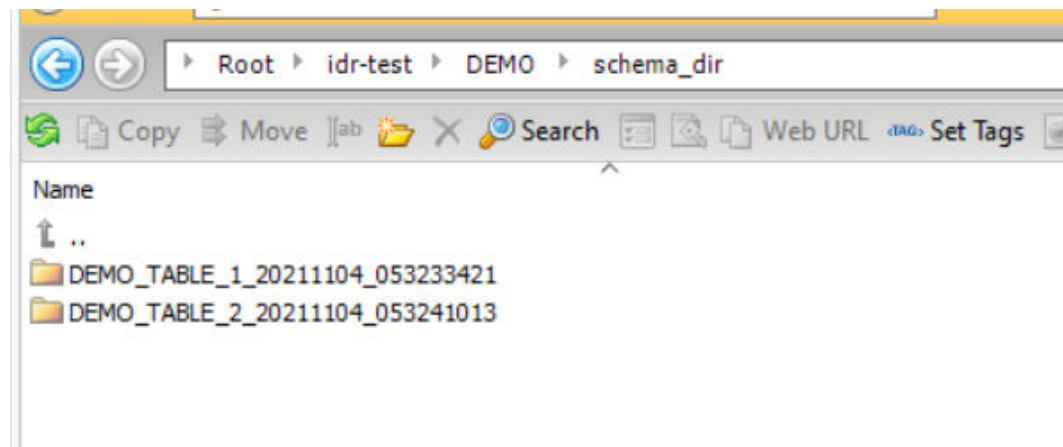
Schema Directory: ?

schema\_dir/{TableName}\_{Timestamp}

Based on this configuration, the resulting data directory structure is:



And the resulting schema directory structure is:



## Incremental loads

By default, incremental load jobs write Cdc-cycle files and Cdc-data files to subdirectories under the parent directory. However, you can create a custom directory structure to organize the files to best suit your organization's requirements.

This feature applies to database ingestion incremental load tasks that have Amazon S3, Google Cloud Storage, or Microsoft Azure Data Lake Storage (ADLS) Gen2 targets. It also applies to application ingestion incremental load jobs that have a Salesforce source and any one of these target types. This feature does not apply to combined initial and incremental load tasks.

For all targets except Google Cloud Storage, the parent directory is set in the target connection properties if the **Connection Directory as Parent** check box is selected on the **Target** page of the task wizard.

- In an Amazon S3 connection, the parent directory is specified in the **Folder Path** field.
- In an ADLS Gen2 connection, the parent directory is specified in the **Directory Path** field.

For Google Cloud Storage targets, the parent directory is the bucket container specified in the **Bucket** field on the **Target** page of the task wizard.

You can customize the directory structure to suit your needs. For example, you can write the cdc-data and cdc-cycle files under a target directory for the task instead of under the parent directory specified in the connection properties. Alternatively, you can 1) consolidate table-specific data and schema files under a subdirectory that includes the table name, 2) partition the data files and summary contents and completed files by CDC cycle, or 3) create a completely customized directory structure by defining a pattern that includes literal values and placeholders. For example, if you want to run SQL-type expressions to process the

data based on time, you can write all data files directly to timestamp subdirectories without partitioning them by CDC cycle.

To configure a custom directory structure for an incremental load task, define a pattern for any of the following optional fields on the **Target** page of the ingestion task wizard:

Field	Description	Default
Task Target Directory	<p>Name of a root directory to use for storing output files for an incremental load task.</p> <p>If you select the <b>Connection Directory as Parent</b> option, you can still optionally specify a task target directory. It will be appended to the parent directory to form the root for the data, schema, cycle completion, and cycle contents directories.</p> <p>This field is required if the {TaskTargetDirectory} placeholder is specified in patterns for any of the following directory fields.</p>	None
Connection Directory as Parent	Select this check box to use the parent directory specified in the connection properties.	Selected
Data Directory	<p>Path to the subdirectory that contains the cdc-data data files.</p> <p>In the directory path, the {TableName} placeholder is required if data and schema files are <i>not</i> partitioned by CDC cycle.</p>	{TaskTargetDirectory}/cdc-data/{TableName}/data
Schema Directory	<p>Path to the subdirectory in which to store the schema file if you do not want to store it in the data directory.</p> <p>In the directory path, the {TableName} placeholder is required if data and schema files are not partitioned by CDC cycle.</p>	{TaskTargetDirectory}/cdc-data/{TableName}/schema
Cycle Completion Directory	Path to the directory that contains the cdc-cycle completed file.	{TaskTargetDirectory}/cdc-cycle/completed
Cycle Contents Directory	Path to the directory that contains the cdc-cycle contents files.	{TaskTargetDirectory}/cdc-cycle/contents
Use Cycle Partitioning for Data Directory	<p>Causes a timestamp subdirectory to be created for each CDC cycle, under each data directory.</p> <p>If this option is not selected, individual data files are written to the same directory without a timestamp, unless you define an alternative directory structure.</p>	Selected
Use Cycle Partitioning for Summary Directories	Causes a timestamp subdirectory to be created for each CDC cycle, under the summary contents and completed subdirectories.	Selected
List Individual Files in Contents	<p>Lists individual data files under the contents subdirectory.</p> <p>If <b>Use Cycle Partitioning for Summary Directories</b> is cleared, this option is selected by default. All of the individual files are listed in the contents subdirectory unless you can configure custom subdirectories by using the placeholders, such as for timestamp or date.</p> <p>If <b>Use Cycle Partitioning for Data Directory</b> is selected, you can still optionally select this check box to list individual files and group them by CDC cycle.</p>	<p>Not selected if <b>Use Cycle Partitioning for Summary Directories</b> is selected.</p> <p>Selected if you cleared <b>Use Cycle Partitioning for Summary Directories</b>.</p>



A directory pattern consists of any combination of case-insensitive placeholders, shown in curly brackets {}, and specific directory names. The following placeholders are supported:

- {TaskTargetDirectory} for a task-specific base directory on the target to use instead of the directory the connection properties
- {TableName} for a target table name
- {Timestamp} for the date and time, in the format yyyyymmdd\_hhmissms
- {Schema} for the target schema name
- {YY} for a two-digit year
- {YYYY} for a four-digit year
- {MM} for a two-digit month value
- {DD} for a two-digit day in the month

**Note:** The timestamp, year, month, and day placeholders indicate when the CDC cycle started when specified in patterns for data, contents, and completed directories, or indicate when the CDC job started when specified in the schema directory pattern.

### Example 1

You want to accept the default directory settings for incremental load jobs as displayed in the task wizard. The target type is Amazon S3. Because the **Connection Directory as Parent** check box is selected by default, the parent directory path that is specified in the **Folder Path** field of the Amazon S3 connection properties is used. This parent directory is `idr-test/dbmi`. You also must specify a task target directory name, in this case, `s3_target`, because the {TaskTargetDirectory} placeholder is used in the default patterns in the subsequent directory fields. The files in the data directory and schema directory will be grouped by table name because the {TableName} placeholder is included in their default patterns. Also, because cycle partitioning is enabled, the files in the data directory, schema directory, and cycle summary directories will be subdivided by CDC cycle. The following image shows the default configuration settings on the **Target** page of the task wizard, except for the specified task target directory name:

s3\_DEMO

< Back Next > Save View Deploy

Add Headers to CSV File: ☐

File Compression Type:

Add Directory Tags: ☐

Task Target Directory:

Connection Directory as Parent: ☒

Data Directory:

Schema Directory:

Cycle Completion Directory:

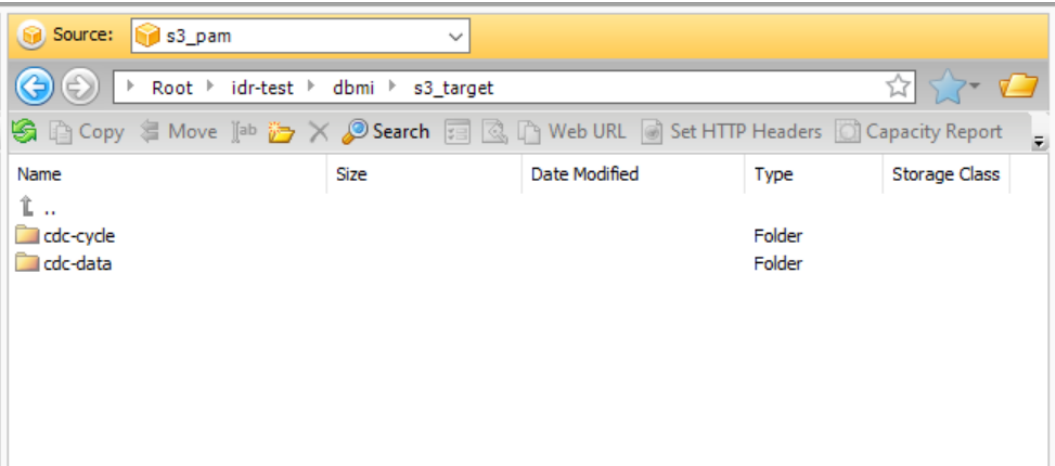
Cycle Contents Directory:

Use Cycle Partitioning for Data Directory: ☒

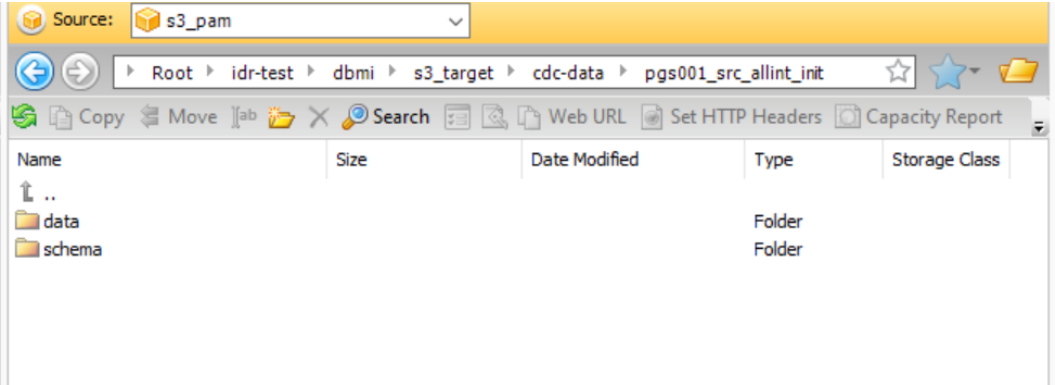
Use Cycle Partitioning for Summary Directories: ☒

List Individual Files in Contents: ☐

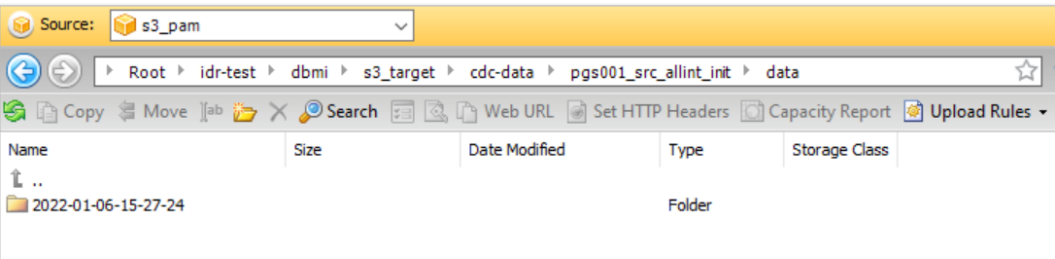
Based on this configuration, the resulting data directory structure is:



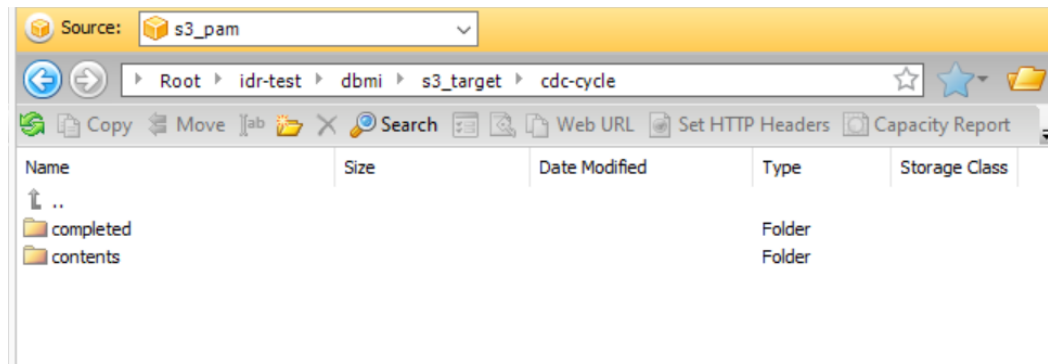
If you drill down on the cdc\_data folder and then on a table in that folder (pgs001\_src\_allint\_init), you can access the data and schema subdirectories:



If drill down on the data folder, you can access the timestamp directories for the data files:



If you drill down on cdc-cycle, you can access the summary contents and completed subdirectories:



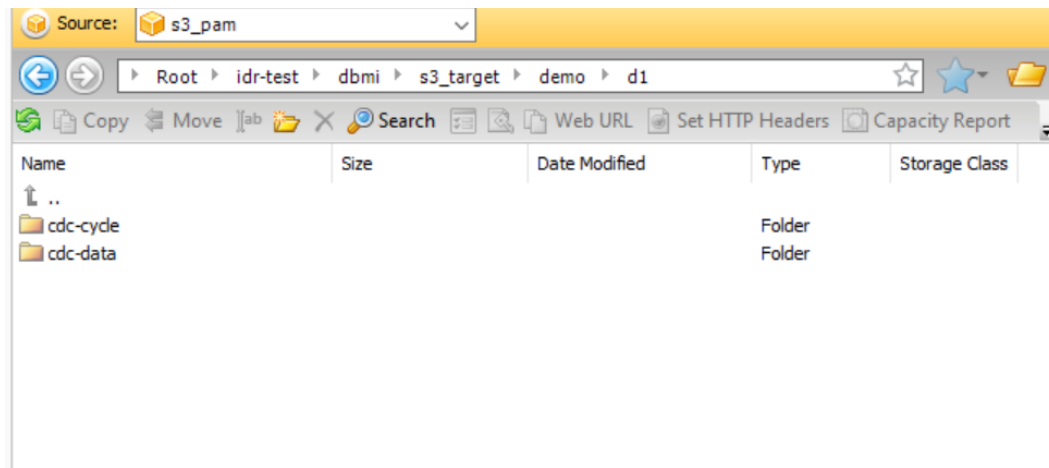
## Example 2

You want to create a custom directory structure for incremental load jobs that adds the subdirectories "demo" and "d1" in all of the directory paths except in the schema directory so that you can easily find the files for your demos. Because the **Connection Directory as Parent** check box is selected, the parent directory path (idr-test/dbmi) that is specified in the **Folder Path** field of the Amazon S3 connection properties is used. You also must specify the task target directory because the {TaskTargetDirectory} placeholder is used in the patterns in the subsequent directory fields. The files in the data directory and schema directory will be grouped by table name. Also, because cycle partitioning is enabled, the files in the data, schema, and cycle summary directories will be subdivided by CDC cycle. The following image shows the custom configuration on the **Target** page of the task wizard:

The screenshot shows the 's3\_DEMO' task wizard configuration page. The page has a title bar with 's3\_DEMO' and navigation buttons: '< Back', 'Next >', 'Save', 'View', and 'Deploy'. The configuration options are as follows:

- Add Headers to CSV File: ☐
- File Compression Type:
- Add Directory Tags: ☐
- Task Target Directory:
- Connection Directory as Parent: ☒
- Data Directory:
- Schema Directory:
- Cycle Completion Directory:
- Cycle Contents Directory:
- Use Cycle Partitioning for Data Directory: ☒
- Use Cycle Partitioning for Summary Directories: ☒
- List Individual Files in Contents: ☐

Based on this configuration, the resulting data directory structure is:



## Supported Avro data types

Mass Ingestion Databases supports some of the primitive and logical data types that Avro schemas provide. These data types pertain to target types that support Avro or Parquet output format.

A primitive data type represents a single data value. A logical data type is an Avro primitive or complex data type that has additional attributes to represent a derived type.

The following table lists the primitive Avro data types that Mass Ingestion Databases supports:

Primitive data type	Description
INT	32-bit signed integer
LONG	64-bit signed integer
FLOAT	Single precision (32-bit) IEEE 754 floating-point number
DOUBLE	Double precision (64-bit) IEEE 754 floating-point number
BYTES	Sequence of 8-bit unsigned bytes
STRING	Unicode character sequence

The following table lists the logical Avro data types that Mass Ingestion Databases supports:

Logical data type	Description
DECIMAL	An arbitrary-precision signed decimal number of the form $\text{unscaled} \times 10^{-\text{scale}}$
DATE	A date, without reference to a time or time zone.
TIME	A time of day that has the precision of 1 millisecond or 1 microsecond, without reference to a time zone or date.
TIMESTAMP	A date and time value that has the precision of 1 millisecond or microsecond, without reference to a particular calendar or time zone.

For Databricks Delta targets, Mass Ingestion Databases does not use the following data types in the intermediate Parquet files:

- **TIMESTAMP**, with millisecond precision
- **TIME**, with either millisecond or microsecond precision

## Schema drift handling

Mass Ingestion Databases can be configured to automatically detect some source schema changes and handle these changes on the target. This process is referred to as *schema drift*.

Mass Ingestion Databases can detect the following types of source schema changes:

- Add column
- Modify column
- Drop column
- Rename column

This feature is available for database ingestion incremental load tasks and combined initial and incremental load tasks that propagate change data from Microsoft SQL Server or Oracle sources to Amazon Redshift, Google BigQuery, Kafka, Microsoft Azure Synapse Analytics, Oracle, or Snowflake targets, or from a PostgreSQL source to Amazon Redshift, Google BigQuery, Kafka, Microsoft Azure Synapse Analytics, or Snowflake targets. By default, this feature is *not* enabled.

When you define a task, on the **Schedule and Runtime Options** page of the database ingestion task wizard, you can configure the types of source schema changes to propagate and how to handle them. For example, you can configure schema drift options to ignore the changes, replicate them, or stop the job or subtask when a schema change occurs. For more information, see [“Configuring schedule and runtime options” on page 267](#).

### Considerations:

- If you try to replicate a type of schema change that is not supported on the target, the database ingestion job ends with an error.
- Mass Ingestion Databases does not replicate source changes that add, remove, or modify primary key or unique key constraints. If these types of changes occur on the source, you must resynchronize the target tables.
- If you configured schema drift options to stop the job when Mass Ingestion Databases detects a schema change, you can use the **Resume With Options** command to resume the job with an override schema drift option.
- Mass Ingestion Databases detects a schema change in a source table only after DML operations occur on the altered source table. If multiple schema changes occur without intervening DML operations, Mass Ingestion Databases detects all of the schema changes at one time, when a DML operation occurs. To ensure that Mass Ingestion Databases detects all of the supported schema changes correctly, Informatica recommends that you apply schema changes to source tables one by one, each followed by at least one DML change.
- Database ingestion tasks that have Microsoft Azure Synapse Analytics targets cannot replicate rename operations on source columns. The Replicate option is not available.

- Database ingestion tasks that have Snowflake targets support modify operations on source columns with the following limitations:
  - Snowflake targets cannot modify the scale of NUMBER columns.
  - Snowflake targets do not support changing the data type of an existing column to a different data type.
- Database ingestion tasks that have Google BigQuery targets cannot replicate rename or modify operations on source columns. The schema drift options for these operations are not available.

## Ability to apply deletes as soft deletes on the target

For database ingestion incremental load and combined initial and incremental load jobs that have any supported source type and a Snowflake target, you can configure the task to process delete operations on the source as soft deletes on the target.

A soft delete marks a deleted row as deleted without actually removing it from the database. The row is applied to the target with the value of "D" in the generated INFA\_OPERATION\_TYPE metadata column.

Example scenario: Your organization wants to use soft deletes in a data warehouse to mark the rows that were deleted at the source while still retaining the rows for audit purposes.

To enable soft deletes, set the **Apply Mode** field to **Soft Deletes on Target** page of the task wizard when you configure the ingestion task. Also, make sure that the source tables have primary keys. Do not change the primary keys after you run the jobs.

## Configuring a database ingestion task

In Mass Ingestion, use the database ingestion task wizard to configure a database ingestion task.

On the wizard pages, complete the following configuration tasks:

1. Define basic task information, such as the task name, project location, runtime environment, and load type.
2. Configure the source.
3. Configure the target.
4. Configure runtime options.

Click **Next** or **Back** to navigate from one page to another. At any point, you can click **Save** to save the information that you have entered so far.

After you complete all wizard pages, save the information and then click **Deploy** to make the task available as an executable job to the Secure Agent.

### Before you begin

Before you begin, complete the following prerequisite tasks in Administrator:

- Check that your organization has licenses for Mass Ingestion Databases and the DBMI packages.
- Verify that the Secure Agent in your runtime environment is running and that you can access the Mass Ingestion service.

- Define the source and target connections.

Also, if you plan to perform incremental load operations with Oracle sources, ensure that the ORACLE\_HOME environment variable is defined on the Secure Agent system.

## Defining basic task information

To begin defining a database ingestion task, you must first enter some basic information about the task, such as a task name, project or project folder location, and load operation type.

1. In Mass Ingestion, click **New > Database Ingestion Task**.



The **Definition** page of the Mass Ingestion Databases Task wizard appears.

2. Configure the following properties:

Property	Description
Name	<p>A name for the database ingestion task.</p> <p>Task names can contain Latin alphanumeric characters, spaces, periods (.), commas (,), underscores (_), plus signs (+), and hyphens (-). Task names cannot include other special characters.</p> <p>Task names are not case sensitive.</p> <p>Maximum length is 50 characters.</p> <p><b>Note:</b> If you include spaces in the database ingestion task name, after you deploy the task, the spaces do not appear in the corresponding job name.</p>
Location	The project or project\folder that will contain the task definition.
Runtime Environment	<p>The runtime environment in which you want to run the task.</p> <p>The runtime environment must be a Secure Agent group that consists of one or more Secure Agents. A Secure Agent is a lightweight program that runs tasks and enables secure communication.</p> <p>You cannot use the Hosted Agent or a serverless runtime environment.</p> <p><b>Tip:</b> Click the <b>Refresh</b> icon to refresh the list of runtime environments.</p>

Property	Description
Description	An optional description for the task. Maximum length is 4,000 characters.
Load Type	The type of load operation that the database ingestion task performs. Options are: <ul style="list-style-type: none"> <li>- <b>Initial Load.</b> Loads data read at a specific point in time from source tables to a target in a batch operation. You can perform an initial load to materialize a target to which incremental change data will be sent.</li> <li>- <b>Incremental Load.</b> Propagates source data changes to a target continuously or until the job is stopped or ends. The job propagates the changes that have occurred since the last time the job ran or from a specific start point for the first job run.</li> <li>- <b>Initial and Incremental Loads.</b> Performs an initial load of point-in-time data to the target and then automatically switches to propagating incremental data changes made to the same source tables on a continuous basis.</li> </ul>

3. Click **Next**.

## Configuring the source

Configure the source on the **Source** page of the database ingestion task wizard.

**Note:** For MongoDB sources only, the task wizard displays *database* instead of *schema* and displays *collection* instead of *table*. However, for simplicity, the terms *schema* and *table* are used in this documentation to cover all source types.

1. In the **Connection** list, select the connection for the source system.

The connection must be predefined in Administrator for a runtime environment that your organization uses.

The list includes only the connection types that are valid for the load type selected on the **Definition** page. No connections are listed if you did not select a load type.

If you change the load type and the selected connection is no longer valid, a warning message is issued and the **Connection** field is cleared. You must select another connection that is valid for the updated load type.

**Note:** After you deploy the ingestion task, you cannot change the connection without first undeploying the associated ingestion job. After you change the connection, you must deploy the task again.

2. In the **Schema** list, select the source schema that includes the source tables.

The list includes only the schemas that are available in the database that is accessed with the specified source connection. When creating a task that has an Oracle, Microsoft SQL Server, Netezza, or PostgreSQL source, the schema name that is specified in the connection properties is displayed by default.

An expanded view of the **Table Selection** area appears.

3. If you are defining a Db2 for i source for an incremental load task, in the **Journal Name** field, select the name of the journal that records the changes made to the source tables.



4. If you are defining a PostgreSQL source for an incremental load task, specify the following fields:

Field	Description
Replication Slot Name	Specify the unique name of a PostgreSQL replication slot. A slot name can contain Latin alphanumeric characters in lowercase and the underscore (_) character. Maximum length is 63 characters. <b>Important:</b> Each database ingestion task must use a different replication slot.
Replication Plugin	Select a PostgreSQL replication plugin. Options are: - <b>pgoutput</b> . You can select this option only for PostgreSQL version 10 and later. - <b>wal2json</b>
Publication	If you selected pgoutput as the replication plugin, specify the publication name that is used by this plugin. <b>Note:</b> This field is not displayed if you selected wal2json as the replication plugin.

5. Under **Table Selection**, select one of the following options to indicate whether you want to select all tables in the specified schema or create table selection rules to define a subset of the source tables:

- **Select All**. Selects all tables in the schema for data replication.
- **Rule-based Selection**. Enables you to define rules to select only the tables you want to replicate. You also must be in this mode to assign actions to columns in any selected table.

Default is **Rule-based Selection**.

**Note:** If you switch to **Select All** and then switch back to **Rule-based Selection**, none of the rules that you previously defined are reinstated and all tables in the schema are listed. You can add new rules to filter the list.

6. If you selected **Rule-based Selection**, create rules to select the tables you want to replicate.

By default, the **Rules** list contains a single *Include* rule with a condition that specifies only the asterisk (\*) wildcard character. This rule selects all tables in the specified source schema.

To create a rule, you can either use the entry boxes under **Create Rule** or click the Add Rule (+) icon and enter the rule within the **Table Rule** list.

To create a rule under **Create Rule**:

- a. Select **Table Selection** as the general rule type.
- b. In the adjacent drop-down list, select **Include** or **Exclude** to create an inclusion or exclusion rule, respectively.
- c. In the condition field, enter a table name or a table-name mask that includes one or more wildcards to identify the source tables to include in or exclude from table selection. Use the following guidelines:
  - A mask can contain one or both of the following wildcards: an asterisk (\*) wildcard to represent one or more characters and a question mark (?) wildcard to represent a single character. A wildcard can occur multiple times in a mask value and can occur anywhere in the value.
  - The task wizard is case sensitive. Enter the table names or masks in the case with which the tables were defined.

- Do not include delimiters such as quotation marks or brackets, even if the source database uses them. For example, Oracle requires quotation marks around lowercase and mixed-case names to force the names to be stored in lowercase or mixed case. However, in the task wizard, you must enter the lowercase or mixed-case names without quotation marks.
  - If a table name includes special characters such as a backslash (\), asterisk(\*), dollar sign (\$), caret (^), or question mark (?) escape each special character with a backslash (\) when you enter the rule.
- d. Click **Add Rule**.

The rule appears in the **Rules** list.

vp\_RuleSelectTablePrvw

1 Definition 2 Source 3 Target 4 Schedule and Runtime Options

Source

Connection: mp\_Oracle184c\_DBM1\_uhl183r4

Schema: AUSGA

Table Selection

Select All Rule-based Selection

Create Rule

Table Selection Exclude Enter the condition Add Rule

Total Tables Selected: 0

Rules

Table Rule	Condition	Tables Affected
Include	*	
Exclude	*NULL*	

Preview Selection

The rules will be processed in the order in which they're listed, from top to bottom. Use the arrow icons to change the order. For an example of using multiple rules, see [“Example of rules for selecting source tables” on page 242](#).

**Tip:** Click the Refresh icon next to **Total Tables Selected** field to refresh the total number of tables that match all rules and to display the number of source tables that match each rule in the **Tables Affected** column.

- e. To preview the tables to be selected based on all rules, click **Preview Selection**.

The tables are listed on the **Selected Tables** tab. The list shows the table names and column count.

**Tip:** Click the Refresh icon next to the **Updated** date to refresh the total tables count and the list of tables based on the current rules. You can check the results of new rules in this manner. Click the Settings icon to control the line spacing in the list of tables, from Comfortable (most spacing) to Compact (least spacing).

- f. To refine the table selection, you can switch back to the **Rules** tab to edit the rules or define additional rules. Then preview the table selection again.

**Important:** Mass Ingestion Databases might exclude an unsupported type of table from processing even if this table matches the selection rules.

7. If you selected **Select All**, you can review the tables in the specified schema that will be replicated.

A list of all tables in the schema appears, which shows the number of columns in each table. The **Tables Selected** field displays the total number of tables selected. For example:

Click the Refresh icon to refresh the list to reflect any added or dropped tables in the schema in the source database. Click the Settings icon to change the line spacing in the list of tables.

**Note:** If you switch back to **Rule-based Selection**, the list of tables remains displayed but you can define rules to filter the list of tables.

8. To perform trim actions on columns in source tables that were selected based on rules, create column action rules.

**Note:** You cannot create column action rules for MongoDB sources.

To create a column action rule:

- a. Ensure that the **Rule-based Selection** option is selected.
- b. Under **Create Rule**, select **Column Action**.
- c. In the adjacent list, select one of the following action types:
  - **LTRIM**. Trims spaces to the left of character column values.
  - **RTRIM**. Trims spaces to the right of character column values.
  - **TRIM**. Trims spaces to the left of and to the right of character column values.
- d. In the condition field, enter a column name or a column name mask that includes one or more asterisk (\*) or question mark (?) wildcards. The value is matched against columns in the selected source tables to identify the columns to which the action applies.
- e. Click **Add Rule**.

The rule appears in the actions list, below the table rules list.

**Note:** You can define multiple rules for different action types or for the same action type with different conditions. The rules are processed in the order in which they're listed, from top to bottom. Use the arrow icons to change the order.

9. If you are defining an initial load task that has a Db2 for i, Db2 for LUW, Db2 for z/OS, Microsoft SQL Server, MySQL, Oracle, PostgreSQL, or Teradata source and want to include database views as sources, select the **Include Views** check box.

The views are then fetched and included in the **Total Tables Selected** or **Tables Selected** count. The views will also be included in the list of table names when you preview selections and when you download the list of table names.

10. If you are defining an incremental load or combined initial and incremental load task that has a Db2 for i, Db2 for z/OS, Microsoft SQL Server, Oracle, PostgreSQL, or SAP HANA source and one or more of the selected source tables are not enabled for change data capture, you can generate a script for enabling CDC and then run or download the script.

- a. In the **CDC Script** field, select one of the following options:

- **Enable CDC for all columns**. Enables CDC for all columns in the selected source tables. This option is the only valid option for a Db2 for i, Db2 for z/OS, PostgreSQL, SAP HANA, or SQL Server source.

**Note:** For source tables without a primary key, including any tables with unique indexes, CDC is enabled for all columns by default, regardless of which option is selected.
- **Enable CDC for primary key columns**. Enables CDC only for primary key columns in the selected source tables. Do not use this option for a Db2 for i, Db2 for z/OS, PostgreSQL, or SQL Server source or for any task that has a Google Big Query target.

The script enables CDC in the following ways, depending on the source type:

- For Db2 for i sources, the script enables journaling on the source tables.
- For Db2 for z/OS, the script sets DATA CAPTURE CHANGES for the source tables.
- For Microsoft SQL Server sources, the script runs the stored procedures sys.sp\_cdc\_enable\_db and sys.sp\_cdc\_enable\_table to enable CDC on the source database and tables.
- For Oracle sources, the script enables supplemental logging for all or primary key columns in the selected source tables to log additional information in the redo logs.
- For PostgreSQL sources, the script sets REPLICATION IDENTITY FULL on the selected source tables to write all column values to the WAL file.

- For SAP HANA sources, the script creates the required PKLOG, PROCESSED, and \_CDC shadow tables. The script also creates three triggers and a sequence for each selected source table.
- For Microsoft SQL Server sources, complete the following fields:
    - In the **Capture Filegroup** field, enter the name of the filegroup to be used for the change table that is created for the capture. If you leave this field empty, the change table is located in the default filegroup of the database.
    - In the **Gating Role** field, enter the name of the database role that is used to gate access to change data. If you leave this field empty, the database does not use the gating role.
  - To run the script, click **Execute**.

If you do not have a database role or privilege that allows you to run the script, click the Download icon to download the script. The script file name has the following format:

`cdc_script_taskname_number.txt`. Then ask your database administrator to run the script.

Make sure the script runs before you run the database ingestion task.

**Note:** If you change to the **CDC Script** option later and run the script again, the script first drops CDC for the original set of columns and then enables CDC for the current set of columns. If the PROCESSED and PKLOG tables already exist, they are omitted from the new script. If the shadow \_CDC table and triggers already exist for any table, the SQL statements for creating those objects are commented out in the new script.

- To create and download a list of the source tables that match the table selection criteria, perform the following substeps:
  - If you used rule-based table selection, in the **Table Names** list, select the type of selection rules that you want to use. Options are:
    - **Include Rules Only**
    - **Exclude Rules Only**
    - **Include And Exclude Rules**
  - To list the columns, regardless of which table selection method you used, select the **Include Columns** check box.
 

**Note:** This option is not available for MongoDB sources.
  - Click the Download icon.

A downloaded list that includes columns has the following format:

`status, schema_name, table_name, object_type, column_name, comment`

The following table describes the information that is displayed in the downloaded list:

Field	Description
status	Indicates whether Mass Ingestion Databases excludes the source table or column from processing because it has an unsupported type. Valid values are: <ul style="list-style-type: none"><li>- <b>E</b>. The object is excluded from processing by an Exclude rule.</li><li>- <b>I</b>. The object is included in processing.</li><li>- <b>X</b>. The object is excluded from processing because it is an unsupported type of object. For example, unsupported types of objects include columns with unsupported data types and tables that include only unsupported columns. The comment field provides detail on unsupported types.</li></ul>
schema_name	Specifies the name of the source schema.
table_name	Specifies the name of the source table.
object_type	Specifies the type of the source object. Valid values are: <ul style="list-style-type: none"><li>- <b>C</b>. Column.</li><li>- <b>T</b>. Table.</li></ul>
column_name	Specifies the name of the source column. This information appears only if you selected the <b>Columns</b> check box.
comment	Specifies the reason why a source object of an unsupported type is excluded from processing even though it matches the selection rules.

12. Under **Advanced**, set the advanced properties that are available for your source type and load type. All of the properties are optional.

Field	Source and Load Type	Description
Disable Flashback	Oracle - Initial loads	<p>Select this check box to disable Mass Ingestion Databases use of Oracle Flashback when fetching data from the database.</p> <p>The use of Oracle Flashback requires users to be granted the EXECUTE ON DBMS_FLASHBACK privilege, which is not necessary for initial loads.</p> <p>This check box is selected by default for new initial load tasks. For existing initial load tasks, this check box is cleared by default, which causes Oracle Flashback to remain enabled.</p>
Include LOBs	<p>Oracle and PostgreSQL - Initial loads with Amazon Redshift, Amazon S3, Google Big Query, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, and Snowflake targets</p> <p>SQL Server with the following load types:</p> <ul style="list-style-type: none"> <li>- Initial loads with Amazon Redshift, Amazon S3, Google Big Query, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, and Snowflake targets</li> <li>- Incremental, and combined initial and incremental loads with Snowflake targets</li> </ul>	<p>Select this check box if the source contains the large-object columns from which you want to replicate data to an Amazon Redshift, Amazon S3, Google Big Query, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, Microsoft Azure Synapse Analytics, or Snowflake target.</p> <p>Large-object data types are:</p> <ul style="list-style-type: none"> <li>- For Oracle: BLOB, CLOB, and NCLOB</li> <li>- For PostgreSQL: BYTEA, TEXT, and XML</li> <li>- For SQL Server: IMAGE, NTEXT, NVARCHAR(MAX), TEXT, VARBINARY(MAX), VARCHAR(MAX), and XML</li> </ul> <p>For initial loads:</p> <ul style="list-style-type: none"> <li>- BLOB, BYTEA, IMAGE, or VARBINARY(MAX) columns are truncated before being written to BINARY columns on the target. <ul style="list-style-type: none"> <li>- For Amazon S3, Google Cloud Storage, and Microsoft Azure Data Lake Storage Gen2, the data is truncated to 16777216 bytes.</li> <li>- For Amazon Redshift, the data is truncated to 1024000 bytes.</li> <li>- For Microsoft Azure Synapse Analytics, the data is truncated to 1000000 bytes.</li> <li>- For Google Big Query and Snowflake, the data is truncated to 8388608 bytes.</li> </ul> </li> <li>- CLOB, NCLOB, TEXT, XML, VARCHAR(MAX), NTEXT, NVARCHAR(MAX), VARCHAR(MAX), or XML columns are truncated before being written to VARCHAR columns on the target. <ul style="list-style-type: none"> <li>- For Amazon S3, Google Cloud Storage, Microsoft Azure Data Lake Storage Gen2, and Snowflake, the data is truncated to 16777216 bytes.</li> <li>- For Amazon Redshift, the data is truncated to 65535 bytes.</li> <li>- For Microsoft Azure Synapse Analytics, the data is truncated to 500000 bytes.</li> <li>- For Google Big Query, the data is truncated to 8388608 bytes.</li> </ul> </li> </ul> <p>For incremental loads and combined loads:</p> <ul style="list-style-type: none"> <li>- If large-object columns contain more than 8 KB of data, the data is truncated to 4000 bytes if stored inline or truncated to approximately 8000 bytes if stored out-of-line, before being written to a BINARY or VARCHAR column on the target.</li> </ul>

Field	Source and Load Type	Description
Enable Persistent Storage	All sources except PostgreSQL - Incremental and combined initial and incremental loads	<p>Select this check box to enable persistent storage of transaction data in a disk buffer so that the data can be consumed continually, even when the writing of data to the target is slow or delayed.</p> <p>Benefits of using persistent storage are faster consumption of the source transaction logs, less reliance on log archives or backups, and the ability to still access the data persisted in disk storage after restarting a database ingestion job.</p> <p><b>Note:</b> For PostgreSQL CDC sources, persistent storage is required and cannot be disabled.</p>



Field	Source and Load Type	Description
Restart Point for Incremental Load	All sources - Incremental and combined initial and incremental loads	<p>Set this field if you want to customize the position in the source logs from which the database ingestion job starts reading change records the first time it runs.</p> <p><b>Note:</b> <b>Earliest Available</b> and <b>Position</b> options are not supported for MongoDB sources.</p> <p>Options are:</p> <ul style="list-style-type: none"> <li>- <b>Earliest Available.</b> The earliest available position in the database log or structure where changes are stored. <ul style="list-style-type: none"> <li>- For Db2 for i, the start of the current journal.</li> <li>- For Db2 for z/OS, the earliest available record in the transaction log.</li> <li>- For Oracle, the earliest available record in the online redo log.</li> <li>- For PostgreSQL, the earliest available record in the replication slot.</li> <li>- For SAP HANA, the earliest available record in the PKLOG table.</li> <li>- For SQL Server, the earliest available record in the active transaction log.</li> </ul> </li> <li>- <b>Latest Available.</b> The latest available position in the database log or structure.</li> <li>- <b>Position.</b> A valid Oracle SCN or SQL Server LSN that Mass Ingestion uses to determine a position in the change stream from which to start retrieving change records. The SCN or LSN must be equal to or less than the current SCN or LSN. An invalid value will cause the job to fail. The default <b>Position</b> value is 0. <p>For Oracle, 0 causes reading to start from the default restart point, which is latest available point.</p> <p>For SQL Server, 0 causes reading to start from the earliest available point. A non-zero LSN that predates the beginning of the active transaction log causes data to be read from the CDC tables instead of from the transaction log.</p> </li> <li>- <b>Specific Date and Time.</b> A date and time, in the format MM/DD/YYYY hh:mm AM PM, that Mass Ingestion uses to determine the position in the change stream from which to start retrieving change records. Mass Ingestion retrieves only the changes that were started after this date and time. If you enter a date and time earlier than the earliest date and time in the available archived logs, the job will fail.</li> </ul> <p>The default is <b>Latest Available</b>.</p> <p><b>Notes:</b></p> <ul style="list-style-type: none"> <li>- This restart point option pertains only to the initial run of a job. Thereafter, if you resume a stopped or aborted job, the job begins propagating source data from where it last left off.</li> <li>- For combined initial and incremental load jobs, initial loading is not performed until the incremental processing of change data reaches the end of the current transaction log. For this reason, Informatica recommends that you select <b>Latest Available</b> as the start point.</li> </ul>

Field	Source and Load Type	Description
		<p>For Oracle combined initial and incremental load jobs, Oracle Flashback queries are used to get committed data that was current at a specific point in the change stream. Ensure that no source table is truncated during the initial load period. If truncation occurs, any DDL change performed during a flashback query causes the query to fail.</p> <p>For SQL Server, data changes are read from the active portion of the transaction log if the requested LSN is available there. If the LSN predates the active transaction log, the data changes are read from previously enabled CDC tables. Ensure that SQL Server CDC is enabled on the source tables.</p>
Fetch Size	MongoDB - Initial and incremental loads	For a MongoDB source, the number of records that a database ingestion job must read at a single time from the source. Valid values are 1 to 2147483647. The default is 5000.

13. Under **Custom Properties**, you can specify custom properties that Informatica provides to meet your special requirements. To add a property, in the **Create Property** fields, enter the property name and value. Then click **Add Property**.

Specify these properties only at the direction of Informatica Global Customer Support. Usually, these properties address unique environments or special processing needs. You can specify multiple properties, if necessary. A property name can contain only alphanumeric characters and the following special characters: periods (.), hyphens (-), and underscores (\_).

**Tip:** To delete a property, click the Delete icon at the right end of the property row in the list.

14. Click **Next**.

## Example of rules for selecting source tables

When you define a source for a database ingestion task, you can optionally define table selection rules to select a subset of the source tables in the specified schema. This simple example demonstrates how to use selection rules to select the tables you want.

Assume that 4,071 tables are in the source schema. You want to exclude the tables from which you do not need to replicate data.

Define the following rules in the order shown:

vp\_RuleSelectTablePrvw

< Back Next > Save

Table Selection [Clear all](#)

☐ Select All ☒ Rule-based Selection

▼ Create Rule ⓘ

Rule type: ▼ Action: ▼ Enter the condition: Add Rule

Total Tables Selected: ⓘ 3057 ↻

Rules [Preview Selection](#)

Table Rule	Condition	Tables Affected
Include	*	4071
Exclude	SYS*	112
Exclude	*TGT*	901
Exclude	*TEST	1

The rules are processed from top to bottom.

- Rule 1 includes all source tables in the schema.
- Rule 2 excludes the source tables that have names beginning with "SYS".
- Rule 3 excludes source tables that include "TGT" anywhere in their names.
- Rule 4 excludes source tables that have names ending with "TEST".

After clicking the Refresh icon, the **Total Tables Selected** field shows 3057, which indicates you filtered out 1,014 tables.

## Configuring the target

Configure the target on the **Target** page of the database ingestion task wizard.

1. In the **Connection** list, select a connection for the target type.

You must have previously defined the connection in Administrator for the runtime environment.

The list includes only the connection types that are valid for the load type selected on the **Definition** page. No connections are listed if you did not select a load type.

If you change the load type and the selected connection is no longer valid, a warning message is issued and the **Connection** field is cleared. You must select another connection that is valid for the updated load type.

**Note:** After you deploy the ingestion task, you cannot change the connection without first undeploying the associated ingestion job. You must then deploy the task again.

2. Under Target, configure the properties that pertain to your target type.

For descriptions of these properties, see the following topics:

- [“Amazon Redshift target properties” on page 246](#)
- [“Amazon S3 target properties” on page 246](#)
- [“Databricks Delta target properties” on page 250](#)
- [“Flat File target properties” on page 250](#)
- [“Google BigQuery target properties” on page 253](#)
- [“Google Cloud Storage target properties” on page 253](#)
- [“Kafka target properties” on page 257](#)
- [“Microsoft Azure Data Lake Storage target properties” on page 260](#)
- [“Microsoft Azure Synapse Analytics target properties” on page 264](#)
- [“Oracle target properties” on page 264](#)
- [“Snowflake Cloud Data Warehouse target properties” on page 265](#)

3. If you want to rename the target objects that are associated with the selected source tables, define table renaming rules.

For example, you can add a prefix such as TGT\_. For more information, see [“Rules for renaming tables on the target” on page 244](#).

4. If you want to override the default mappings of source data types to target data types, define data type rules.

This feature is available only for tasks that have an Oracle source and an SQL-based target type. For more information, see [“Rules for customizing data type mappings” on page 245](#).

5. Under **Custom Properties**, you can specify custom properties that Informatica provides to meet your special requirements. To add a property, in the **Create Property** fields, enter the property name and value. Then click **Add Property**.

Specify these properties only at the direction of Informatica Global Customer Support. Usually, these properties address unique environments or special processing needs. You can specify multiple properties, if necessary. A property name can contain only alphanumeric characters and the following special characters: periods (.), hyphens (-), and underscores (\_).

**Tip:** To delete a property, click the Delete icon button at the right end of the property row in the list.

6. Click **Next** if available, or click **Save**.

## Rules for renaming tables on the target

When you configure a target with an existing schema, you can optionally define rules for renaming the target tables that correspond to the selected source tables.

For target messaging systems, such as Apache Kafka, the rule renames the table name in the output messages.

To create a rule for renaming tables:

1. Under **Table Renaming Rules**, in the **Create Rule** fields, enter a source table name or a table name mask that includes one or more wildcards. Then enter the corresponding target table name or table name mask.

### Notes:

- For the source, you can enter only the asterisk (\*) wildcard to select all source tables that match the table selection criteria on the **Source** page. Or you can enter a specific source table name or a table-name mask that includes one or more of the following wildcards: an asterisk (\*) to represent one or more characters or a question mark (?) to represent a single character.
- To use a table-name mask with the wildcard character for the target, you must also use the wildcard character in the source. If you use a specific source table name with a target table mask that includes the wildcard character, the task deployment will fail.
- If a table name includes special characters, such as a backslash (\), asterisk(\*), dot (.), or question mark (?), escape each special character in the name with a backslash (\).

2. Click **Add Rule**.  
The rule appears in the rules list.

▼ Table Renaming Rules ⓘ

Create Rule:

Source Table	Target Table
*	PROD_*

You can define multiple table rules. The order of the rules does not matter with regard to how they are processed unless a table matches multiple rules. In this case, the last matching rule determines the name of the table.

To delete a rule, click the Delete icon at the right end of the rule row.

### Example:

Assume that you want to add the prefix "PROD\_" to the names of target tables that correspond to all selected source tables. Enter the following values:

- For the source, enter only the asterisk (\*) wildcard character to specify all of the selected source tables.

- For the target, enter PROD\_\* to add this prefix to the names of all target tables that match the source tables by name.

## Rules for customizing data type mappings

When you configure a target for a database ingestion task, you can optionally define data-type mapping rules to override the default mappings of source data types to target data types.

The default mappings are described in [“Default Data Type Mappings” on page 276](#).

This feature is supported for tasks that have an Oracle source and a target type that supports SQL, including Databricks Delta, Google BigQuery, Microsoft Azure Synapse Analytics, Oracle, and Snowflake.

For example, you can create a data-type rule that maps Oracle NUMBER columns that have no precision to Snowflake target NUMBER() columns that also have no precision, instead of using the default mapping to the Snowflake VARCHAR(255) data type.

To create a data-type mapping rule:

1. Expand **Data Type Rules**.
2. In the **Create Rule** fields, enter a source data type and the target data type that you want to map it to. In the *Source* field only, you can include the percent (%) wildcard to represent the data type precision, scale, or size, for example, NUMBER(%,4), NUMBER(8,%), or NUMBER(%). Use the wildcard to cover all source columns that have the same data type but use different precision, scale, or size values, instead of specifying each one individually. For example, enter FLOAT(%) to cover FLOAT(16), FLOAT(32), and FLOAT(84). You cannot enter the % wildcard in the target data type. A source data type that uses the % wildcard must map to a target data type that uses specific precision, scale, or size value. For example, you could map the source data type FLOAT(%) to a target data type specification such as NUMBER(38,10).
3. Click **Add Rule**.  
The rule appears in the list of rules.

▼ Data Type Rules ⓘ

Create Rule:

Source

Target

Add Rule

Source Data Type	Target Data Type
FLOAT(126)	CHAR(100)

To delete a rule, click the Delete icon at the right end of the rule row.

After you deploy a task with custom mapping rules, you cannot edit the rules until the task is undeployed.

### Notes:

- If you define multiple data-type rules for the same source data type with the same length or same precision and scale values, you will not be able to save the database ingestion task.
- If you define multiple data-type rules for the same source data type but use the % wildcard to represent the length or precision and scale value in one rule and a specific length or precision and scale value in the second rule, the rule that contains the specific value is processed first, before the rule with the % wildcard. For example, if you map the source data types FLOAT(84) and FLOAT(%), the FLOAT(84) rule is processed first and then the FLOAT(%) rule is processed to cover any other FLOAT source columns with different sizes.
- If a source data type requires a length or precision and scale value, make sure that you set the required attribute by using the % wildcard or a specific value, for example, VARCHAR(%) or VARCHAR(10).

- If you define an invalid mapping, an error message is written to the log. You can then correct the mapping error, with assistance from your DBA if necessary.
- For Oracle sources, you must use the data types that are returned by the following query for the source object:

```
select dbms_metadata.get_ddl('TABLE', 'YOUR_TABLE_NAME','TABLE_OWNER_NAME') from dual;
```

- Mass Ingestion Databases does not differentiate between Oracle TIMESTAMP(0) and DATE data types when you define data-type mapping rules. If you create a rule for TIMESTAMP(0), the database ingestion job will also apply this rule to DATE columns.
- Mass Ingestion Databases does not support the BYTE and CHAR semantics in data-type mappings rules.
- If a source data type has a default value, you must specify it in your rule. For example, you must use TIMESTAMP(6) instead of TIMESTAMP.

## Amazon Redshift target properties

When you define a database ingestion task that has an Amazon Redshift target, you must enter some target properties on the **Target** tab of the task wizard.

The following table describes the Amazon Redshift target properties that appear under **Target**:

Property	Description
Target Creation	The only available option is <b>Create Target Tables</b> , which generates the target tables based on the source tables. <b>Note:</b> After the target table is created, Mass Ingestion Databases intelligently handles the target tables on subsequent job runs. Mass Ingestion Databases might truncate or re-create the target tables depending on the specific circumstances.
Schema	Select the target schema in which Mass Ingestion Databases creates the target tables.
Bucket	Specifies the name of the Amazon S3 bucket that stores, organizes, and controls access to the data objects that you load to Amazon Redshift.
Directory	Specifies the subdirectory where Mass Ingestion Databases stores output files for this job.

## Amazon S3 target properties

When you define a database ingestion task that has an Amazon S3 target, you must enter some target properties on the **Target** tab of the task wizard.

Under **Target**, you can enter the following Amazon S3 target properties:

Property	Description
Output Format	Select the format of the output file. Options are: <ul style="list-style-type: none"> <li>- <b>CSV</b></li> <li>- <b>AVRO</b></li> <li>- <b>PARQUET</b></li> </ul> The default value is <b>CSV</b> . <b>Note:</b> Output files in CSV format use double-quotation marks (") as the delimiter for each field.
Add Headers to CSV File	If <b>CSV</b> is selected as the output format, select this check box to add a header with source column names to the output CSV file.

Property	Description
Avro Format	<p>If you selected <b>AVRO</b> as the output format, select the format of the Avro schema that will be created for each source table. Options are:</p> <ul style="list-style-type: none"> <li>- <b>Avro-Flat</b>. This Avro schema format lists all Avro fields in one record.</li> <li>- <b>Avro-Generic</b>. This Avro schema format lists all columns from a source table in a single array of Avro fields.</li> <li>- <b>Avro-Nested</b>. This Avro schema format organizes each type of information in a separate record.</li> </ul> <p>The default value is <b>Avro-Flat</b>.</p>
Avro Serialization Format	<p>If <b>AVRO</b> is selected as the output format, select the serialization format of the Avro output file. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Binary</b></li> <li>- <b>JSON</b></li> </ul> <p>The default value is <b>Binary</b>.</p>
Avro Schema Directory	<p>If <b>AVRO</b> is selected as the output format, specify the local directory where Mass Ingestion Databases stores Avro schema definitions for each source table. Schema definition files have the following naming pattern:</p> <p><i>schemaname_tablename.txt</i></p> <p><b>Note:</b> If this directory is not specified, no Avro schema definition file is produced.</p>
File Compression Type	<p>Select a file compression type for output files in CSV or AVRO output format. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Deflate</b></li> <li>- <b>Gzip</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Avro Compression Type	<p>If <b>AVRO</b> is selected as the output format, select an Avro compression type. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Bzip2</b></li> <li>- <b>Deflate</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Parquet Compression Type	<p>If the <b>PARQUET</b> output format is selected, you can select a compression type that is supported by Parquet. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Gzip</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Deflate Compression Level	<p>If <b>Deflate</b> is selected in the <b>Avro Compression Type</b> field, specify a compression level from 0 to 9. The default value is 0.</p>
Add Directory Tags	<p>For incremental load tasks, select this check box to add the "dt=" prefix to the names of apply cycle directories to be compatible with the naming convention for Hive partitioning. This check box is cleared by default.</p>

Property	Description
Task Target Directory	<p>For incremental load tasks, the root directory for the other directories that hold output data files, schema files, and CDC cycle contents and completed files. You can use it to specify a custom root directory for the task. If you enable the <b>Connection Directory as Parent</b> option, you can still optionally specify a task target directory to use with the parent directory specified in the connection properties.</p> <p>This field is required if the {TaskTargetDirectory} placeholder is specified in patterns for any of the following directory fields.</p>
Connection Directory as Parent	<p>For initial load and incremental load tasks, select this check box to use the directory value that is specified in the target connection properties as the parent directory for the custom directory paths specified in the task target properties. For initial load tasks, the parent directory is used in the <b>Data Directory</b> and <b>Schema Directory</b>. For incremental load tasks, the parent directory is used in the <b>Data Directory</b>, <b>Schema Directory</b>, <b>Cycle Completion Directory</b>, and <b>Cycle Contents Directory</b>.</p> <p>This check box is selected by default. If you clear it, for initial loads, define the full path to the output files in the <b>Data Directory</b> field. For incremental loads, optionally specify a root directory for the task in the <b>Task Target Directory</b>.</p>
Data Directory	<p><i>For initial load tasks</i>, define a directory structure for the directories where Mass Ingestion Databases stores output data files and optionally stores the schema. To define directory pattern, you can use the following types of entries:</p> <ul style="list-style-type: none"> <li>- The placeholders {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The {Timestamp} values are in the format yyyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the initial load job starts to transfer data to the target.</li> <li>- Specific directory names.</li> <li>- The toUpper() and toLower() functions, which force the values for an associated (<i>placeholder</i>) to uppercase or lowercase.</li> </ul> <p><b>Note:</b> Placeholder values are not case sensitive.</p> <p>Examples:</p> <pre>myDir1/{SchemaName}/{TableName} myDir1/myDir2/{SchemaName}/{YYYY}/{MM}/{TableName}_{Timestamp} myDir1/{toLower(SchemaName)}/{TableName}_{Timestamp}</pre> <p>The default directory pattern is {TableName}_{Timestamp}.</p> <p><i>For incremental load tasks</i>, define a custom path to the subdirectory that contains the cdc-data data files. To define the directory pattern, you can use the following types of entries:</p> <ul style="list-style-type: none"> <li>- The placeholders {TaskTargetDirectory}, {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The {Timestamp} values are in the format yyyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the CDC cycle started.</li> </ul> <p>If you include the toUpper or toLower function, put the placeholder name in parentheses and enclose the both the function and placeholder in curly brackets, as shown in the preceding example.</p> <ul style="list-style-type: none"> <li>- Specific directory names.</li> </ul> <p>The default directory pattern is {TaskTargetDirectory}/cdc-data/{TableName}/data</p> <p><b>Note:</b> For Amazon S3, Flat File, and Microsoft Azure Data Lake Storage Gen2 targets, Mass Ingestion Databases uses the directory specified in the target connection properties as the root for the data directory path when <b>Connection Directory as Parent</b> is selected. For Google Cloud Storage targets, Mass Ingestion Databases uses the <b>Bucket</b> name that you specify in the target properties for the ingestion task.</p>



Property	Description
Schema Directory	<p>For initial load and incremental load tasks, you can specify a custom directory in which to store the schema file if you want to store it in a directory other than the default directory. For initial loads, previously used values if available are shown in a drop-down list for your convenience. This field is optional.</p> <p>For initial loads, the schema is stored in the data directory by default. For incremental loads, the default directory for the schema file is <code>{TaskTargetDirectory}/cdc-data/{TableName}/schema</code></p> <p>You can use the same placeholders as for the <b>Data Directory</b> field. Ensure that you enclose placeholders with curly brackets <code>{ }</code>.</p> <p>If you include the <code>toUpper</code> or <code>toLower</code> function, put the placeholder name in parentheses and enclose the both the function and placeholder in curly brackets, for example:  <code>{toLower(SchemaName)}</code></p> <p><b>Note:</b> Schema is written only to output data files in CSV format. Data files in Parquet and Avro formats contain their own embedded schema.</p>
Cycle Completion Directory	For incremental load tasks, the path to the directory that contains the cdc-cycle completed file. Default is <code>{TaskTargetDirectory}/cdc-cycle/completed</code> .
Cycle Contents Directory	For incremental load tasks, the path to the directory that contains the cdc-cycle contents files. Default is <code>{TaskTargetDirectory}/cdc-cycle/contents</code> .
Use Cycle Partitioning for Data Directory	<p>For incremental load tasks, causes a timestamp subdirectory to be created for each CDC cycle, under each data directory.</p> <p>If this option is not selected, individual data files are written to the same directory without a timestamp, unless you define an alternative directory structure.</p>
Use Cycle Partitioning for Summary Directories	For incremental load tasks, causes a timestamp subdirectory to be created for each CDC cycle, under the summary contents and completed subdirectories.
List Individual Files in Contents	<p>For incremental load tasks, lists individual data files under the contents subdirectory.</p> <p>If <b>Use Cycle Partitioning for Summary Directories</b> is cleared, this option is selected by default. All of the individual files are listed in the contents subdirectory unless you can configure custom subdirectories by using the placeholders, such as for timestamp or date.</p> <p>If <b>Use Cycle Partitioning for Data Directory</b> is selected, you can still optionally select this check box to list individual files and group them by CDC cycle.</p>

Under **Advanced**, you can enter the following Amazon S3 advanced target properties, which primarily apply to incremental loads:

Field	Description
Add Operation Type	<p>Select this check box to add a metadata column that includes the source SQL operation type in the output that the job propagates to the target.</p> <p>For incremental loads, the job writes "I" for insert, "U" for update, or "D" for delete. For initial loads, the job always writes "I" for insert.</p> <p>By default, this check box is cleared.</p>
Add Operation Time	<p>Select this check box to add a metadata column that includes the source SQL operation time in the output that the job propagates to the target.</p> <p>For initial loads, the job always writes the current date and time.</p> <p>By default, this check box is cleared.</p>

Field	Description
Add Operation Owner	Select this check box to add a metadata column that includes the owner of the source SQL operation in the output that the job propagates to the target. For initial loads, the job always writes "INFA" as the owner. By default, this check box is cleared. <b>Note:</b> This property is not available for jobs that have a PostgreSQL source.
Add Operation Transaction Id	Select this check box to add a metadata column that includes the source transaction ID in the output that the job propagates to the target for SQL operations. For initial loads, the job always writes "1" as the ID. By default, this check box is cleared.
Add Before Images	Select this check box to include UNDO data in the output that an incremental load job writes to the target. For initial loads, the job writes nulls. By default, this check box is cleared.

## Databricks Delta target properties

When you define a database ingestion task that has a Databricks Delta target, you must enter some target properties on the **Target** tab of the task wizard.

The following table describes the Databricks Delta target properties that appear under **Target**:

Property	Description
Target Creation	The only available option is <b>Create Target Tables</b> , which generates the target tables based on the source tables. <b>Note:</b> After the target table is created, Mass Ingestion Databases intelligently handles the target tables on subsequent job runs. Mass Ingestion Databases might truncate or re-create the target tables depending on the specific circumstances.
Schema	Select the target schema in which Mass Ingestion Databases creates the target tables.
Directory	Specifies the subdirectory where Mass Ingestion Databases stores output files for this job.

## Flat File target properties

When you define a database ingestion task, you must enter some properties for your Flat File target on the **Target** page of the task wizard.

**Note:** For flat file targets, these properties apply to initial load jobs only.

Under **Target**, you can enter the following Flat File target properties:

Property	Description
Output Format	<p>Select the format of the output file. Options are:</p> <ul style="list-style-type: none"> <li>- <b>CSV</b></li> <li>- <b>AVRO</b></li> </ul> <p>The default value is <b>CSV</b>.</p> <p><b>Note:</b> Output files in CSV format use double-quotation marks (") as the delimiter for each field.</p>
Add Headers to CSV File	<p>If <b>CSV</b> is selected as the output format, select this check box to add a header with source column names to the output CSV file.</p>
Avro Format	<p>If you selected <b>AVRO</b> as the output format, select the format of the Avro schema that will be created for each source table. Options are:</p> <ul style="list-style-type: none"> <li>- <b>Avro-Flat</b>. This Avro schema format lists all Avro fields in one record.</li> <li>- <b>Avro-Generic</b>. This Avro schema format lists all columns from a source table in a single array of Avro fields.</li> <li>- <b>Avro-Nested</b>. This Avro schema format organizes each type of information in a separate record.</li> </ul> <p>The default value is <b>Avro-Flat</b>.</p>
Avro Serialization Format	<p>If <b>AVRO</b> is selected as the output format, select the serialization format of the Avro output file. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Binary</b></li> <li>- <b>JSON</b></li> </ul> <p>The default value is <b>Binary</b>.</p>
Avro Schema Directory	<p>If <b>AVRO</b> is selected as the output format, specify the local directory where Mass Ingestion Databases stores Avro schema definitions for each source table. Schema definition files have the following naming pattern:</p> <p><i>schemaname_tablename.txt</i></p> <p><b>Note:</b> If this directory is not specified, no Avro schema definition file is produced.</p>
File Compression Type	<p>Select a file compression type for output files in CSV or AVRO output format. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Deflate</b></li> <li>- <b>Gzip</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Avro Compression Type	<p>If <b>AVRO</b> is selected as the output format, select an Avro compression type. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Bzip2</b></li> <li>- <b>Deflate</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Deflate Compression Level	<p>If <b>Deflate</b> is selected in the <b>Avro Compression Type</b> field, specify a compression level from 0 to 9. The default value is 0.</p>

Property	Description
Data Directory	<p>For initial load tasks, define a directory structure for the directories where Mass Ingestion Databases stores output data files and optionally stores the schema. To define directory pattern, you can use the following types of entries:</p> <ul style="list-style-type: none"> <li>- The placeholders {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The {Timestamp} values are in the format yyyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the initial load job starts to transfer data to the target.</li> <li>- Specific directory names.</li> <li>- The toUpper() and toLower() functions, which force the values for an associated (<i>placeholder</i>) to uppercase or lowercase.</li> </ul> <p><b>Note:</b> Placeholder values are not case sensitive.</p> <p>Examples:</p> <pre>myDir1/{SchemaName}/{TableName} myDir1/myDir2/{SchemaName}/{YYYY}/{MM}/{TableName}_{Timestamp} myDir1/{toLower(SchemaName)}/{TableName}_{Timestamp}</pre> <p>The default directory pattern is {TableName}_{Timestamp}.</p> <p><b>Note:</b> For Amazon S3, Flat File, and Microsoft Azure Data Lake Storage Gen2 targets, Mass Ingestion Databases uses the directory specified in the target connection properties as the root for the data directory path when <b>Connection Directory as Parent</b> is selected. For Google Cloud Storage targets, Mass Ingestion Databases uses the <b>Bucket</b> name that you specify in the target properties for the ingestion task.</p>
Connection Directory as Parent	<p>For initial load tasks, select this check box to use the directory value that is specified in the target connection properties as the parent directory for the custom directory paths specified in the task target properties. The parent directory is used in the <b>Data Directory</b> and <b>Schema Directory</b>.</p>
Schema Directory	<p>For initial load tasks, you can specify a custom directory in which to store the schema file if you want to store it in a directory other than the default directory. This field is optional.</p> <p>The schema is stored in the data directory by default. For incremental loads, the default directory for the schema file is {TaskTargetDirectory}/data/{TableName}/schema.</p> <p>You can use the same placeholders as for the <b>Data Directory</b> field. Ensure the placeholders are enclosed in curly brackets {}.</p>

Under **Advanced**, the following table describes the Flat File target advanced properties that appear:

Field	Description
Add Operation Type	<p>Select this check box to add a metadata column that includes the source SQL operation type in the output that the job propagates to the target.</p> <p>For incremental loads, the job writes "I" for insert, "U" for update, or "D" for delete. For initial loads, the job always writes "I" for insert.</p> <p>By default, this check box is cleared.</p>
Add Operation Time	<p>Select this check box to add a metadata column that includes the source SQL operation time in the output that the job propagates to the target.</p> <p>For initial loads, the job always writes the current date and time.</p> <p>By default, this check box is cleared.</p>
Add Operation Owner	<p>Select this check box to add a metadata column that includes the owner of the source SQL operation in the output that the job propagates to the target.</p> <p>For initial loads, the job always writes "INFA" as the owner.</p> <p>By default, this check box is cleared.</p> <p><b>Note:</b> This property is not available for jobs that have a PostgreSQL source.</p>

Field	Description
Add Operation Transaction Id	Select this check box to add a metadata column that includes the source transaction ID in the output that the job propagates to the target for SQL operations. For initial loads, the job always writes "1" as the ID. By default, this check box is cleared.
Add Before Images	Select this check box to include UNDO data in the output that an incremental load job writes to the target. For initial loads, the job writes nulls. By default, this check box is cleared.

## Google BigQuery target properties

When you define a database ingestion task that has a Google BigQuery target, you must enter some target properties on the **Target** tab of the task wizard.

The following table describes the Google BigQuery target properties that appear under **Target**:

Property	Description
Target Creation	The only available option is <b>Create Target Tables</b> , which generates the target tables based on the source tables. <b>Note:</b> After the target table is created, Mass Ingestion Databases intelligently handles the target tables on subsequent job runs. Mass Ingestion Databases might truncate or re-create the target tables depending on the specific circumstances.
Schema	Select the target schema in which Mass Ingestion Databases creates the target tables.
Bucket	Specifies the name of an existing bucket container that stores, organizes, and controls access to the data objects that you load to Google Cloud Storage.
Directory	Specifies the virtual directory for the Google Cloud Storage target objects that contain the data.

## Google Cloud Storage target properties

When you define a database ingestion task that has a Google Cloud Storage target, you must enter some target properties on the **Target** tab of the task wizard.

Under **Target**, you can enter the following Google Cloud Storage target properties:

Property	Description
Output Format	Select the format of the output file. Options are: - <b>CSV</b> - <b>AVRO</b> - <b>PARQUET</b> The default value is <b>CSV</b> . <b>Note:</b> Output files in CSV format use double-quotation marks (") as the delimiter for each field.
Add Headers to CSV File	If <b>CSV</b> is selected as the output format, select this check box to add a header with source column names to the output CSV file.

Property	Description
Avro Format	<p>If you selected <b>AVRO</b> as the output format, select the format of the Avro schema that will be created for each source table. Options are:</p> <ul style="list-style-type: none"> <li>- <b>Avro-Flat</b>. This Avro schema format lists all Avro fields in one record.</li> <li>- <b>Avro-Generic</b>. This Avro schema format lists all columns from a source table in a single array of Avro fields.</li> <li>- <b>Avro-Nested</b>. This Avro schema format organizes each type of information in a separate record.</li> </ul> <p>The default value is <b>Avro-Flat</b>.</p>
Avro Serialization Format	<p>If <b>AVRO</b> is selected as the output format, select the serialization format of the Avro output file. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Binary</b></li> <li>- <b>JSON</b></li> </ul> <p>The default value is <b>Binary</b>.</p>
Avro Schema Directory	<p>If <b>AVRO</b> is selected as the output format, specify the local directory where Mass Ingestion Databases stores Avro schema definitions for each source table. Schema definition files have the following naming pattern:</p> <p><i>schemaname_tablename.txt</i></p> <p><b>Note:</b> If this directory is not specified, no Avro schema definition file is produced.</p>
File Compression Type	<p>Select a file compression type for output files in CSV or AVRO output format. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Deflate</b></li> <li>- <b>Gzip</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Avro Compression Type	<p>If <b>AVRO</b> is selected as the output format, select an Avro compression type. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Bzip2</b></li> <li>- <b>Deflate</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Parquet Compression Type	<p>If the <b>PARQUET</b> output format is selected, you can select a compression type that is supported by Parquet. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Gzip</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Deflate Compression Level	<p>If <b>Deflate</b> is selected in the <b>Avro Compression Type</b> field, specify a compression level from 0 to 9. The default value is 0.</p>
Add Directory Tags	<p>For incremental load tasks, select this check box to add the "dt=" prefix to the names of apply cycle directories to be compatible with the naming convention for Hive partitioning. This check box is cleared by default.</p>
Bucket	<p>Specifies the name of an existing bucket container that stores, organizes, and controls access to the data objects that you load to Google Cloud Storage.</p>

Property	Description
Task Target Directory	<p>For incremental load tasks, the root directory for the other directories that hold output data files, schema files, and CDC cycle contents and completed files. You can use it to specify a custom root directory for the task. If you enable the <b>Connection Directory as Parent</b> option, you can still optionally specify a task target directory to use with the parent directory specified in the connection properties.</p> <p>This field is required if the {TaskTargetDirectory} placeholder is specified in patterns for any of the following directory fields.</p>
Data Directory	<p><i>For initial load tasks</i>, define a directory structure for the directories where Mass Ingestion Databases stores output data files and optionally stores the schema. To define directory pattern, you can use the following types of entries:</p> <ul style="list-style-type: none"> <li>- The placeholders {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The {Timestamp} values are in the format yyyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the initial load job starts to transfer data to the target.</li> <li>- Specific directory names.</li> <li>- The toUpper() and toLower() functions, which force the values for an associated (<i>placeholder</i>) to uppercase or lowercase.</li> </ul> <p><b>Note:</b> Placeholder values are not case sensitive.</p> <p>Examples:</p> <pre>myDir1/{SchemaName}/{TableName} myDir1/myDir2/{SchemaName}/{YYYY}/{MM}/{TableName}_{Timestamp} myDir1/{toLower(SchemaName)}/{TableName}_{Timestamp}</pre> <p>The default directory pattern is {TableName}_{Timestamp}.</p> <p><i>For incremental load tasks</i>, define a custom path to the subdirectory that contains the cdc-data data files. To define the directory pattern, you can use the following types of entries:</p> <ul style="list-style-type: none"> <li>- The placeholders {TaskTargetDirectory}, {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The {Timestamp} values are in the format yyyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the CDC cycle started.</li> </ul> <p>If you include the toUpper or toLower function, put the placeholder name in parentheses and enclose the both the function and placeholder in curly brackets, as shown in the preceding example.</p> <ul style="list-style-type: none"> <li>- Specific directory names.</li> </ul> <p>The default directory pattern is {TaskTargetDirectory}/cdc-data/{TableName}/data</p> <p><b>Note:</b> For Amazon S3, Flat File, and Microsoft Azure Data Lake Storage Gen2 targets, Mass Ingestion Databases uses the directory specified in the target connection properties as the root for the data directory path when <b>Connection Directory as Parent</b> is selected. For Google Cloud Storage targets, Mass Ingestion Databases uses the <b>Bucket</b> name that you specify in the target properties for the ingestion task.</p>
Schema Directory	<p>For initial load and incremental load tasks, you can specify a custom directory in which to store the schema file if you want to store it in a directory other than the default directory. For initial loads, previously used values if available are shown in a drop-down list for your convenience. This field is optional.</p> <p>For initial loads, the schema is stored in the data directory by default. For incremental loads, the default directory for the schema file is {TaskTargetDirectory}/cdc-data/{TableName}/schema</p> <p>You can use the same placeholders as for the <b>Data Directory</b> field. Ensure that you enclose placeholders with curly brackets {}.</p> <p>If you include the toUpper or toLower function, put the placeholder name in parentheses and enclose the both the function and placeholder in curly brackets, for example:</p> <pre>{toLower(SchemaName)}</pre> <p><b>Note:</b> Schema is written only to output data files in CSV format. Data files in Parquet and Avro formats contain their own embedded schema.</p>

Property	Description
Cycle Completion Directory	For incremental load tasks, the path to the directory that contains the cdc-cycle completed file. Default is <code>{TaskTargetDirectory}/cdc-cycle/completed</code> .
Cycle Contents Directory	For incremental load tasks, the path to the directory that contains the cdc-cycle contents files. Default is <code>{TaskTargetDirectory}/cdc-cycle/contents</code> .
Use Cycle Partitioning for Data Directory	For incremental load tasks, causes a timestamp subdirectory to be created for each CDC cycle, under each data directory.  If this option is not selected, individual data files are written to the same directory without a timestamp, unless you define an alternative directory structure.
Use Cycle Partitioning for Summary Directories	For incremental load tasks, causes a timestamp subdirectory to be created for each CDC cycle, under the summary contents and completed subdirectories.
List Individual Files in Contents	For incremental load tasks, lists individual data files under the contents subdirectory.  If <b>Use Cycle Partitioning for Summary Directories</b> is cleared, this option is selected by default. All of the individual files are listed in the contents subdirectory unless you can configure custom subdirectories by using the placeholders, such as for timestamp or date.  If <b>Use Cycle Partitioning for Data Directory</b> is selected, you can still optionally select this check box to list individual files and group them by CDC cycle.

Under **Advanced**, you can enter the following Google Cloud Storage advanced target properties, which are primarily for incremental load jobs:

Field	Description
Add Operation Type	Select this check box to add a metadata column that includes the source SQL operation type in the output that the job propagates to the target.  For incremental loads, the job writes "I" for insert, "U" for update, or "D" for delete. For initial loads, the job always writes "I" for insert.  By default, this check box is cleared.
Add Operation Time	Select this check box to add a metadata column that includes the source SQL operation time in the output that the job propagates to the target.  For initial loads, the job always writes the current date and time.  By default, this check box is cleared.
Add Operation Owner	Select this check box to add a metadata column that includes the owner of the source SQL operation in the output that the job propagates to the target.  For initial loads, the job always writes "INFA" as the owner.  By default, this check box is cleared. <b>Note:</b> This property is not available for jobs that have a PostgreSQL source.



Field	Description
Add Operation Transaction Id	<p>Select this check box to add a metadata column that includes the source transaction ID in the output that the job propagates to the target for SQL operations.</p> <p>For initial loads, the job always writes "1" as the ID.</p> <p>By default, this check box is cleared.</p>
Add Before Images	<p>Select this check box to include UNDO data in the output that an incremental load job writes to the target.</p> <p>For initial loads, the job writes nulls.</p> <p>By default, this check box is cleared.</p>

## Kafka target properties

When you define a database ingestion task, you must enter some properties for your Kafka target on the **Target** page of the task wizard.

These properties apply to incremental load operations only.

The following table describes the Kafka target properties that appear under **Target**:

Property	Description
Use Table Name as Topic Name	<p>Indicates whether Mass Ingestion Databases writes messages that contain source data to separate topics, one for each source table, or writes all messages to a single topic.</p> <p>Select this check box to write messages to separate table-specific topics. The topic names match the source table names, unless you add the source schema name, a prefix, or a suffix in the <b>Include Schema Name</b>, <b>Table Prefix</b>, or <b>Table Suffix</b> properties.</p> <p>By default, this check box is cleared. With the default setting, you must specify the name of the single topic to which all messages are written in the <b>Topic Name</b> property.</p>
Include Schema Name	<p>When <b>Use Table Name as Topic Name</b> is selected, this check box appears and is selected by default. This setting adds the source schema name in the table-specific topic names. The topic names then have the format <i>schemaname.tablename</i>.</p> <p>If you do <i>not</i> want to include the schema name, clear this check box.</p>
Table Prefix	<p>When <b>Use Table Name as Topic Name</b> is selected, this property appears so that you can optionally enter a prefix to add to the table-specific topic names. For example, if you specify <i>myprefix_</i>, the topic names have the format <i>myprefix.tablename</i>. If you omit the underscore ( <i>_</i> ) after the prefix, the prefix is prepended to the table name.</p>
Table Suffix	<p>When <b>Use Table Name as Topic Name</b> is selected, this property appears so that you can optionally enter a suffix to add to the table-specific topic names. For example, if you specify <i>_mysuffix</i>, the topic names have the format <i>tablename_mysuffix</i>. If you omit the underscore ( <i>_</i> ) before the suffix, the suffix is appended to the table name.</p>
Topic Name	<p>If you do <i>not</i> select <b>Use table name as topic name</b>, you must enter the name of the single Kafka topic to which all messages that contain source data will be written.</p>

Property	Description
Output Format	<p>Select the format of the output file. Options are:</p> <ul style="list-style-type: none"> <li>- <b>CSV</b></li> <li>- <b>AVRO</b></li> <li>- <b>JSON</b></li> </ul> <p>The default value is <b>CSV</b>.</p> <p><b>Note:</b> Output files in CSV format use double-quotation marks (") as the delimiter for each field.</p> <p>If your Kafka target uses Confluent Schema Registry to store schemas for incremental load jobs, you must select <b>AVRO</b> as the format.</p>
JSON Format	<p>If <b>JSON</b> is selected as the output format, select the level of detail of the output. Options are:</p> <ul style="list-style-type: none"> <li>- <b>Concise</b>. This format records only the most relevant data in the output, such as the operation type and the column names and values.</li> <li>- <b>Verbose</b>. This format records detailed information, such as the table name and column types.</li> </ul>
Avro Format	<p>If you selected <b>AVRO</b> as the output format, select the format of the Avro schema that will be created for each source table. Options are:</p> <ul style="list-style-type: none"> <li>- <b>Avro-Flat</b>. This Avro schema format lists all Avro fields in one record.</li> <li>- <b>Avro-Generic</b>. This Avro schema format lists all columns from a source table in a single array of Avro fields.</li> <li>- <b>Avro-Nested</b>. This Avro schema format organizes each type of information in a separate record.</li> </ul> <p>The default value is <b>Avro-Flat</b>.</p>
Avro Serialization Format	<p>If <b>AVRO</b> is selected as the output format, select the serialization format of the Avro output file. Options are:</p> <ul style="list-style-type: none"> <li>- <b>Binary</b></li> <li>- <b>JSON</b></li> <li>- <b>None</b></li> </ul> <p>The default value is <b>Binary</b>.</p> <p>If you have a Confluent Kafka target that uses Confluent Schema Registry to store schemas, select <b>None</b>. Otherwise, Confluent Schema Registry does not register the schema. Do not select <b>None</b> if you are not using Confluent Schema Registry.</p>
Avro Schema Directory	<p>If <b>AVRO</b> is selected as the output format, specify the local directory where Mass Ingestion Databases stores Avro schema definitions for each source table. Schema definition files have the following naming pattern:</p> <p><i>schemaname_tablename.txt</i></p> <p><b>Note:</b> If this directory is not specified, no Avro schema definition file is produced.</p> <p>If a source schema change is expected to alter the target, the Avro schema definition file is regenerated with a unique name that includes a timestamp, in the following format:</p> <p><i>schemaname_tablename_YYYYMMDDhhmmss.txt</i></p> <p>This unique naming pattern ensures that older schema definition files are preserved for audit purposes.</p>
Avro Compression Type	<p>If <b>AVRO</b> is selected as the output format, select an Avro compression type. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Bzip2</b></li> <li>- <b>Deflate</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Deflate Compression Level	<p>If <b>Deflate</b> is selected in the <b>Avro Compression Type</b> field, specify a compression level from 0 to 9. The default value is 0.</p>

The following table describes the advanced Kafka target properties that appear under **Advanced**:

Property	Description
Add Operation Type	Select this check box to add a metadata column that includes the source SQL operation type in the output that the job propagates to the target. For incremental loads, the job writes "I" for insert, "U" for update, or "D" for delete. For initial loads, the job always writes "I" for insert. By default, this check box is cleared.
Add Operation Time	Select this check box to add a metadata column that includes the source SQL operation time in the output that the job propagates to the target. For initial loads, the job always writes the current date and time. By default, this check box is cleared.
Add Operation Owner	Select this check box to add a metadata column that includes the owner of the source SQL operation in the output that the job propagates to the target. For initial loads, the job always writes "INFA" as the owner. By default, this check box is cleared. <b>Note:</b> This property is not available for jobs that have a PostgreSQL source.
Add Operation Transaction Id	Select this check box to add a metadata column that includes the source transaction ID in the output that the job propagates to the target for SQL operations. For initial loads, the job always writes "1" as the ID. By default, this check box is cleared.
Add Before Images	Select this check box to include UNDO data in the output that an incremental load job writes to the target. For initial loads, the job writes nulls. By default, this check box is cleared.

Property	Description
Async Write	<p>Controls whether to use synchronous delivery of messages to Kafka.</p> <ul style="list-style-type: none"> <li>- Clear this check box to use synchronous delivery. Kafka must acknowledge each message as received before Mass Ingestion Databases sends the next message. In this mode, Kafka is unlikely to receive duplicate messages. However, performance might be slower.</li> <li>- Select this check box to use asynchronous delivery. Mass Ingestion Databases sends messages as soon as possible, without regard for the order in which the changes were retrieved from the source.</li> </ul> <p>By default, this check box is selected.</p>
Producer Configuration Properties	<p>Specify a comma-separated list of <i>key=value</i> pairs to enter Kafka producer properties for Apache Kafka, Confluent Kafka, Amazon Managed Streaming for Apache Kafka (MSK), or Kafka-enabled Azure Event Hubs targets.</p> <p>If you have a Confluent target that uses Confluent Schema Registry to store schemas, you must specify the following properties:</p> <pre>schema.registry.url=url, key.serializer=org.apache.kafka.common.serialization.StringSerializer, value.serializer=io.confluent.kafka.serializers.KafkaAvroSerializer</pre> <p>You can specify Kafka producer properties in either this field or in the <b>Additional Connection Properties</b> field in the Kafka connection.</p> <p>If you enter the producer properties in this field, the properties pertain to the database ingestion jobs associated with this task only. If you enter the producer properties for the connection, the properties pertain to jobs for all tasks that use the connection definition, unless you override the connection-level properties for specific tasks by also specifying properties in the <b>Producer Configuration Properties</b> field.</p> <p>For information about Kafka producer properties, see the Apache Kafka, Confluent Kafka, Amazon MSK, or Azure Event Hubs documentation.</p>

## Microsoft Azure Data Lake Storage target properties

When you define a database ingestion task that has a Microsoft Azure Data Lake Storage target, you must enter some target properties on the **Target** page of the task wizard.

Under **Target**, you can enter the following Microsoft Azure Data Lake Storage target properties:

Property	Description
Output Format	<p>Select the format of the output file. Options are:</p> <ul style="list-style-type: none"> <li>- <b>CSV</b></li> <li>- <b>AVRO</b></li> <li>- <b>PARQUET</b></li> </ul> <p>The default value is <b>CSV</b>.</p> <p><b>Note:</b> Output files in CSV format use double-quotation marks (") as the delimiter for each field.</p>
Add Headers to CSV File	<p>If <b>CSV</b> is selected as the output format, select this check box to add a header with source column names to the output CSV file.</p>

Property	Description
Avro Format	<p>If you selected <b>AVRO</b> as the output format, select the format of the Avro schema that will be created for each source table. Options are:</p> <ul style="list-style-type: none"> <li>- <b>Avro-Flat</b>. This Avro schema format lists all Avro fields in one record.</li> <li>- <b>Avro-Generic</b>. This Avro schema format lists all columns from a source table in a single array of Avro fields.</li> <li>- <b>Avro-Nested</b>. This Avro schema format organizes each type of information in a separate record.</li> </ul> <p>The default value is <b>Avro-Flat</b>.</p>
Avro Serialization Format	<p>If <b>AVRO</b> is selected as the output format, select the serialization format of the Avro output file. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Binary</b></li> <li>- <b>JSON</b></li> </ul> <p>The default value is <b>Binary</b>.</p>
Avro Schema Directory	<p>If <b>AVRO</b> is selected as the output format, specify the local directory where Mass Ingestion Databases stores Avro schema definitions for each source table. Schema definition files have the following naming pattern:</p> <p><i>schemaname_tablename.txt</i></p> <p><b>Note:</b> If this directory is not specified, no Avro schema definition file is produced.</p>
File Compression Type	<p>Select a file compression type for output files in CSV or AVRO output format. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Deflate</b></li> <li>- <b>Gzip</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Avro Compression Type	<p>If <b>AVRO</b> is selected as the output format, select an Avro compression type. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Bzip2</b></li> <li>- <b>Deflate</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Parquet Compression Type	<p>If the <b>PARQUET</b> output format is selected, you can select a compression type that is supported by Parquet. Options are:</p> <ul style="list-style-type: none"> <li>- <b>None</b></li> <li>- <b>Gzip</b></li> <li>- <b>Snappy</b></li> </ul> <p>The default value is <b>None</b>, which means no compression is used.</p>
Deflate Compression Level	<p>If <b>Deflate</b> is selected in the <b>Avro Compression Type</b> field, specify a compression level from 0 to 9. The default value is 0.</p>
Add Directory Tags	<p>For incremental load tasks, select this check box to add the "dt=" prefix to the names of apply cycle directories to be compatible with the naming convention for Hive partitioning. This check box is cleared by default.</p>

Property	Description
Task Target Directory	<p>For incremental load tasks, the root directory for the other directories that hold output data files, schema files, and CDC cycle contents and completed files. You can use it to specify a custom root directory for the task. If you enable the <b>Connection Directory as Parent</b> option, you can still optionally specify a task target directory to use with the parent directory specified in the connection properties.</p> <p>This field is required if the {TaskTargetDirectory} placeholder is specified in patterns for any of the following directory fields.</p>
Connection Directory as Parent	<p>For initial load and incremental load tasks, select this check box to use the directory value that is specified in the target connection properties as the parent directory for the custom directory paths specified in the task target properties. For initial load tasks, the parent directory is used in the <b>Data Directory</b> and <b>Schema Directory</b>. For incremental load tasks, the parent directory is used in the <b>Data Directory</b>, <b>Schema Directory</b>, <b>Cycle Completion Directory</b>, and <b>Cycle Contents Directory</b>.</p> <p>This check box is selected by default. If you clear it, for initial loads, define the full path to the output files in the <b>Data Directory</b> field. For incremental loads, optionally specify a root directory for the task in the <b>Task Target Directory</b>.</p>
Data Directory	<p><i>For initial load tasks</i>, define a directory structure for the directories where Mass Ingestion Databases stores output data files and optionally stores the schema. To define directory pattern, you can use the following types of entries:</p> <ul style="list-style-type: none"> <li>- The placeholders {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The {Timestamp} values are in the format yyyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the initial load job starts to transfer data to the target.</li> <li>- Specific directory names.</li> <li>- The toUpper() and toLower() functions, which force the values for an associated (<i>placeholder</i>) to uppercase or lowercase.</li> </ul> <p><b>Note:</b> Placeholder values are not case sensitive.</p> <p>Examples:</p> <pre>myDir1/{SchemaName}/{TableName} myDir1/myDir2/{SchemaName}/{YYYY}/{MM}/{TableName}_{Timestamp} myDir1/{toLower(SchemaName)}/{TableName}_{Timestamp}</pre> <p>The default directory pattern is {TableName}_{Timestamp}.</p> <p><i>For incremental load tasks</i>, define a custom path to the subdirectory that contains the cdc-data data files. To define the directory pattern, you can use the following types of entries:</p> <ul style="list-style-type: none"> <li>- The placeholders {TaskTargetDirectory}, {SchemaName}, {TableName}, {Timestamp}, {YY}, {YYYY}, {MM}, and {DD}, where {YY}, {YYYY}, {MM}, and {DD} are for date elements. The {Timestamp} values are in the format yyyyymmdd_hhmissms. The generated dates and times in the directory paths indicate when the CDC cycle started.</li> </ul> <p>If you include the toUpper or toLower function, put the placeholder name in parentheses and enclose the both the function and placeholder in curly brackets, as shown in the preceding example.</p> <ul style="list-style-type: none"> <li>- Specific directory names.</li> </ul> <p>The default directory pattern is {TaskTargetDirectory}/cdc-data/{TableName}/data</p> <p><b>Note:</b> For Amazon S3, Flat File, and Microsoft Azure Data Lake Storage Gen2 targets, Mass Ingestion Databases uses the directory specified in the target connection properties as the root for the data directory path when <b>Connection Directory as Parent</b> is selected. For Google Cloud Storage targets, Mass Ingestion Databases uses the <b>Bucket</b> name that you specify in the target properties for the ingestion task.</p>

Property	Description
Schema Directory	<p>For initial load and incremental load tasks, you can specify a custom directory in which to store the schema file if you want to store it in a directory other than the default directory. For initial loads, previously used values if available are shown in a drop-down list for your convenience. This field is optional.</p> <p>For initial loads, the schema is stored in the data directory by default. For incremental loads, the default directory for the schema file is <code>{TaskTargetDirectory}/cdc-data/{TableName}/schema</code></p> <p>You can use the same placeholders as for the <b>Data Directory</b> field. Ensure that you enclose placeholders with curly brackets <code>{ }</code>.</p> <p>If you include the <code>toUpper</code> or <code>toLower</code> function, put the placeholder name in parentheses and enclose the both the function and placeholder in curly brackets, for example:  <code>{toLower(SchemaName)}</code></p> <p><b>Note:</b> Schema is written only to output data files in CSV format. Data files in Parquet and Avro formats contain their own embedded schema.</p>
Cycle Completion Directory	<p>For incremental load tasks, the path to the directory that contains the cdc-cycle completed file. Default is <code>{TaskTargetDirectory}/cdc-cycle/completed</code>.</p>
Cycle Contents Directory	<p>For incremental load tasks, the path to the directory that contains the cdc-cycle contents files. Default is <code>{TaskTargetDirectory}/cdc-cycle/contents</code>.</p>
Use Cycle Partitioning for Data Directory	<p>For incremental load tasks, causes a timestamp subdirectory to be created for each CDC cycle, under each data directory.</p> <p>If this option is not selected, individual data files are written to the same directory without a timestamp, unless you define an alternative directory structure.</p>
Use Cycle Partitioning for Summary Directories	<p>For incremental load tasks, causes a timestamp subdirectory to be created for each CDC cycle, under the summary contents and completed subdirectories.</p>
List Individual Files in Contents	<p>For incremental load tasks, lists individual data files under the contents subdirectory.</p> <p>If <b>Use Cycle Partitioning for Summary Directories</b> is cleared, this option is selected by default. All of the individual files are listed in the contents subdirectory unless you can configure custom subdirectories by using the placeholders, such as for timestamp or date.</p> <p>If <b>Use Cycle Partitioning for Data Directory</b> is selected, you can still optionally select this check box to list individual files and group them by CDC cycle.</p>

Under **Advanced**, you can enter the following Microsoft Azure Data Lake Storage advanced target properties, which are primarily for incremental load and combined initial and incremental load jobs:

Field	Description
Add Operation Type	<p>Select this check box to add a metadata column that includes the source SQL operation type in the output that the job propagates to the target.</p> <p>For incremental loads, the job writes "I" for insert, "U" for update, or "D" for delete. For initial loads, the job always writes "I" for insert.</p> <p>By default, this check box is cleared.</p>
Add Operation Time	<p>Select this check box to add a metadata column that includes the source SQL operation time in the output that the job propagates to the target.</p> <p>For initial loads, the job always writes the current date and time.</p> <p>By default, this check box is cleared.</p>

Field	Description
Add Operation Owner	Select this check box to add a metadata column that includes the owner of the source SQL operation in the output that the job propagates to the target. For initial loads, the job always writes "INFA" as the owner. By default, this check box is cleared. <b>Note:</b> This property is not available for jobs that have a PostgreSQL source.
Add Operation Transaction Id	Select this check box to add a metadata column that includes the source transaction ID in the output that the job propagates to the target for SQL operations. For initial loads, the job always writes "1" as the ID. By default, this check box is cleared.
Add Before Images	Select this check box to include UNDO data in the output that an incremental load job writes to the target. For initial loads, the job writes nulls. By default, this check box is cleared.

## Microsoft Azure Synapse Analytics target properties

When you define a database ingestion task, you must enter some properties for your Microsoft Azure Synapse Analytics target on the **Target** page of the task wizard.

These properties apply to initial load, incremental load, and combined initial and incremental load operations.

The following table describes the target properties that appear under **Target**:

Property	Description
Target Creation	The only available option is <b>Create Target Tables</b> , which generates the target tables based on the source tables. <b>Note:</b> After the target table is created, Mass Ingestion Databases intelligently handles the target tables on subsequent job runs. Mass Ingestion Databases might truncate or re-create the target tables depending on the specific circumstances.
Schema	Select the target schema in which Mass Ingestion Databases creates the target tables. The schema name that is specified in the connection properties is displayed by default. Because this field is case sensitive, ensure that you entered the schema name in the connection properties in the correct case.

## Oracle target properties

When you define a database ingestion task, you must enter some properties for your Oracle target on the **Target** page of the task wizard.

These properties apply to initial load operations.



The following table describes the target properties that appear under **Target**:

Property	Description
Target Creation	The only available option is <b>Create Target Tables</b> , which generates the target tables based on the source tables. <b>Note:</b> After the target table is created, Mass Ingestion Databases intelligently handles the target tables on subsequent job runs. Mass Ingestion Databases might truncate or re-create the target tables depending on the specific circumstances.
Schema	Select the target schema in which Mass Ingestion Databases creates the target tables.

## Snowflake Cloud Data Warehouse target properties

When you define a database ingestion task, you must enter some properties for your Snowflake Cloud Data Warehouse target on the **Target** page of the task wizard. The properties vary slightly by load type.

The following table describes the Snowflake target properties that appear under **Target**:

Property	Description
Apply Mode	For incremental load and combined initial and incremental load jobs with Snowflake targets, indicates how source DML changes, including inserts, updates, and deletes, are applied to the target. Options are: <ul style="list-style-type: none"> <li>- <b>Standard.</b> Accumulate the changes in a single apply cycle and intelligently merge them into fewer SQL statements before applying them to the target. For example, if an update followed by a delete occurs on the source row, no row is applied to the target. If multiple updates occur on the same column or field, only the last update is applied to the target. If multiple updates occur on different columns or fields, the updates are merged into a single update record before being applied to the target.</li> <li>- <b>Soft Delete.</b> Apply source delete operations to the target as soft deletes. A soft delete marks the deleted row as deleted without actually removing it from the database. For example, a delete on the source results in a change record on the target with "D" displayed in the INFA_OPERATION_TYPE column. If an update followed by a delete occurs on the source, two records are written to the target both with "D" displayed in the INFA_OPERATION_TYPE column. Consider using soft deletes if you have a long-running business process that needs the soft-deleted data to finish processing, to restore data after an accidental delete operation, or to track deleted values for audit purposes.</li> <li>- <b>Audit.</b> For Snowflake targets only, ingest change data into an audit table on the target system by using insert operations, instead of merging and applying the changes to the target database. Consider using audit tables if you want to perform computations or other downstream processing on the data before applying it to the target database or if you want to examine the changes. You can add metadata columns for SQL change operations to the audit table by setting options under the <b>Advanced</b> section.</li> </ul> Default is Standard.
Target Creation	The only available option is <b>Create Target Tables</b> , which generates the target tables based on the source tables. <b>Note:</b> After the target table is created, Mass Ingestion Databases intelligently handles the target tables on subsequent job runs. Mass Ingestion Databases might truncate or re-create the target tables depending on the specific circumstances.
Schema	Select the target schema in which Mass Ingestion Databases creates the target tables.
Stage	The name of internal staging area that holds the data read from the source before the data is written to the target tables. This name must not include spaces. If the staging area does not exist, it will be automatically created.

Under **Advanced**, you can enter the following advanced target properties if you set **Apply Mode** to **Soft Deletes** or **Audit** to add metadata columns for each delete operation or each DML change recorded in the audit table:

Field	Description
Add Operation Type	<p>Select this check box to add a metadata column that records the source SQL operation type in the output that the job propagates to the target database or inserts into the audit table on the target system.</p> <p>For incremental loads and combined initial and incremental loads, the job writes "I" for insert, "U" for update, or "D" for delete. For initial loads, the job always writes "I" for insert.</p> <p>By default, this check box is selected. You cannot deselect it if you are using soft deletes.</p>
Add Operation Time	<p>Select this check box to add a metadata column that records the source SQL operation timestamp in the output that the job propagates to the target database or inserts into the audit table on the target system.</p> <p>By default, this check box is not selected.</p>
Add Operation Owner	<p>Select this check box to add a metadata column that records the owner of the source SQL operation in the output that the job propagates to the target database or inserts into the audit table on the target system.</p> <p>By default, this check box is not selected.</p> <p><b>Note:</b> This property is not available for jobs that have a PostgreSQL source.</p>
Add Operation Transaction Id	<p>Select this check box to add a metadata column that records the transaction ID of the source transaction with the SQL operation that the job propagates to the target database.</p> <p>By default, this check box is not selected.</p>
Add Operation Sequence	<p>Select this check box to add a metadata column that records a generated, ascending sequence number for each change operation that the job inserts into the audit table on the target system. The sequence number reflects the change stream position of the operation.</p> <p>By default, this check box is not selected.</p>
Add Before Images	<p>Select this check box to add _OLD columns with UNDO "before image" data in the output that the job inserts into the audit table on the target system. You can then compare the old and current values for each data column. For a delete operation, the current value will be null.</p> <p>This field is available only when <b>Apply Mode</b> is set to <b>Audit</b>.</p> <p>By default, this check box is not selected.</p>
Add Columns Prefix	<p>Select this check box to add a prefix to the names of the added metadata columns to easily identify them and to prevent conflicts with the names of existing columns.</p> <p>The default value is INFA_.</p>

## Configuring schedule and runtime options

On the **Schedule and Runtime Options** page of the database ingestion task wizard, you can specify a schedule for running initial load jobs periodically and configure runtime options for jobs of any load type.

1. Under **Advanced**, optionally edit the **Number of Rows in Output File** value to specify the maximum number of rows that the database ingestion task writes to an output data file for a Flat File, Amazon Redshift, Amazon S3, Microsoft Azure Data Lake Storage, Microsoft Azure Synapse Analytics, or Snowflake target.

**Note:** Advanced options are not displayed for incremental load tasks that have an Apache Kafka target.

For incremental load operations and combined initial and incremental load operations, change data is flushed to the target either when this number of rows is reached or when the flush latency period expires and the job is *not* in the middle of processing a transaction. The flush latency period is the time that the job waits for more change data before flushing data to the target. The latency period is internally set to 10 seconds and cannot be changed.

Valid values are 1 through 100000000. The default value for Amazon S3 and Microsoft Azure Data Lake Storage Gen2 targets is 1000 rows. For the other targets, the default value is 100000 rows.

**Note:** For Microsoft Azure Synapse Analytics targets, the data is first sent to a Microsoft Azure Data Lake Storage staging file before being written to the target tables. After data is written to the target, the entire contents of the table-specific directory that includes the staging files are emptied. For Snowflake targets, the data is first stored in an internal stage area before being written to the target tables.

2. For initial load jobs only, optionally clear the **File extension based on file type** check box if you want the output data files for Flat File, Amazon S3, and Microsoft Azure Data Lake Storage targets to have the .dat extension. This check box is selected by default, which causes the output files to have file-name extensions based on their file types.

**Note:** For incremental load jobs with these target types, this option is not available. Mass Ingestion Databases always uses output file-name extensions based on file type.

3. For database ingestion incremental load tasks that have Amazon S3 or Microsoft Azure Data Lake Storage Gen2 targets, configure the following apply cycle options:

Option	Description
Apply Cycle Interval	Specifies the amount of time that must elapse before a database ingestion job ends an apply cycle. You can specify days, hours, minutes, and seconds or specify values for a subset of these time fields leaving the other fields blank.  The default value is 15 minutes.
Apply Cycle Change Limit	Specifies the number of records that must be processed before a database ingestion job ends an apply cycle. When this record limit is reached, the database ingestion job ends the apply cycle and writes the change data to the target.  The default value is 10000 records.
Low Activity Flush Interval	Specifies the amount of time, in hours, minutes, or both, that must elapse during a period of no change activity on the source before a database ingestion job ends an apply cycle. When this time limit is reached, the database ingestion job ends the apply cycle and writes the change data to the target.  If you do not specify a value for this option, a database ingestion job ends apply cycles only after either the <b>Apply Cycle Change Limit</b> or <b>Apply Cycle Interval</b> limit is reached.  No default value is provided.

**Notes:**

- Either the **Apply Cycle Interval** or **Apply Cycle Change Limit** field must have a non-zero value or use the default value.
  - An apply cycle ends when the job reaches any of the three limits, whichever limit is met first.
4. Under **Schema Drift Options**, if the detection of schema drift is supported for your source and target combination, specify the schema drift option to use for each of the supported types of DDL operations.

Schema drift options are supported for database ingestion incremental load tasks that propagate change data from Microsoft SQL Server, Oracle, or PostgreSQL sources to Amazon Redshift, Amazon S3, Databricks Delta, Google BigQuery, Google Cloud Storage, Kafka, Microsoft Azure Data Lake Storage, Microsoft Azure Synapse Analytics, or Snowflake targets. Schema drift options are also supported for database ingestion combined initial and incremental load tasks with the same source types and with Amazon Redshift, Databricks Delta, Google BigQuery, Kafka, Microsoft Azure Synapse Analytics, or Snowflake targets.

The types of supported DDL operations are:

- Add Column
- Modify Column
- Drop Column
- Rename Column

**Note:** The Modify Column and Rename Column options are not supported and not displayed for database ingestion jobs that have Google BigQuery targets.

The following table describes the schema drift options that you can set for a DDL operation type:

Option	Description
Ignore	<p>Does not replicate DDL changes that occur on the source database to the target. For Amazon Redshift, Kafka, Microsoft Azure Synapse Analytics, or Snowflake targets, this option is the default option for the Drop Column and Rename Column operation types.</p> <p>For Amazon S3, Google Cloud Storage, and Microsoft Azure Data Lake Storage targets that use the CSV output format, the <b>Ignore</b> option is disabled. For the AVRO output format, this option is enabled.</p>
Replicate	<p>Allows the database ingestion job to replicate the DDL change to the target. For Amazon S3, Google Cloud Storage, and Microsoft Azure Data Lake Storage targets, this option is the default option for all operation types. For other targets, this option is the default option for the Add Column and Modify Column operation types.</p> <p><b>Restrictions:</b></p> <ul style="list-style-type: none"><li>- If you try to replicate a type of schema change that is not supported on the target, database ingestion jobs associated with the task will end with an error. For example, if you select <b>Replicate</b> for Rename Column operations on Microsoft Azure Synapse Analytics targets, the jobs will end.</li><li>- Add Column operations that add a primary-key column are not supported and can cause unpredictable results.</li><li>- For Databricks Delta targets, the <b>Replicate</b> option is not available for Drop Column.</li><li>- Modify Column operations that change the NULL or NOT NULL constraint for a column are not replicated to the target by design because changing the nullability of a target column can cause problems when subsequent changes are applied.</li></ul>

Option	Description
Stop Job	Stops the entire database ingestion job.
Stop Table	<p>Stops processing the source table on which the DDL change occurred. When one or more of the tables are excluded from replication because of the <b>Stop Table</b> schema drift option, the job state changes to <b>Running with Warning</b>.</p> <p><b>Important:</b> The database ingestion job cannot retrieve the data changes that occurred on the source table after the job stopped processing it. Consequently, data loss might occur on the target. To avoid data loss, you will need to resynchronize the source and target objects that the job stopped processing. Use the <b>Resume With Options &gt; Resync</b> option. For more information, see <a href="#">“Overriding schema drift options when resuming a database ingestion job” on page 273</a>.</p>

- For incremental load jobs that have an Apache Kafka target, configure the following checkpointing options:

Option	Description
Checkpoint All Rows	<p>Indicates whether a database ingestion job performs checkpoint processing for every message that is sent to the Kafka target.</p> <p><b>Note:</b> If this check box is selected, the <b>Checkpoint Every Commit</b>, <b>Checkpoint Row Count</b>, and <b>Checkpoint Frequency (secs)</b> options are ignored.</p>
Checkpoint Every Commit	Indicates whether a database ingestion job performs checkpoint processing for every commit that occurs on the source.
Checkpoint Row Count	Specifies the maximum number of messages that a database ingestion job sends to the target before adding a checkpoint. If you set this option to 0, a database ingestion job does not perform checkpoint processing based on the number of messages. If you set this option to 1, a database ingestion jobs add a checkpoint for each message.
Checkpoint Frequency (secs)	Specifies the maximum number of seconds that must elapse before a database ingestion job adds a checkpoint. If you set this option to 0, a database ingestion job does not perform checkpoint processing based on elapsed time.

- Under **Schedule**, if you want to run job instances for an initial load task based on an existing schedule instead of manually starting the job after it is deployed from one of the monitoring interfaces, select **Run this task based on a schedule** and then select a predefined schedule. The default option is **Do not run this task based on a schedule**.

This field is unavailable for incremental load and combined initial and incremental load tasks.

You can view and edit the schedule options in Administrator. If you edit the schedule, the changes will apply to all jobs that use the schedule. If you edit the schedule after deploying the task, you do not need to redeploy the task.

If the schedule criteria for running the job is met but the previous job run is still active, Mass Ingestion Databases skips the new job run.

- Under **Custom Properties**, you can specify custom properties that Informatica provides to meet your special requirements. To add a property, in the **Create Property** fields, enter the property name and value. Then click **Add Property**.

Specify these properties only at the direction of Informatica Global Customer Support. Usually, these properties address unique environments or special processing needs. You can specify multiple properties, if necessary. A property name can contain only alphanumeric characters and the following special characters: periods (.), hyphens (-), and underscores (\_).

**Tip:** To delete a property, click the Delete icon button at the right end of the property row in the list.

8. Click **Save**.

## Deploying a database ingestion task

After you define a database ingestion task and save it, deploy the task to create an executable job instance on the on-premises system that contains the Secure Agent and the Database Ingestion agent service and DBMI packages. You must deploy the task before you can run the job. The deploy process also validates the task definition.

Before you deploy a task that has a Microsoft Azure Synapse Analytics or Snowflake target, drop any existing target tables that do not match the structure of the source tables, for example, because of added or dropped source columns or altered column null constraints or data types. When you deploy the task, the target tables are generated based the latest source structure.

- To deploy a task, in the database ingestion task wizard, save the completed task definition and then click **Deploy**.

After you deploy a task successfully, the associated job instance is in the Deployed state. You can run it from the **My Jobs** page in Mass Ingestion or from the **All Jobs** tab on the **Mass Ingestion** page in Operational Insights or Monitor.

### Deployment considerations:

- If you included spaces in the database ingestion task name, the spaces are omitted from the corresponding job name for the generated job instance.
- If the deployment process fails, the job status switches to Failed. To diagnose the error, download the error log from the **My Jobs** page in Mass Ingestion or the **All Jobs** tab on the **Mass Ingestion** page in Operational Insights or Monitor. In the Actions menu for the job, click **Error Log**. After you resolve the problem, deploy the task again either from the database ingestion task wizard or from the Actions menu for the failed job.
- If you undeploy a job and then want to run a job for the associated ingestion task again, you must deploy the task again to create a new job instance. The new job instance name ends with an incremented number in the format *taskname-job\_instance\_number*. The job instance number is incremented each time you deploy the ingestion task by adding 1 to the maximum instance number across all ingestion jobs.
- If the Secure Agent is restarted while the task is deploying, the job status switches to Failed. Avoid restarting the Secure Agent while tasks are being deployed.
- If a task appears to be hung in the Deploying state, restart the Secure Agent. The associated job instance acquires the status of Failed. You can then deploy it again.

## Running database ingestion jobs

You can run a deployed database ingestion job from one of the monitoring interfaces, or when you create a database ingestion initial load task, you can specify a schedule for running the job instances associated with a task.

## Running a deployed database ingestion job

You can run a database ingestion job that has been previously deployed and is in a state other than Undeployed from the **My Jobs** page in the Mass Ingestion service or from the **All Jobs** tab on the **Mass Ingestion** page in Operational Insights or Monitor.

1. Navigate to the row for the job that you want to run.
2. In the Actions menu for the row, click **Run**.

A subtask is started for each source table.

### Notes:

- If the initial load portion of a combined initial and incremental load job fails to load data from a source table to a target table, the database ingestion job retries the subtask for the table up to three times. The interval between retries is a minimum of 60 seconds. If all of the initial load retries fail, the subtask acquires the state of **Error** and the table is excluded from replication. The job then tries to proceed with incremental loading. In this case, the job status changes to **Running with Warning**.
- If an initial load job detects inconsistencies between column definitions in the source and target tables, the job drops the target table and then re-creates it to be consistent with the source table, prior to loading source data to the target.
- For initial load and combined tasks, the initial load might take a long time to perform if the source tables contain many rows.

## Running initial load jobs based on a schedule

When you configure a database ingestion initial load task, you can specify a schedule for running job instances associated with the task. You must have previously defined the schedule in Administrator.

1. In Administrator, create the schedule if it does not already exist.  
You can set schedule options to start a job instance on a specific date and time, plus 10 seconds, based on a specific time zone and to run job instances on a reoccurring basis.
2. When you configure a database ingestion initial load task in Mass Ingestion, on the **Schedule and Runtime Options** page, select **Run this task based on a schedule** and then select the schedule.

For more information, see "Scheduling" in Administrator help.

## Managing database ingestion jobs

After you configure and run databases tasks, you might occasionally need to perform some job management tasks such stopping, resuming, undeploying, or redeploying jobs.

## Stopping a database ingestion job

You can stop a database ingestion job of any load type that is in the **Up and Running**, **Running with Warning**, or **On Hold** state.

For an incremental load job, the job stops after a checkpoint is taken. A checkpoint records the point in the change stream where incremental processing left off for recovery purposes.

For a combined initial and incremental load job, initial load subtasks that are running are allowed to run to completion and initial load subtasks that are not running remain in their current states. For the incremental load portion of the job, a checkpoint is written to the checkpoint file or target recovery table before the job stops. The database ingestion job will not be able to record a checkpoint unless a change record has been processed for at least one of the tables in the job during the first job run after deployment. If a checkpoint is not available, the job resumes processing from the configured restart point, which is the latest available position in the change stream by default.

For an initial load job, any *running* subtasks are allowed to run to completion and then the job stops. Non-running subtasks remain in their current states.

- From the Actions menu for the job, select **Stop**.

The job state switches to **Stopping** and then to **Stopped**.

**Tip:** If the Stop operation is taking too long, you can abort the job.

## Aborting a database ingestion job

You can abort a database ingestion job that is in the **Up and Running**, **Running with Warning**, **On Hold**, or **Stopping** state.

For an incremental load job, the job stops immediately after a checkpoint is taken. A checkpoint records the point in the change stream where incremental processing left off for recovery purposes.

For a combined initial and incremental load job, any *running* initial load subtasks stop immediately. For the incremental load portion of the job, a checkpoint is taken and then the job stops.

For an initial load job, any *running* subtasks stop immediately and then the job stops. Non-running subtasks remain in their current states.

- From the Actions menu for the job, select **Abort**.

The job state switches to **Aborting** and then to **Aborted**.

For initial load jobs, the state of started and running subtasks switches to **Aborted**. For incremental load or combined initial and incremental load jobs, the state of subtasks switches to **Stopped**.

## Resuming a database ingestion job

You can resume a database ingestion job that is in the Stopped, Aborted, or Failed state.

You can resume a job from either the **My Jobs** page in the Mass Ingestion service or from the **All Jobs** tab on the **Mass Ingestion** page in Operational Insights or Monitor.

When you resume an initial load job that has multiple subtasks, Mass Ingestion Databases starts only the subtasks that are in a Failed, Stopped, Aborted, or Queued state.

When you resume an incremental load job or a combined initial and incremental load job, Mass Ingestion Databases resumes replicating source data change from the last position recorded in the checkpoint file or target recovery table. A checkpoint will not be available unless a change record was processed for at least one of the tables during the first job run after deployment. If a checkpoint is not available, the job resumes processing from the configured restart point, which is the latest available position in the change stream by default.

**Note:** For initial load jobs, the **Run** command might also be available. Click **Run** if you want the ingestion job to truncate all of the target tables and then reload the source data to the target tables.

1. Navigate to the row for the job that you want to resume.



2. In the Actions menu for the row, click **Resume**.

**Note:** The **Resume** command is not available if the job is in the **Failed** state because the task deployment failed.

A subtask is started for each source table.

If an error occurs, an error message is displayed at the top of the page.

## Overriding schema drift options when resuming a database ingestion job

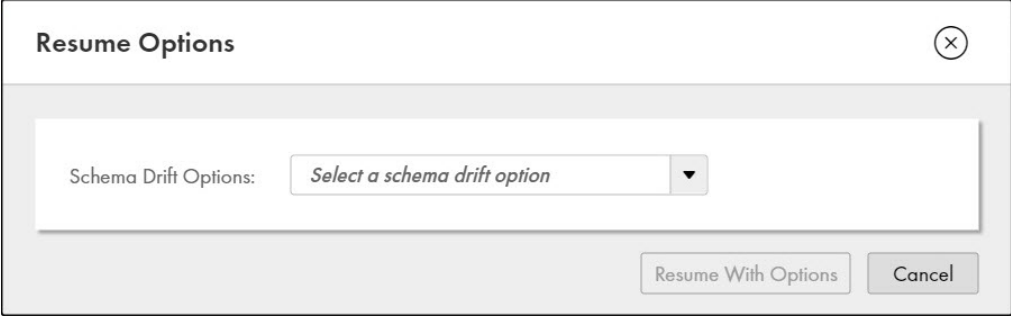
You can override the schema drift options when you resume a database ingestion job that is in the Stopped, Aborted, or Failed state. The overrides affect only those tables that are currently in the Error state because of the **Stop Table** or **Stop Job** Schema Drift option. Use the overrides to correct or resolve these errors.

You can override schema drift options and resume an incremental load job or a combined initial and incremental load job either from the **My Jobs** page in the Mass Ingestion service or from the **All Jobs** tab on the **Mass Ingestion** page in Operational Insights or Monitor.

1. Navigate to the row for the job that you want to resume with an override.
2. In the Actions menu for the row, click **Resume With Options**.

**Note:** The **Resume With Options** command is not available if the job is in the **Failed** state because the task deployment failed.

The **Resume Options** dialog box appears.

A screenshot of the 'Resume Options' dialog box. The dialog has a title bar with 'Resume Options' and a close button (X). Inside, there is a label 'Schema Drift Options:' followed by a dropdown menu with the text 'Select a schema drift option'. At the bottom right, there are two buttons: 'Resume With Options' and 'Cancel'.

3. In the **Schema Drift Options** list, select the schema drift option that will be used to process the DDL operation on the source that caused the database ingestion job to stop.

The following table describes the schema drift options:

Option	Description
Ignore	Do not replicate DDL changes that occur on the source database to the target.
Stop Table	Stop processing the source table on which the DDL change occurred. <b>Important:</b> The database ingestion job cannot retrieve the data changes that occurred on the source table after the job stopped processing it. Consequently, data loss might occur on the target. To avoid data loss, you will need to resynchronize the source and target objects that the job stopped processing. Use the <b>Resume With Options &gt; Resync</b> option.

Option	Description
Resync	Resynchronize the target table with the source table. Use this option for tables that the job stopped processing because of the <b>Stop Table</b> setting for a <b>Schema Drift</b> option. <b>Important:</b> This option is available only for combined initial and incremental load jobs.
Replicate	Allow the database ingestion job to replicate the DDL change to the target. <b>Important:</b> If you specify the <b>Replicate</b> option for Rename Column operations on Microsoft Azure Synapse Analytics targets, the job will end with an error.

4. Click **Resume With Options**.

The resumed job will use the schema drift option that you specified in step 3 to process the schema change that caused the job to stop. Thereafter, the schema drift options that you specified when creating the task take effect again.

**Important:** Mass Ingestion Databases processes a schema change to a source table only after a DML operation occurs on the table. Therefore, after you resume a job, the table subtask state remains unchanged until the first DML operation occurs on the table.

## Redeploying a database ingestion job

Redeploy a database ingestion job after editing available fields in the associated database ingestion ingestion task so that the new settings can take effect.

You can edit some but not all of fields in an ingestion task definition that has been previously deployed, without first undeploying the job. You can add a source table and change any of the runtime and target options that are available for editing. For example, you might want to reset some target options to test the effects of different settings.

The redeploy operation stops each job subtask for a source table, deploys the updated ingestion task, and automatically starts the subtasks that were stopped and any new subtasks for added source tables.

**Note:** For incremental load jobs and combined initial and incremental load jobs, the redeploy operation does not change the list of selected tables that was created during a previous deployment. To update the list of tables, edit the table selection rules in the associated task and then redeploy the job. You must make an update to the table selection rules, even if you added a table that matches the existing table selection rules.

1. On the **My Jobs** page, navigate to the row for the job that you want to redeploy.
2. In the Actions menu for the row, select **Redeploy**.

The job instance automatically starts running.

If the job is running when you select **Redeploy**, Mass Ingestion Databases stops the job and then redeploys the ingestion task and restarts the job.

For jobs with Microsoft Azure Synapse Analytics or Snowflake Cloud Data Warehouse targets, the redeploy operation also validates that the target tables exist and creates new ones if table selection rules have changed.

## Undeploying a database ingestion job

Undeploy a database ingestion job that was previously deployed if you no longer need to run the job or if you need to change a connection or property in the associated ingestion task that cannot be edited without first undeploying the job.

Before you attempt to undeploy a job, ensure that it is not running.

After the job is undeployed, you cannot run it again or redeploy it. If you want to run a job for the associated ingestion task again, you must deploy the task again from the task wizard to create a new job instance. For example, if you want to change the target connection, undeploy the job, edit the ingestion task to change the connection, deploy the task again, and then run the new job instance.

1. On the **My Jobs** page in Mass Ingestion or on the **All Jobs** tab of the Mass Ingestion page in Operational Insights or Monitor, navigate to the row for the job that you want to undeploy.
2. In the **Actions** menu for the row, click **Undeploy**.

The Undeploy command is not available for a Failed job if the job has not been previously deployed.

If the undeploy operation fails, the job state switches to Failed, even if it was in the Aborted state.

## Resynchronizing source and target objects

You can resynchronize source and target objects for a subtask that is part of a running database ingestion combined initial and incremental load job. The subtask must be in a state other than Queued or Starting.

For example, you might want to resynchronize the target with the source if initial load or incremental load processing failed or if you want to start the job over again from a specific restart point.

**Important:** To resynchronize tables that stopped and are currently in the **Error** state because of the **Schema Drift** setting of **Stop Table**, you must use the **Resume With Options > Resync** option in the Actions menu. For more information, see [“Overriding schema drift options when resuming a database ingestion job” on page 273](#).

1. On the **My Jobs** page in the Mass Ingestion service or on the **All Jobs** tab of the **Mass Ingestion** page in Operational Insights or Monitor, drill down on an ingestion job to display job details.

The job must be in the **Up and Running** state and be for a combined initial and incremental load operation.

2. Click the **Object Detail** tab.
3. In the subtask row for the source and target objects that you want to resynchronize, click the Actions menu and select **Resync**.

For the Actions menu and **Resync** option to be available, the subtask must be in a state other than Queued or Starting.

If the source table schema does not match the target table schema, the ingestion subtask drops the target table and creates a new table that matches the source schema. Regardless of whether the target tables are re-created, the subtask truncates the target tables and then reloads source data to the tables.

**Important:** If the source table contains many rows, the resynchronization might take a long time to perform.

## Restart and recovery for incremental load jobs

Mass Ingestion Databases can restart incremental load and combined initial and incremental load jobs that stopped because of an error or a user stop request without losing change data.

After the first job run, Mass Ingestion Databases continually records an identifier for the processing position in the change stream as changes are applied to the target. For file-based targets such as Amazon S3, Azure Data Lake Storage, Google Cloud Storage, and Kafka, the identifier is stored in a checkpoint file. For database targets, the identifier is stored in a generated recovery table, called `INFORMATICA_CDC_RECOVERY`, on the target.

**Note:** For the first job run, Mass Ingestion Databases uses the start point that you set in the **Restart Point for Incremental Load** field when defining the database ingestion task.

If incremental job processing ends abnormally or in response to a user stop or abort request and you then resume the job, the job resumes from the last position saved to the checkpoint file or recovery table. A

checkpoint will not be available unless a change record was processed for at least one of the tables during the first job run after deployment. If a checkpoint is not available, the job resumes processing from the configured restart point, which is the latest available position in the change stream by default.

## Default Data Type Mappings

This reference provides default data-type mappings for relational sources and Amazon Redshift, Databricks Delta, Google BigQuery, Microsoft Azure Synapse Analytics, Oracle, and Snowflake targets. When Mass Ingestion Databases generates the target tables, it uses these mappings.

When you configure a target, you can optionally define data-type mapping rules to customize the default mappings of source data types to target data tables. For more information, see [“Rules for customizing data type mappings” on page 245](#).

If a source data type is not listed, Mass Ingestion Databases either cannot extract data from the source columns with this data type or cannot apply the extracted data to any appropriate target data type.

### Db2 for i Source and Amazon Redshift Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Db2 for i source and an Amazon Redshift target:

Db2 for i Source Data Type	Amazon Redshift Target Data Type
bigint	bigint
binary(size), 1 <= s <= 32766	character varying(size), 2 <= s <= 65532
char(size) for bit data, 1 <= s <= 32766	character varying(size), 2 <= s <= 65532
char(size), 1 <= s <= 32766	character varying(size), 4 <= s <= 65535
date	date
decfloat(precision), 16 <= p <= 34	character varying(255)
decimal(p,s), 1 <= p <= 38, 0 <= s <= 37	numeric(p,s), 1 <= p <= 38, 0 <= s <= 37
decimal(p,s), 38 <= p <= 63, 1 <= s <= 62	character varying(size), 40 <= s <= 65
float	double precision
integer	integer
long varbinary	character varying(65535)
long varchar	character varying(65535)
long varchar for bit data	character varying(65535)
numeric(p,s), 1 <= p <= 38, 0 <= s <= 37	numeric(p,s), 1 <= p <= 38, 0 <= s <= 37

Db2 for i Source Data Type	Amazon Redshift Target Data Type
numeric(p,s), 38 <= p <= 63, 1 <= s <= 62	character varying(size), 40 <= s <= 65
real	real
rowid	character varying(80)
smallint	smallint
time	time without time zone
timestamp(precision), 0 <= p <= 6	timestamp without time zone
timestamp(precision), 7 <= p <= 12	character varying(size), 27 <= s <= 32
varbinary(size), 1 <= s <= 32740	character varying(size), 2 <= s <= 65480
varchar(size) for bit data, 1 <= s <= 32740	character varying(size), 2 <= s <= 65480
varchar(size), 1 <= s <= 32740	character varying(size), 4 <= s <= 65535

**Note:** The maximum supported length of DECFLOAT values on the target is 255 characters.

## Db2 for i Source and Databricks Delta Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Db2 for i source and a Databricks Delta target:

Db2 for i Source Data Type	Databricks Delta Target Data Type
bigint	long
binary(size), 1 <= s <= 32766	binary
char(size) for bit data, 1 <= s <= 32766	binary
char(size), 1 <= s <= 32766	string
date	string
decfloat(precision), 16 <= p <= 34	string
decimal(p,s), 1 <= p <= 38, 0 <= s <= 38	decimal(p,s), 1 <= p <= 38, 0 <= s <= 38
decimal(p,s), 39 <= p <= 63, 0 <= s <= 62	string
float	double
integer	integer
long varbinary	binary
long varchar	string

Db2 for i Source Data Type	Databricks Delta Target Data Type
long varchar for bit data	binary
numeric(p,s), 1 <= p <= 38, 0 <= s <= 38	decimal(p,s), 1 <= p <= 38, 0 <= s <= 38
numeric(p,s), 39 <= p <= 63, 0 <= s <= 62	string
real	float
rowid	binary
smallint	integer
time	string
timestamp(precision), 0 <= p <= 6	timestamp
timestamp(precision), 7 <= p <= 12	string
varbinary(size), 1 <= s <= 32740	binary
varchar(size) for bit data, 1 <= s <= 32740	binary
varchar(size), 1 <= s <= 32740	string

## Db2 for i Source and Google BigQuery Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Db2 for i source and a Google BigQuery target:

Db2 for i Source Data Type	Google BigQuery Target Data Type
bigint	int64
binary(size), 1 <= s <= 32766	bytes
char(size) for bit data, 1 <= s <= 32766	bytes
char(size), 1 <= s <= 32766	string
date	date
decfloat(precision), 16 <= p <= 34	string
decimal	int64
decimal(1)	int64
decimal(1,1)	numeric
decimal(10)	int64
decimal(10,10)	bignumeric

Db2 for i Source Data Type	Google BigQuery Target Data Type
decimal(11)	int64
decimal(11,s), 10 <= s <= 11	bignumeric
decimal(12)	int64
decimal(12,s), 10 <= s <= 12	bignumeric
decimal(13)	int64
decimal(13,s), 10 <= s <= 13	bignumeric
decimal(14)	int64
decimal(14,s), 10 <= s <= 14	bignumeric
decimal(15)	int64
decimal(15,s), 10 <= s <= 15	bignumeric
decimal(16)	int64
decimal(16,s), 10 <= s <= 16	bignumeric
decimal(17)	int64
decimal(17,s), 10 <= s <= 17	bignumeric
decimal(18)	int64
decimal(18,s), 10 <= s <= 18	bignumeric
decimal(19,s), 10 <= s <= 19	bignumeric
decimal(2)	int64
decimal(2,s), 1 <= s <= 2	numeric
decimal(20,s), 10 <= s <= 20	bignumeric
decimal(21,s), 10 <= s <= 21	bignumeric
decimal(22,s), 10 <= s <= 22	bignumeric
decimal(23,s), 10 <= s <= 23	bignumeric
decimal(24,s), 10 <= s <= 24	bignumeric
decimal(25,s), 10 <= s <= 25	bignumeric
decimal(26,s), 10 <= s <= 26	bignumeric
decimal(27,s), 10 <= s <= 27	bignumeric

Db2 for i Source Data Type	Google BigQuery Target Data Type
decimal(28,s), 10 <= s <= 28	bignumeric
decimal(3)	int64
decimal(3,s), 1 <= s <= 3	numeric
decimal(38,9)	numeric
decimal(4)	int64
decimal(4,s), 1 <= s <= 4	numeric
decimal(40,s), 1 <= s <= 21	bignumeric
decimal(41,s), 2 <= s <= 22	bignumeric
decimal(42,s), 3 <= s <= 23	bignumeric
decimal(43,s), 4 <= s <= 24	bignumeric
decimal(44,s), 5 <= s <= 25	bignumeric
decimal(45,s), 6 <= s <= 26	bignumeric
decimal(46,s), 7 <= s <= 27	bignumeric
decimal(47,s), 8 <= s <= 28	bignumeric
decimal(48,s), 9 <= s <= 29	bignumeric
decimal(49,s), 10 <= s <= 30	bignumeric
decimal(5,s), 1 <= s <= 5	numeric
decimal(50,s), 11 <= s <= 31	bignumeric
decimal(51,s), 12 <= s <= 32	bignumeric
decimal(52,s), 13 <= s <= 33	bignumeric
decimal(53,s), 14 <= s <= 34	bignumeric
decimal(54,s), 15 <= s <= 35	bignumeric
decimal(55,s), 16 <= s <= 36	bignumeric
decimal(56,38)	bignumeric
decimal(56,s), 17 <= s <= 37	bignumeric
decimal(6)	int64
decimal(6,s), 1 <= s <= 6	numeric



Db2 for i Source Data Type	Google BigQuery Target Data Type
decimal(63,s), 24 <= s <= 38	bignumeric
decimal(7)	int64
decimal(7,s), 1 <= s <= 7	numeric
decimal(8)	int64
decimal(8,s), 1 <= s <= 8	numeric
decimal(9)	int64
decimal(p,39), 40 <= p <= 39	string
decimal(p,s), 10 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 11 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 12 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 13 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 14 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 15 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 16 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 17 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 18 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 19 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 20 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 21 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 22 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 23 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 24 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 25 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 26 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 27 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 28 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 29 <= p <= 30, 10 <= s <= 29	bignumeric

Db2 for i Source Data Type	Google BigQuery Target Data Type
decimal(p,s), 29 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 30 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 30 <= p <= 49, 1 <= s <= 30	bignumeric
decimal(p,s), 31 <= p <= 38, 2 <= s <= 9	numeric
decimal(p,s), 31 <= p <= 50, 1 <= s <= 31	bignumeric
decimal(p,s), 32 <= p <= 38, 3 <= s <= 9	numeric
decimal(p,s), 32 <= p <= 51, 1 <= s <= 32	bignumeric
decimal(p,s), 33 <= p <= 38, 4 <= s <= 9	numeric
decimal(p,s), 33 <= p <= 52, 1 <= s <= 33	bignumeric
decimal(p,s), 34 <= p <= 38, 5 <= s <= 9	numeric
decimal(p,s), 34 <= p <= 53, 1 <= s <= 34	bignumeric
decimal(p,s), 35 <= p <= 38, 6 <= s <= 9	numeric
decimal(p,s), 35 <= p <= 54, 1 <= s <= 35	bignumeric
decimal(p,s), 36 <= p <= 38, 7 <= s <= 9	numeric
decimal(p,s), 36 <= p <= 55, 1 <= s <= 36	bignumeric
decimal(p,s), 37 <= p <= 38, 8 <= s <= 9	numeric
decimal(p,s), 37 <= p <= 56, 1 <= s <= 37	bignumeric
decimal(p,s), 38 <= p <= 56, 1 <= s <= 38	bignumeric
decimal(p,s), 39 <= p <= 56, 21 <= s <= 38	bignumeric
decimal(p,s), 40 <= p <= 41, 1 <= s <= 40	string
decimal(p,s), 41 <= p <= 42, 1 <= s <= 41	string
decimal(p,s), 41 <= p <= 56, 23 <= s <= 38	bignumeric
decimal(p,s), 42 <= p <= 43, 1 <= s <= 42	string
decimal(p,s), 42 <= p <= 56, 24 <= s <= 38	bignumeric
decimal(p,s), 43 <= p <= 44, 1 <= s <= 43	string
decimal(p,s), 43 <= p <= 56, 25 <= s <= 38	bignumeric
decimal(p,s), 44 <= p <= 56, 26 <= s <= 38	bignumeric

Db2 for i Source Data Type	Google BigQuery Target Data Type
decimal(p,s), 44 <= p <= 63, 1 <= s <= 44	string
decimal(p,s), 45 <= p <= 56, 27 <= s <= 38	bignumeric
decimal(p,s), 45 <= p <= 63, 1 <= s <= 45	string
decimal(p,s), 46 <= p <= 56, 28 <= s <= 38	bignumeric
decimal(p,s), 46 <= p <= 63, 1 <= s <= 46	string
decimal(p,s), 47 <= p <= 56, 29 <= s <= 38	bignumeric
decimal(p,s), 47 <= p <= 63, 1 <= s <= 47	string
decimal(p,s), 48 <= p <= 56, 30 <= s <= 38	bignumeric
decimal(p,s), 48 <= p <= 63, 1 <= s <= 48	string
decimal(p,s), 49 <= p <= 56, 31 <= s <= 38	bignumeric
decimal(p,s), 49 <= p <= 63, 1 <= s <= 49	string
decimal(p,s), 50 <= p <= 56, 32 <= s <= 38	bignumeric
decimal(p,s), 50 <= p <= 63, 1 <= s <= 50	string
decimal(p,s), 51 <= p <= 56, 33 <= s <= 38	bignumeric
decimal(p,s), 51 <= p <= 63, 1 <= s <= 51	string
decimal(p,s), 52 <= p <= 56, 34 <= s <= 38	bignumeric
decimal(p,s), 52 <= p <= 63, 1 <= s <= 52	string
decimal(p,s), 53 <= p <= 56, 35 <= s <= 38	bignumeric
decimal(p,s), 53 <= p <= 63, 1 <= s <= 53	string
decimal(p,s), 54 <= p <= 56, 36 <= s <= 38	bignumeric
decimal(p,s), 54 <= p <= 63, 1 <= s <= 54	string
decimal(p,s), 55 <= p <= 56, 37 <= s <= 38	bignumeric
decimal(p,s), 55 <= p <= 63, 1 <= s <= 55	string
decimal(p,s), 56 <= p <= 63, 1 <= s <= 56	string
decimal(p,s), 56 <= p <= 63, 39 <= s <= 62	string
decimal(p,s), 57 <= p <= 63, 1 <= s <= 57	string
decimal(p,s), 57 <= p <= 63, 18 <= s <= 38	bignumeric

Db2 for i Source Data Type	Google BigQuery Target Data Type
decimal(p,s), 58 <= p <= 63, 1 <= s <= 58	string
decimal(p,s), 58 <= p <= 63, 19 <= s <= 38	bignumeric
decimal(p,s), 59 <= p <= 63, 1 <= s <= 59	string
decimal(p,s), 59 <= p <= 63, 20 <= s <= 38	bignumeric
decimal(p,s), 60 <= p <= 63, 1 <= s <= 60	string
decimal(p,s), 60 <= p <= 63, 21 <= s <= 38	bignumeric
decimal(p,s), 61 <= p <= 63, 1 <= s <= 61	string
decimal(p,s), 61 <= p <= 63, 22 <= s <= 38	bignumeric
decimal(p,s), 62 <= p <= 63, 1 <= s <= 62	string
decimal(p,s), 62 <= p <= 63, 23 <= s <= 38	bignumeric
decimal(p,s), 9 <= p <= 38, 1 <= s <= 9	numeric
float	float64
integer	int64
long varbinary	bytes
long varchar	string
long varchar for bit data	bytes
numeric	int64
numeric(1)	int64
numeric(1,1)	numeric
numeric(10)	int64
numeric(10,10)	bignumeric
numeric(11)	int64
numeric(11,s), 10 <= s <= 11	bignumeric
numeric(12)	int64
numeric(12,s), 10 <= s <= 12	bignumeric
numeric(13)	int64
numeric(13,s), 10 <= s <= 13	bignumeric

Db2 for i Source Data Type	Google BigQuery Target Data Type
numeric(14)	int64
numeric(14,s), 10 <= s <= 14	bignumeric
numeric(15)	int64
numeric(15,s), 10 <= s <= 15	bignumeric
numeric(16)	int64
numeric(16,s), 10 <= s <= 16	bignumeric
numeric(17)	int64
numeric(17,s), 10 <= s <= 17	bignumeric
numeric(18)	int64
numeric(18,s), 10 <= s <= 18	bignumeric
numeric(19,s), 10 <= s <= 19	bignumeric
numeric(2)	int64
numeric(2,s), 1 <= s <= 2	numeric
numeric(20,s), 10 <= s <= 20	bignumeric
numeric(21,s), 10 <= s <= 21	bignumeric
numeric(22,s), 10 <= s <= 22	bignumeric
numeric(23,s), 10 <= s <= 23	bignumeric
numeric(24,s), 10 <= s <= 24	bignumeric
numeric(25,s), 10 <= s <= 25	bignumeric
numeric(26,s), 10 <= s <= 26	bignumeric
numeric(27,s), 10 <= s <= 27	bignumeric
numeric(28,s), 10 <= s <= 28	bignumeric
numeric(3)	int64
numeric(3,s), 1 <= s <= 3	numeric
numeric(38,9)	numeric
numeric(4)	int64
numeric(4,s), 1 <= s <= 4	numeric

Db2 for i Source Data Type	Google BigQuery Target Data Type
numeric(40,s), 1 <= s <= 21	bignumeric
numeric(41,s), 2 <= s <= 22	bignumeric
numeric(42,s), 3 <= s <= 23	bignumeric
numeric(43,s), 4 <= s <= 24	bignumeric
numeric(44,s), 5 <= s <= 25	bignumeric
numeric(45,s), 6 <= s <= 26	bignumeric
numeric(46,s), 7 <= s <= 27	bignumeric
numeric(47,s), 8 <= s <= 28	bignumeric
numeric(48,s), 9 <= s <= 29	bignumeric
numeric(49,s), 10 <= s <= 30	bignumeric
numeric(5,s), 1 <= s <= 5	numeric
numeric(50,s), 11 <= s <= 31	bignumeric
numeric(51,s), 12 <= s <= 32	bignumeric
numeric(52,s), 13 <= s <= 33	bignumeric
numeric(53,s), 14 <= s <= 34	bignumeric
numeric(54,s), 15 <= s <= 35	bignumeric
numeric(55,s), 16 <= s <= 36	bignumeric
numeric(56,38)	bignumeric
numeric(56,s), 17 <= s <= 37	bignumeric
numeric(6)	int64
numeric(6,s), 1 <= s <= 6	numeric
numeric(63,s), 24 <= s <= 38	bignumeric
numeric(7)	int64
numeric(7,s), 1 <= s <= 7	numeric
numeric(8)	int64
numeric(8,s), 1 <= s <= 8	numeric
numeric(9)	int64

Db2 for i Source Data Type	Google BigQuery Target Data Type
numeric(p,39), 40 <= p <= 39	string
numeric(p,s), 10 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 11 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 12 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 13 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 14 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 15 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 16 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 17 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 18 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 19 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 20 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 21 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 22 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 23 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 24 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 25 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 26 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 27 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 28 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 29 <= p <= 30, 10 <= s <= 29	bignumeric
numeric(p,s), 29 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 30 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 30 <= p <= 49, 1 <= s <= 30	bignumeric
numeric(p,s), 31 <= p <= 38, 2 <= s <= 9	numeric
numeric(p,s), 31 <= p <= 50, 1 <= s <= 31	bignumeric
numeric(p,s), 32 <= p <= 38, 3 <= s <= 9	numeric

Db2 for i Source Data Type	Google BigQuery Target Data Type
numeric(p,s), 32 <= p <= 51, 1 <= s <= 32	bignumeric
numeric(p,s), 33 <= p <= 38, 4 <= s <= 9	numeric
numeric(p,s), 33 <= p <= 52, 1 <= s <= 33	bignumeric
numeric(p,s), 34 <= p <= 38, 5 <= s <= 9	numeric
numeric(p,s), 34 <= p <= 53, 1 <= s <= 34	bignumeric
numeric(p,s), 35 <= p <= 38, 6 <= s <= 9	numeric
numeric(p,s), 35 <= p <= 54, 1 <= s <= 35	bignumeric
numeric(p,s), 36 <= p <= 38, 7 <= s <= 9	numeric
numeric(p,s), 36 <= p <= 55, 1 <= s <= 36	bignumeric
numeric(p,s), 37 <= p <= 38, 8 <= s <= 9	numeric
numeric(p,s), 37 <= p <= 56, 1 <= s <= 37	bignumeric
numeric(p,s), 38 <= p <= 56, 1 <= s <= 38	bignumeric
numeric(p,s), 39 <= p <= 56, 21 <= s <= 38	bignumeric
numeric(p,s), 40 <= p <= 41, 1 <= s <= 40	string
numeric(p,s), 41 <= p <= 42, 1 <= s <= 41	string
numeric(p,s), 41 <= p <= 56, 23 <= s <= 38	bignumeric
numeric(p,s), 42 <= p <= 43, 1 <= s <= 42	string
numeric(p,s), 42 <= p <= 56, 24 <= s <= 38	bignumeric
numeric(p,s), 43 <= p <= 44, 1 <= s <= 43	string
numeric(p,s), 43 <= p <= 56, 25 <= s <= 38	bignumeric
numeric(p,s), 44 <= p <= 56, 26 <= s <= 38	bignumeric
numeric(p,s), 44 <= p <= 63, 1 <= s <= 44	string
numeric(p,s), 45 <= p <= 56, 27 <= s <= 38	bignumeric
numeric(p,s), 45 <= p <= 63, 1 <= s <= 45	string
numeric(p,s), 46 <= p <= 56, 28 <= s <= 38	bignumeric
numeric(p,s), 46 <= p <= 63, 1 <= s <= 46	string
numeric(p,s), 47 <= p <= 56, 29 <= s <= 38	bignumeric



Db2 for i Source Data Type	Google BigQuery Target Data Type
numeric(p,s), 47 <= p <= 63, 1 <= s <= 47	string
numeric(p,s), 48 <= p <= 56, 30 <= s <= 38	bignumeric
numeric(p,s), 48 <= p <= 63, 1 <= s <= 48	string
numeric(p,s), 49 <= p <= 56, 31 <= s <= 38	bignumeric
numeric(p,s), 49 <= p <= 63, 1 <= s <= 49	string
numeric(p,s), 50 <= p <= 56, 32 <= s <= 38	bignumeric
numeric(p,s), 50 <= p <= 63, 1 <= s <= 50	string
numeric(p,s), 51 <= p <= 56, 33 <= s <= 38	bignumeric
numeric(p,s), 51 <= p <= 63, 1 <= s <= 51	string
numeric(p,s), 52 <= p <= 56, 34 <= s <= 38	bignumeric
numeric(p,s), 52 <= p <= 63, 1 <= s <= 52	string
numeric(p,s), 53 <= p <= 56, 35 <= s <= 38	bignumeric
numeric(p,s), 53 <= p <= 63, 1 <= s <= 53	string
numeric(p,s), 54 <= p <= 56, 36 <= s <= 38	bignumeric
numeric(p,s), 54 <= p <= 63, 1 <= s <= 54	string
numeric(p,s), 55 <= p <= 56, 37 <= s <= 38	bignumeric
numeric(p,s), 55 <= p <= 63, 1 <= s <= 55	string
numeric(p,s), 56 <= p <= 63, 1 <= s <= 56	string
numeric(p,s), 56 <= p <= 63, 39 <= s <= 62	string
numeric(p,s), 57 <= p <= 63, 1 <= s <= 57	string
numeric(p,s), 57 <= p <= 63, 18 <= s <= 38	bignumeric
numeric(p,s), 58 <= p <= 63, 1 <= s <= 58	string
numeric(p,s), 58 <= p <= 63, 19 <= s <= 38	bignumeric
numeric(p,s), 59 <= p <= 63, 1 <= s <= 59	string
numeric(p,s), 59 <= p <= 63, 20 <= s <= 38	bignumeric
numeric(p,s), 60 <= p <= 63, 1 <= s <= 60	string
numeric(p,s), 60 <= p <= 63, 21 <= s <= 38	bignumeric

Db2 for i Source Data Type	Google BigQuery Target Data Type
numeric(p,s), 61 <= p <= 63, 1 <= s <= 61	string
numeric(p,s), 61 <= p <= 63, 22 <= s <= 38	bignumeric
numeric(p,s), 62 <= p <= 63, 1 <= s <= 62	string
numeric(p,s), 62 <= p <= 63, 23 <= s <= 38	bignumeric
numeric(p,s), 9 <= p <= 38, 1 <= s <= 9	numeric
real	float64
rowid	bytes
smallint	int64
time	time
timestamp(precision), 0 <= p <= 6	datetime
timestamp(precision), 7 <= p <= 12	string
varbinary(size), 1 <= s <= 32740	bytes
varchar(size) for bit data, 1 <= s <= 32740	bytes
varchar(size), 1 <= s <= 32740	string

## Db2 for i Source and Microsoft Azure Synapse Analytics Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Db2 for i source and a Microsoft Azure Synapse Analytics target:

Db2 for i Source Data Type	Synapse Analytics Target Data Type
bigint	bigint
binary(size), 1 <= s <= 8000	binary(size), 1 <= s <= 8000
binary(size), 8001 <= s <= 32766	varbinary(max)
char(size) for bit data, 1 <= s <= 8000	binary(size), 1 <= s <= 8000
char(size) for bit data, 8001 <= s <= 32766	varbinary(max)
char(size), 1 <= s <= 32000	char(size), 4 <= s <= 32000
char(size), 32001 <= s <= 32766	nvarchar(max)
date	date
decimal(precision), 16 <= p <= 34	varchar(255)

Db2 for i Source Data Type	Synapse Analytics Target Data Type
decimal(p,s), 1 <= p <= 38, 0 <= s <= 38	decimal(p,s), 1 <= p <= 38, 0 <= s <= 38
decimal(p,s), 39 <= p <= 63, 0 <= s <= 62	char(size), 40 <= s <= 65
float	float
integer	int
long varbinary	varbinary(max)
long varchar	varchar(max)
long varchar for bit data	varbinary(max)
numeric(p,s), 1 <= p <= 38, 0 <= s <= 38	decimal(p,s), 1 <= p <= 38, 0 <= s <= 38
numeric(p,s), 39 <= p <= 63, 0 <= s <= 62	char(size), 40 <= s <= 65
real	real
rowid	varbinary(40)
smallint	smallint
time	time(0)
timestamp(precision), 0 <= p <= 7	datetime2(precision), 0 <= p <= 7
timestamp(precision), 8 <= p <= 12	char(size), 28 <= s <= 32
varbinary(size), 1 <= s <= 32740	varbinary(size), 1 <= s <= max
varchar(size) for bit data, 1 <= s <= 32740	varbinary(size), 1 <= s <= max
varchar(size), 1 <= s <= 32740	varchar(size), 4 <= s <= max

**Note:** The maximum supported length of DECFLOAT values on the target is 255 characters.

## Db2 for i Source and Oracle Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Db2 for i source and an Oracle target:

Db2 for i Source Data Type	Oracle Target Data Type
bigint	number(19)
binary(size), 1 <= s <= 32766	blob
char(size) for bit data, 1 <= s <= 32766	blob
char(size), 1 <= s <= 500	char(s char), 1 <= s <= 500

Db2 for i Source Data Type	Oracle Target Data Type
char( <i>size</i> ), 501 <= s <= 32766	clob
date	date
decfloat( <i>precision</i> ), 16 <= p <= 34	char(255 char)
decimal(p,s), 1 <= p <= 38, 0 <= s <= 38	number(p,s), 1 <= p <= 38, 0 <= s <= 38
decimal(p,s), 39 <= p <= 63, 0 <= s <= 62	char(s char), 40 <= s <= 65
float	binary_double
integer	number(10)
long varbinary	blob
long varchar	clob
long varchar for bit data	blob
numeric(p,s), 1 <= p <= 38, 0 <= s <= 38	number(p,s), 1 <= p <= 38, 0 <= s <= 38
numeric(p,s), 39 <= p <= 63, 0 <= s <= 62	char(s char), 40 <= s <= 65
real	binary_double
rowid	blob
smallint	number(5)
time	char(8 char)
timestamp(0)	date
timestamp( <i>precision</i> ), 1 <= p <= 9	timestamp( <i>precision</i> ), 1 <= p <= 9
timestamp( <i>precision</i> ), 10 <= p <= 12	char(s char), 30 <= s <= 32
varbinary( <i>size</i> ), 1 <= s <= 32740	blob
varchar( <i>size</i> ) for bit data, 1 <= s <= 32740	blob
varchar( <i>size</i> ), 1 <= s <= 401	char(s char), 1 <= s <= 401
varchar( <i>size</i> ), 501 <= s <= 32740	clob

**Note:** The maximum supported length of DECFLOAT values on the target is 255 characters.

## Db2 for i Source and Snowflake Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Db2 for i source and a Snowflake target:

Db2 for i Source Data Type	Snowflake Target Data Type
bigint	integer
binary(size), 1 <= s <= 32766	binary(size), 1 <= s <= 32766
char(size) for bit data, 1 <= s <= 32766	binary(size), 1 <= s <= 32766
char(size), 1 <= s <= 32766	char(size), 4 <= s <= 131064
date	date
decfloat(precision), 16 <= p <= 34	char(255)
decimal(p,s), 1 <= p <= 38, 0 <= s <= 37	number(p,s), 1 <= p <= 38, 0 <= s <= 37
decimal(p,s), 38 <= p <= 63, 1 <= s <= 62	char(size), 40 <= s <= 65
float	float
integer	integer
long varbinary	binary
long varchar	varchar
long varchar for bit data	binary
numeric(p,s), 1 <= p <= 38, 0 <= s <= 37	number(p,s), 1 <= p <= 38, 0 <= s <= 37
numeric(p,s), 38 <= p <= 63, 1 <= s <= 62	char(size), 40 <= s <= 65
real	float
rowid	binary(40)
smallint	integer
time	time(0)
timestamp(precision), 0 <= p <= 9	timestamp_ntz(precision), 0 <= p <= 9
timestamp(precision), 10 <= p <= 12	char(size), 30 <= s <= 32
varbinary(size), 1 <= s <= 32740	binary(size), 1 <= s <= 32740
varchar(size) for bit data, 1 <= s <= 32740	binary(size), 1 <= s <= 32740
varchar(size), 1 <= s <= 32740	varchar(size), 4 <= s <= 130960

**Note:** The maximum supported length of DECFLOAT values on the target is 255 characters.

## Db2 for Linux, UNIX, and Windows Source and Amazon Redshift Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Db2 for Linux, UNIX, and Windows (LUW) source and an Amazon Redshift target:

DB2 for Linux, UNIX, and Windows Source Data Type	Amazon Redshift Target Data Type
bigint	bigint
binary( <i>size</i> ), 1 <= <i>s</i> <= 255	character varying( <i>size</i> ), 2 <= <i>s</i> <= 510
boolean	boolean
char for bit data	character varying(2)
char( <i>size</i> ) for bit data, 2 <= <i>s</i> <= 255	character varying( <i>size</i> ), 4 <= <i>s</i> <= 510
character( <i>size</i> ), 1 <= <i>s</i> <= 255	character varying( <i>size</i> ), 4 <= <i>s</i> <= 1020
date	date
decfloat( <i>precision</i> ), 16 <= <i>p</i> <= 34	character varying(255)
decimal( <i>p,s</i> ), 1 <= <i>p</i> <= 31, 0 <= <i>s</i> <= 31	numeric( <i>p,s</i> ), 1 <= <i>p</i> <= 31, 0 <= <i>s</i> <= 31
double	double precision
integer	integer
long varchar	character varying(65535)
long varchar for bit data	character varying(65400)
real	real
smallint	smallint
time	time without time zone
timestamp( <i>precision</i> ), 0 <= <i>p</i> <= 6	timestamp without time zone
timestamp( <i>precision</i> ), 7 <= <i>p</i> <= 12	character varying( <i>size</i> ), 27 <= <i>s</i> <= 32
varbinary( <i>size</i> ), 1 <= <i>s</i> <= 32672	character varying( <i>size</i> ), 2 <= <i>s</i> <= 65344
varchar( <i>size</i> ) for bit data, 1 <= <i>s</i> <= 32672	character varying( <i>size</i> ), 2 <= <i>s</i> <= 65344
varchar( <i>size</i> ), 1 <= <i>s</i> <= 32672	character varying( <i>size</i> ), 4 <= <i>s</i> <= 65535

**Note:** The maximum supported length of DECFLOAT values on the target is 255 characters.

## DB2 for Linux, UNIX, and Windows Source and Databricks Delta Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a DB2 for Linux, UNIX, and Windows source and a Databricks Delta target:

DB2 for Linux, UNIX, and Windows Source Data Type	Databricks Delta Target Data Type
bigint	long
binary( <i>size</i> ), 1 <= s <= 255	binary
boolean	boolean
char for bit data	binary
char( <i>size</i> ) for bit data, 2 <= s <= 255	binary
character( <i>size</i> ), 1 <= s <= 255	string
date	string
decfloat( <i>precision</i> ), 16 <= p <= 34	string
decimal(p,s), 1 <= p <= 31, 0 <= s <= 31	decimal(p,s), 1 <= p <= 31, 0 <= s <= 31
double	double
integer	integer
long varchar	string
long varchar for bit data	binary
real	float
smallint	integer
time	string
timestamp( <i>precision</i> ), 0 <= p <= 6	timestamp
timestamp( <i>precision</i> ), 7 <= p <= 12	string
varbinary( <i>size</i> ), 1 <= s <= 32672	binary
varchar( <i>size</i> ) for bit data, 1 <= s <= 32672	binary
varchar( <i>size</i> ), 1 <= s <= 32672	string

## DB2 for Linux, UNIX, and Windows Source and Google BigQuery Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a DB2 for Linux, UNIX, and Windows source and a Google BigQuery target:

DB2 for Linux, UNIX, and Windows Source Data Type	Google BigQuery Target Data Type
bigint	int64
binary( <i>size</i> ), 1 <= s <= 255	bytes
boolean	bool
char for bit data	bytes
char( <i>size</i> ) for bit data, 2 <= s <= 255	bytes
character( <i>size</i> ), 1 <= s <= 255	string
date	date
decimal( <i>precision</i> ), 16 <= p <= 34	string
decimal	int64
decimal(1)	int64
decimal(1,1)	numeric
decimal(10)	int64
decimal(10,10)	bignumeric
decimal(11)	int64
decimal(11,s), 10 <= s <= 11	bignumeric
decimal(12)	int64
decimal(12,s), 10 <= s <= 12	bignumeric
decimal(13)	int64
decimal(13,s), 10 <= s <= 13	bignumeric
decimal(14)	int64
decimal(14,s), 10 <= s <= 14	bignumeric
decimal(15)	int64
decimal(15,s), 10 <= s <= 15	bignumeric
decimal(16)	int64



DB2 for Linux, UNIX, and Windows Source Data Type	Google BigQuery Target Data Type
decimal(16,s), 10 <= s <= 16	bignumeric
decimal(17)	int64
decimal(17,s), 10 <= s <= 17	bignumeric
decimal(18)	int64
decimal(18,s), 10 <= s <= 18	bignumeric
decimal(19,s), 10 <= s <= 19	bignumeric
decimal(2)	int64
decimal(2,s), 1 <= s <= 2	numeric
decimal(20,s), 10 <= s <= 20	bignumeric
decimal(21,s), 10 <= s <= 21	bignumeric
decimal(22,s), 10 <= s <= 22	bignumeric
decimal(23,s), 10 <= s <= 23	bignumeric
decimal(24,s), 10 <= s <= 24	bignumeric
decimal(25,s), 10 <= s <= 25	bignumeric
decimal(26,s), 10 <= s <= 26	bignumeric
decimal(27,s), 10 <= s <= 27	bignumeric
decimal(28,s), 10 <= s <= 28	bignumeric
decimal(3)	int64
decimal(3,s), 1 <= s <= 3	numeric
decimal(31,s), 10 <= s <= 31	bignumeric
decimal(31,s), 2 <= s <= 9	numeric
decimal(4)	int64
decimal(4,s), 1 <= s <= 4	numeric
decimal(5,s), 1 <= s <= 5	numeric
decimal(6)	int64
decimal(6,s), 1 <= s <= 6	numeric
decimal(7)	int64

DB2 for Linux, UNIX, and Windows Source Data Type	Google BigQuery Target Data Type
decimal(7,s), 1 <= s <= 7	numeric
decimal(8)	int64
decimal(8,s), 1 <= s <= 8	numeric
decimal(9)	int64
decimal(p,s), 10 <= p <= 31, 1 <= s <= 9	numeric
decimal(p,s), 11 <= p <= 31, 1 <= s <= 9	numeric
decimal(p,s), 12 <= p <= 31, 1 <= s <= 9	numeric
decimal(p,s), 13 <= p <= 31, 1 <= s <= 9	numeric
decimal(p,s), 14 <= p <= 31, 1 <= s <= 9	numeric
decimal(p,s), 15 <= p <= 31, 1 <= s <= 9	numeric
decimal(p,s), 16 <= p <= 31, 1 <= s <= 9	numeric
decimal(p,s), 17 <= p <= 31, 1 <= s <= 9	numeric
decimal(p,s), 18 <= p <= 31, 1 <= s <= 9	numeric
decimal(p,s), 19 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 20 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 21 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 22 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 23 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 24 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 25 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 26 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 27 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 28 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 29 <= p <= 30, 10 <= s <= 29	bignumeric
decimal(p,s), 29 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 30 <= p <= 31, 1 <= s <= 30	bignumeric
decimal(p,s), 30 <= p <= 31, 1 <= s <= 9	numeric

DB2 for Linux, UNIX, and Windows Source Data Type	Google BigQuery Target Data Type
decimal(p,s), 9 <= p <= 31, 1 <= s <= 9	numeric
double	float64
integer	int64
long varchar	string
long varchar for bit data	bytes
real	float64
smallint	int64
time	time
timestamp(precision), 0 <= p <= 6	datetime
timestamp(precision), 7 <= p <= 12	string
varbinary(size), 1 <= s <= 32672	bytes
varchar(size) for bit data, 1 <= s <= 32672	bytes
varchar(size), 1 <= s <= 32672	string

## DB2 for Linux, UNIX, and Windows Source and Microsoft Azure Synapse Analytics Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a DB2 for Linux, UNIX, and Windows source and a Microsoft Azure Synapse Analytics target:

DB2 for Linux, UNIX, and Windows Source Data Type	Synapse Analytics Target Data Type
bigint	bigint
binary(size), 1 <= s <= 255	binary(size), 1 <= s <= 255
boolean	bit
char for bit data	binary
char(size) for bit data, 2 <= s <= 255	binary(size), 2 <= s <= 255
character(size), 1 <= s <= 255	char(size), 4 <= s <= 1020
date	date

DB2 for Linux, UNIX, and Windows Source Data Type	Synapse Analytics Target Data Type
decfloat( <i>precision</i> ), 16 <= p <= 34	varchar(255)
decimal(p,s), 1 <= p <= 31, 0 <= s <= 31	decimal(p,s), 1 <= p <= 31, 0 <= s <= 31
double	float
integer	int
long varchar	varchar(max)
long varchar for bit data	varbinary(max)
real	real
smallint	smallint
time	time(0)
timestamp( <i>precision</i> ), 0 <= p <= 7	datetime2( <i>precision</i> ), 0 <= p <= 7
timestamp( <i>precision</i> ), 8 <= p <= 12	char( <i>size</i> ), 28 <= s <= 32
varbinary( <i>size</i> ), 1 <= s <= 32672	varbinary( <i>size</i> ), 1 <= s <= max
varchar( <i>size</i> ) for bit data, 1 <= s <= 32672	varbinary( <i>size</i> ), 1 <= s <= max
varchar( <i>size</i> ), 1 <= s <= 32672	varchar( <i>size</i> ), 4 <= s <= max

**Note:** The maximum supported length of DECFLOAT values on the target is 255 characters.

## DB2 for Linux, UNIX, and Windows Source and Snowflake Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a DB2 for Linux, UNIX, and Windows source and a Snowflake target:

DB2 for Linux, UNIX, and Windows Source Data Type	Snowflake Target Data Type
bigint	integer
binary( <i>size</i> ), 1 <= s <= 255	binary( <i>size</i> ), 1 <= s <= 255
boolean	boolean
char for bit data	binary(1)
char( <i>size</i> ) for bit data, 2 <= s <= 255	binary( <i>size</i> ), 2 <= s <= 255
character( <i>size</i> ), 1 <= s <= 255	char( <i>size</i> ), 4 <= s <= 1020
date	date

DB2 for Linux, UNIX, and Windows Source Data Type	Snowflake Target Data Type
decfloat( <i>precision</i> ), 16 <= p <= 34	char(255)
decimal(p,s), 1 <= p <= 31, 0 <= s <= 31	number(p,s), 1 <= p <= 31, 0 <= s <= 31
double	float
integer	integer
long varchar	varchar(130800)
long varchar for bit data	binary(32700)
real	float
smallint	integer
time	time(0)
timestamp( <i>precision</i> ), 0 <= p <= 9	timestamp_ntz( <i>precision</i> ), 0 <= p <= 9
timestamp( <i>precision</i> ), 10 <= p <= 12	char( <i>size</i> ), 30 <= s <= 32
varbinary( <i>size</i> ), 1 <= s <= 32672	binary( <i>size</i> ), 1 <= s <= 32672
varchar( <i>size</i> ) for bit data, 1 <= s <= 32672	binary( <i>size</i> ), 1 <= s <= 32672
varchar( <i>size</i> ), 1 <= s <= 32672	varchar( <i>size</i> ), 4 <= s <= 130688

**Note:** The maximum supported length of DECFLOAT values on the target is 255 characters.

## Db2 for z/OS Source and Amazon Redshift Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Db2 for z/OS source and an Amazon Redshift target:

Db2 for zOS Source Data Type	Amazon Redshift Target Data Type
bigint	bigint
binary( <i>size</i> ), 1 <= s <= 255	character varying( <i>size</i> ), 2 <= s <= 510
char for bit data	character varying(2)
char( <i>size</i> ) for bit data, 2 <= s <= 255	character varying( <i>size</i> ), 4 <= s <= 510
char( <i>size</i> ), 1 <= s <= 255	character varying( <i>size</i> ), 4 <= s <= 1020
date	date
decfloat( <i>precision</i> ), 16 <= p <= 34	character varying(255)

Db2 for zOS Source Data Type	Amazon Redshift Target Data Type
decimal(p,s), 1 <= p <= 31, 0 <= s <= 31	numeric(p,s), 1 <= p <= 31, 0 <= s <= 31
float	double precision
integer	integer
long varchar	character varying(65535)
long varchar for bit data	character varying(65408)
real	real
rowid	character varying(80)
smallint	smallint
time	time without time zone
timestamp( <i>precision</i> ) with time zone, 0 <= p <= 6	timestamp with time zone
timestamp( <i>precision</i> ) with time zone, 7 <= p <= 12	character varying( <i>size</i> ), 67 <= s <= 72
timestamp( <i>precision</i> ), 0 <= p <= 6	timestamp without time zone
timestamp( <i>precision</i> ), 7 <= p <= 12	character varying( <i>size</i> ), 27 <= s <= 32
varbinary( <i>size</i> ), 1 <= s <= 32704	character varying( <i>size</i> ), 2 <= s <= 65408
varchar( <i>size</i> ) for bit data, 1 <= s <= 32704	character varying( <i>size</i> ), 2 <= s <= 65408
varchar( <i>size</i> ), 1 <= s <= 32704	character varying( <i>size</i> ), 4 <= s <= 65535

**Note:** The maximum supported length of DECFLOAT values on the target is 255 characters.

## Db2 for zOS Source and Databricks Delta Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Db2 for zOS source and a Databricks Delta target:

Db2 for zOS Source Data Type	Databricks Delta Target Data Type
bigint	long
binary( <i>size</i> ), 1 <= s <= 255	binary
char for bit data	binary
char( <i>size</i> ) for bit data, 2 <= s <= 255	binary
char( <i>size</i> ), 1 <= s <= 255	string

Db2 for zOS Source Data Type	Databricks Delta Target Data Type
date	string
decfloat( <i>precision</i> ), 16 <= p <= 34	string
decimal(p,s), 1 <= p <= 31, 0 <= s <= 31	decimal(p,s), 1 <= p <= 31, 0 <= s <= 31
float	double
integer	integer
long varchar	string
long varchar for bit data	binary
real	float
rowid	binary
smallint	integer
time	string
timestamp( <i>precision</i> ) with time zone, 0 <= p <= 12	string
timestamp( <i>precision</i> ), 0 <= p <= 6	timestamp
timestamp( <i>precision</i> ), 7 <= p <= 12	string
varbinary( <i>size</i> ), 1 <= s <= 32704	binary
varchar( <i>size</i> ) for bit data, 1 <= s <= 32704	binary
varchar( <i>size</i> ), 1 <= s <= 32704	string

## Db2 for zOS Source and Google BigQuery Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Db2 for zOS source and a Google BigQuery target:

Db2 for zOS Source Data Type	Google BigQuery Target Data Type
bigint	int64
binary( <i>size</i> ), 1 <= s <= 255	bytes
char for bit data	bytes
char( <i>size</i> ) for bit data, 2 <= s <= 255	bytes
char( <i>size</i> ), 1 <= s <= 255	string

Db2 for zOS Source Data Type	Google BigQuery Target Data Type
date	date
decfloat( <i>precision</i> ), 16 <= p <= 34	string
decimal	int64
decimal(1)	int64
decimal(1,1)	numeric
decimal(10)	int64
decimal(10,10)	bignumeric
decimal(11)	int64
decimal(11,s), 10 <= s <= 11	bignumeric
decimal(12)	int64
decimal(12,s), 10 <= s <= 12	bignumeric
decimal(13)	int64
decimal(13,s), 10 <= s <= 13	bignumeric
decimal(14)	int64
decimal(14,s), 10 <= s <= 14	bignumeric
decimal(15)	int64
decimal(15,s), 10 <= s <= 15	bignumeric
decimal(16)	int64
decimal(16,s), 10 <= s <= 16	bignumeric
decimal(17)	int64
decimal(17,s), 10 <= s <= 17	bignumeric
decimal(18)	int64
decimal(18,s), 10 <= s <= 18	bignumeric
decimal(19,s), 10 <= s <= 19	bignumeric
decimal(2)	int64
decimal(2,s), 1 <= s <= 2	numeric
decimal(20,s), 10 <= s <= 20	bignumeric



Db2 for zOS Source Data Type	Google BigQuery Target Data Type
decimal(21,s), 10 <= s <= 21	bignumeric
decimal(22,s), 10 <= s <= 22	bignumeric
decimal(23,s), 10 <= s <= 23	bignumeric
decimal(24,s), 10 <= s <= 24	bignumeric
decimal(25,s), 10 <= s <= 25	bignumeric
decimal(26,s), 10 <= s <= 26	bignumeric
decimal(27,s), 10 <= s <= 27	bignumeric
decimal(28,s), 10 <= s <= 28	bignumeric
decimal(3)	int64
decimal(3,s), 1 <= s <= 3	numeric
decimal(31,s), 10 <= s <= 31	bignumeric
decimal(31,s), 2 <= s <= 9	numeric
decimal(4)	int64
decimal(4,s), 1 <= s <= 4	numeric
decimal(5,s), 1 <= s <= 5	numeric
decimal(6)	int64
decimal(6,s), 1 <= s <= 6	numeric
decimal(7)	int64
decimal(7,s), 1 <= s <= 7	numeric
decimal(8)	int64
decimal(8,s), 1 <= s <= 8	numeric
decimal(9)	int64
decimal(p,s), 10 <= p <= 31, 1 <= s <= 9	numeric
decimal(p,s), 11 <= p <= 31, 1 <= s <= 9	numeric
decimal(p,s), 12 <= p <= 31, 1 <= s <= 9	numeric
decimal(p,s), 13 <= p <= 31, 1 <= s <= 9	numeric
decimal(p,s), 14 <= p <= 31, 1 <= s <= 9	numeric

Db2 for zOS Source Data Type	Google BigQuery Target Data Type
decimal(p,s), 15 <= p <= 31, 1 <= s <= 9	numeric
decimal(p,s), 16 <= p <= 31, 1 <= s <= 9	numeric
decimal(p,s), 17 <= p <= 31, 1 <= s <= 9	numeric
decimal(p,s), 18 <= p <= 31, 1 <= s <= 9	numeric
decimal(p,s), 19 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 20 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 21 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 22 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 23 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 24 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 25 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 26 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 27 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 28 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 29 <= p <= 30, 10 <= s <= 29	bignumeric
decimal(p,s), 29 <= p <= 31, 0 <= s <= 9	numeric
decimal(p,s), 30 <= p <= 31, 1 <= s <= 30	bignumeric
decimal(p,s), 30 <= p <= 31, 1 <= s <= 9	numeric
decimal(p,s), 9 <= p <= 31, 1 <= s <= 9	numeric
float	float64
integer	int64
long varchar	string
long varchar for bit data	bytes
real	float64
rowid	bytes
smallint	int64
time	time

Db2 for z/OS Source Data Type	Google BigQuery Target Data Type
timestamp( <i>precision</i> ) with time zone, 0 <= p <= 6	timestamp
timestamp( <i>precision</i> ) with time zone, 7 <= p <= 12	string
timestamp( <i>precision</i> ), 0 <= p <= 6	datetime
timestamp( <i>precision</i> ), 7 <= p <= 12	string
varbinary( <i>size</i> ), 1 <= s <= 32704	bytes
varchar( <i>size</i> ) for bit data, 1 <= s <= 32704	bytes
varchar( <i>size</i> ), 1 <= s <= 32704	string

## Db2 for z/OS Source and Microsoft Azure Synapse Analytics Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Db2 for z/OS source and a Microsoft Azure Synapse Analytics target:

Db2 for z/OS Source Data Type	Synapse Analytics Target Data Type
bigint	bigint
binary( <i>size</i> ), 1 <= s <= 255	binary( <i>size</i> ), 1 <= s <= 255
char for bit data	binary
char( <i>size</i> ) for bit data, 2 <= s <= 255	binary( <i>size</i> ), 2 <= s <= 255
char( <i>size</i> ), 1 <= s <= 255	char( <i>size</i> ), 4 <= s <= 1020
date	date
decimal( <i>precision</i> ), 16 <= p <= 34	varchar(255)
decimal(p,s), 1 <= p <= 31, 0 <= s <= 31	decimal(p,s), 1 <= p <= 31, 0 <= s <= 31
float	float
integer	int
long varchar	varchar(max)
long varchar for bit data	varbinary(max)
real	real
rowid	varbinary(40)

Db2 for z/OS Source Data Type	Synapse Analytics Target Data Type
smallint	smallint
time	time(0)
timestamp( <i>precision</i> ) with time zone, 0 <= p <= 7	datetimeoffset( <i>precision</i> ), 0 <= p <= 7
timestamp( <i>precision</i> ) with time zone, 8 <= p <= 12	char( <i>size</i> ), 68 <= s <= 72
timestamp( <i>precision</i> ), 0 <= p <= 7	datetime2( <i>precision</i> ), 0 <= p <= 7
timestamp( <i>precision</i> ), 8 <= p <= 12	char( <i>size</i> ), 28 <= s <= 32
varbinary( <i>size</i> ), 1 <= s <= 32704	varbinary( <i>size</i> ), 1 <= s <= max
varchar( <i>size</i> ) for bit data, 1 <= s <= 32704	varbinary( <i>size</i> ), 1 <= s <= max
varchar( <i>size</i> ), 1 <= s <= 32704	varchar( <i>size</i> ), 4 <= s <= max

**Note:** The maximum supported length of DECFLOAT values on the target is 255 characters.

## Db2 for z/OS Source and Snowflake Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Db2 for z/OS source and a Snowflake target:

Db2 for z/OS Source Data Type	Snowflake Target Data Type
bigint	integer
binary( <i>size</i> ), 1 <= s <= 255	binary( <i>size</i> ), 1 <= s <= 255
char for bit data	binary(1)
char( <i>size</i> ) for bit data, 2 <= s <= 255	binary( <i>size</i> ), 2 <= s <= 255
char( <i>size</i> ), 1 <= s <= 255	char( <i>size</i> ), 4 <= s <= 1020
date	date
decfloat( <i>precision</i> ), 16 <= p <= 34	char(255)
decimal(p,s), 1 <= p <= 31, 0 <= s <= 31	number(p,s), 1 <= p <= 31, 0 <= s <= 31
float	float
integer	integer
long varchar	varchar(130816)
long varchar for bit data	binary(32704)

Db2 for zOS Source Data Type	Snowflake Target Data Type
real	float
rowid	binary(40)
smallint	integer
time	time(0)
timestamp( <i>precision</i> ) with time zone, 0 <= p <= 9	timestamp_tz( <i>precision</i> ), 0 <= p <= 9
timestamp( <i>precision</i> ) with time zone, 10 <= p <= 12	char( <i>size</i> ), 70 <= s <= 72
timestamp( <i>precision</i> ), 0 <= p <= 9	timestamp_ntz( <i>precision</i> ), 0 <= p <= 9
timestamp( <i>precision</i> ), 10 <= p <= 12	char( <i>size</i> ), 30 <= s <= 32
varbinary( <i>size</i> ), 1 <= s <= 32704	binary( <i>size</i> ), 1 <= s <= 32704
varchar( <i>size</i> ) for bit data, 1 <= s <= 32704	binary( <i>size</i> ), 1 <= s <= 32704
varchar( <i>size</i> ), 1 <= s <= 32704	varchar( <i>size</i> ), 4 <= s <= 130816

**Note:** The maximum supported length of DECFLOAT values on the target is 255 characters.

## Microsoft SQL Server or Azure SQL Database Source and Amazon Redshift Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Microsoft SQL Server or Azure SQL Database source and an Amazon Redshift target:

Microsoft SQL Server Source Data Type	Amazon Redshift Target Data Type
bigint	bigint
binary( <i>size</i> ), 1 <= s <= 8000	character varying( <i>size</i> ), 2 <= s <= 16000
bit	boolean
char( <i>size</i> ), 1 <= s <= 8000	character varying( <i>size</i> ), 1 <= s <= 8000
date	date
datetime	timestamp without time zone
datetime2(7)	character varying(27)
datetime2( <i>precision</i> ), 0 <= p <= 6	timestamp without time zone
datetimeoffset(7)	character varying(67)

Microsoft SQL Server Source Data Type	Amazon Redshift Target Data Type
datetimeoffset( <i>precision</i> ), 0 <= p <= 6	timestamp with time zone
decimal(38,38)	character varying(41)
decimal(p,s), 1 <= p <= 38, 0 <= s <= 37	numeric(p,s), 1 <= p <= 38, 0 <= s <= 37
float	double precision
geography	character varying(65535)
geometry	character varying(65535)
hierarchyid	character varying(1784)
image	character varying(65535)
int	integer
money	numeric(20,4)
nchar( <i>size</i> ), 1 <= s <= 4000	character varying( <i>size</i> ), 1 <= s <= 8000
ntext	character varying(65535)
numeric(38,38)	character varying(41)
numeric(p,s), 1 <= p <= 38, 0 <= s <= 37	numeric(p,s), 1 <= p <= 38, 0 <= s <= 37
nvarchar( <i>size</i> ), 1 <= s <= 4000	character varying( <i>size</i> ), 1 <= s <= 8000
real	real
smalldatetime	timestamp without time zone
smallint	smallint
smallmoney	numeric(10,4)
sql_variant	character varying(16032)
text	character varying(65535)
time( <i>precision</i> ), 0 <= p <= 7	time without time zone
timestamp(8)	character varying(16)
tinyint	smallint
uniqueidentifier	character(36)
varbinary( <i>size</i> ), 1 <= s <= 8000	character varying( <i>size</i> ), 2 <= s <= 16000

Microsoft SQL Server Source Data Type	Amazon Redshift Target Data Type
varchar(size), 1 <= s <= 8000	character varying(size), 1 <= s <= 8000
xml	character varying(65535)

## Microsoft SQL Server Source and Databricks Delta Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Microsoft SQL Server source and a Databricks Delta target:

Microsoft SQL Server Source Data Type	Databricks Delta Target Data Type
bigint	long
binary(size), 1 <= s <= 8000	binary
bit	boolean
char(size), 1 <= s <= 8000	string
date	string
datetime	timestamp
datetime2(7)	string
datetime2(precision), 0 <= p <= 6	timestamp
datetimeoffset(7)	string
datetimeoffset(precision), 0 <= p <= 6	timestamp
decimal(p,s), 1 <= p <= 38, 0 <= s <= 38	decimal(p,s), 1 <= p <= 38, 0 <= s <= 38
float	double
geography	binary
geometry	binary
hierarchyid	binary
image	binary
int	integer
money	decimal(19,4)
nchar(size), 1 <= s <= 4000	string
ntext	string
numeric(p,s), 1 <= p <= 38, 0 <= s <= 38	decimal(p,s), 1 <= p <= 38, 0 <= s <= 38

Microsoft SQL Server Source Data Type	Databricks Delta Target Data Type
nvarchar(size), 1 <= s <= 4000	string
real	float
smalldatetime	timestamp
smallint	integer
smallmoney	decimal(10,4)
sql_variant	binary
text	string
time(precision), 0 <= p <= 7	string
timestamp(8)	binary
tinyint	integer
uniqueidentifier	string
varbinary(size), 1 <= s <= 8000	binary
varchar(size), 1 <= s <= 8000	string
xml	string

## Microsoft SQL Server Source and Google BigQuery Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Microsoft SQL Server source and a Google BigQuery target:

Microsoft SQL Server Source Data Type	Google BigQuery Target Data Type
bigint	int64
binary(size), 1 <= s <= 8000	bytes
bit	bool
char(size), 1 <= s <= 8000	string
date	date
datetime	datetime
datetime2(7)	string
datetime2(precision), 0 <= p <= 6	datetime
datetimeoffset(7)	string



Microsoft SQL Server Source Data Type	Google BigQuery Target Data Type
datetimeoffset( <i>precision</i> ), 0 <= p <= 6	timestamp
decimal(1,1)	numeric
decimal(10,10)	bignumeric
decimal(11,s), 10 <= s <= 11	bignumeric
decimal(12,s), 10 <= s <= 12	bignumeric
decimal(13,s), 10 <= s <= 13	bignumeric
decimal(14,s), 10 <= s <= 14	bignumeric
decimal(15,s), 10 <= s <= 15	bignumeric
decimal(16,s), 10 <= s <= 16	bignumeric
decimal(17,s), 10 <= s <= 17	bignumeric
decimal(18,s), 10 <= s <= 18	bignumeric
decimal(19,s), 10 <= s <= 19	bignumeric
decimal(2,s), 1 <= s <= 2	numeric
decimal(20,s), 10 <= s <= 20	bignumeric
decimal(21,s), 10 <= s <= 21	bignumeric
decimal(22,s), 10 <= s <= 22	bignumeric
decimal(23,s), 10 <= s <= 23	bignumeric
decimal(24,s), 10 <= s <= 24	bignumeric
decimal(25,s), 10 <= s <= 25	bignumeric
decimal(26,s), 10 <= s <= 26	bignumeric
decimal(27,s), 10 <= s <= 27	bignumeric
decimal(28,s), 10 <= s <= 28	bignumeric
decimal(3,s), 1 <= s <= 3	numeric
decimal(38,s), 10 <= s <= 38	bignumeric
decimal(4,s), 1 <= s <= 4	numeric
decimal(5,s), 1 <= s <= 5	numeric
decimal(6,s), 1 <= s <= 6	numeric

Microsoft SQL Server Source Data Type	Google BigQuery Target Data Type
decimal(7,s), 1 <= s <= 7	numeric
decimal(8,s), 1 <= s <= 8	numeric
decimal(p,1), 18 <= p <= 0	int64
decimal(p,s), 19 <= p <= 29, 0 <= s <= 9	numeric
decimal(p,s), 29 <= p <= 30, 0 <= s <= 29	bignumeric
decimal(p,s), 30 <= p <= 31, 0 <= s <= 30	bignumeric
decimal(p,s), 31 <= p <= 32, 0 <= s <= 31	bignumeric
decimal(p,s), 32 <= p <= 33, 0 <= s <= 32	bignumeric
decimal(p,s), 33 <= p <= 34, 0 <= s <= 33	bignumeric
decimal(p,s), 34 <= p <= 35, 0 <= s <= 34	bignumeric
decimal(p,s), 35 <= p <= 36, 0 <= s <= 35	bignumeric
decimal(p,s), 36 <= p <= 37, 0 <= s <= 36	bignumeric
decimal(p,s), 37 <= p <= 38, 0 <= s <= 37	bignumeric
decimal(p,s), 9 <= p <= 38, 1 <= s <= 9	numeric
float	float64
geography	bytes
geometry	bytes
hierarchyid	bytes
image	bytes
int	int64
money	numeric
nchar(size), 1 <= s <= 4000	string
ntext	string
numeric(1,1)	numeric
numeric(10,10)	bignumeric
numeric(11,s), 10 <= s <= 11	bignumeric
numeric(12,s), 10 <= s <= 12	bignumeric

Microsoft SQL Server Source Data Type	Google BigQuery Target Data Type
numeric(13,s), 10 <= s <= 13	bignumeric
numeric(14,s), 10 <= s <= 14	bignumeric
numeric(15,s), 10 <= s <= 15	bignumeric
numeric(16,s), 10 <= s <= 16	bignumeric
numeric(17,s), 10 <= s <= 17	bignumeric
numeric(18,s), 10 <= s <= 18	bignumeric
numeric(19,s), 10 <= s <= 19	bignumeric
numeric(2,s), 1 <= s <= 2	numeric
numeric(20,s), 10 <= s <= 20	bignumeric
numeric(21,s), 10 <= s <= 21	bignumeric
numeric(22,s), 10 <= s <= 22	bignumeric
numeric(23,s), 10 <= s <= 23	bignumeric
numeric(24,s), 10 <= s <= 24	bignumeric
numeric(25,s), 10 <= s <= 25	bignumeric
numeric(26,s), 10 <= s <= 26	bignumeric
numeric(27,s), 10 <= s <= 27	bignumeric
numeric(28,s), 10 <= s <= 28	bignumeric
numeric(3,s), 1 <= s <= 3	numeric
numeric(38,s), 10 <= s <= 38	bignumeric
numeric(4,s), 1 <= s <= 4	numeric
numeric(5,s), 1 <= s <= 5	numeric
numeric(6,s), 1 <= s <= 6	numeric
numeric(7,s), 1 <= s <= 7	numeric
numeric(8,s), 1 <= s <= 8	numeric
numeric(p,1), 18 <= p <= 0	int64
numeric(p,s), 28 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 29 <= p <= 30, 0 <= s <= 29	bignumeric

Microsoft SQL Server Source Data Type	Google BigQuery Target Data Type
numeric(p,s), 30 <= p <= 31, 0 <= s <= 30	bignumeric
numeric(p,s), 31 <= p <= 32, 0 <= s <= 31	bignumeric
numeric(p,s), 32 <= p <= 33, 0 <= s <= 32	bignumeric
numeric(p,s), 33 <= p <= 34, 0 <= s <= 33	bignumeric
numeric(p,s), 34 <= p <= 35, 0 <= s <= 34	bignumeric
numeric(p,s), 35 <= p <= 36, 0 <= s <= 35	bignumeric
numeric(p,s), 36 <= p <= 37, 0 <= s <= 36	bignumeric
numeric(p,s), 37 <= p <= 38, 0 <= s <= 37	bignumeric
numeric(p,s), 9 <= p <= 38, 1 <= s <= 9	numeric
nvarchar(size), 1 <= s <= 4000	string
real	float64
smalldatetime	datetime
smallint	int64
smallmoney	numeric
sql_variant	bytes
text	string
time(7)	string
time(precision), 0 <= p <= 6	time
timestamp(8)	bytes
tinyint	int64
uniqueidentifier	string
varbinary(size), 1 <= s <= 8000	bytes
varchar(size), 1 <= s <= 8000	string
xml	string

## Microsoft SQL Server or Azure SQL Database Source and Microsoft Azure Synapse Analytics Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Microsoft SQL Server or Azure SQL Database source and a Microsoft Azure Synapse Analytics target:

Microsoft SQL Server Source Data Type	Synapse Analytics Target Data Type
bigint	bigint
binary(size), 1 <= s <= 8000	binary(size), 1 <= s <= 8000
bit	bit
char(size), 1 <= s <= 8000	char(size), 1 <= s <= 8000
date	date
datetime	datetime2(3)
datetime2(precision), 0 <= p <= 7	datetime2(precision), 0 <= p <= 7
datetimeoffset(precision), 0 <= p <= 7	datetimeoffset(precision), 0 <= p <= 7
decimal(p,s), 1 <= p <= 38, 0 <= s <= 38	decimal(p,s), 1 <= p <= 38, 0 <= s <= 38
float	float
geography	varbinary(max)
geometry	varbinary(max)
hierarchyid	varbinary(892)
image	varbinary(max)
int	int
money	money
nchar(size), 1 <= s <= 4000	nchar(size), 1 <= s <= 4000
ntext	nvarchar(max)
numeric(p,s), 1 <= p <= 38, 0 <= s <= 38	numeric(p,s), 1 <= p <= 38, 0 <= s <= 38
nvarchar(size), 1 <= s <= 4000	nvarchar(size), 1 <= s <= 4000
real	real
smalldatetime	datetime2(0)
smallint	smallint
smallmoney	smallmoney

Microsoft SQL Server Source Data Type	Synapse Analytics Target Data Type
sql_variant	varbinary(max)
text	varchar(max)
time( <i>precision</i> ), 0 <= p <= 7	time( <i>precision</i> ), 0 <= p <= 7
timestamp(8)	varbinary(8)
tinyint	tinyint
uniqueidentifier	uniqueidentifier
varbinary( <i>size</i> ), 1 <= s <= 8000	varbinary( <i>size</i> ), 1 <= s <= 8000
varchar( <i>size</i> ), 1 <= s <= 8000	varchar( <i>size</i> ), 1 <= s <= 8000
xml	varchar(max)

## Microsoft SQL Server Source and Oracle Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Microsoft SQL Server or Azure SQL Database source and an Oracle target:

Microsoft SQL Server Source Data Type	Oracle Target Data Type
bigint	number(19)
binary( <i>size</i> ), 1 <= s <= 2000	raw( <i>size</i> ), 1 <= s <= 2000
binary( <i>size</i> ), 2001 <= s <= 8000	blob
bit	char(1 char)
char( <i>size</i> ), 1 <= s <= 2000	char(s char), 1 <= s <= 2000
char( <i>size</i> ), 2001 <= s <= 4000	varchar2(s char), 2001 <= s <= 4000
char( <i>size</i> ), 4001 <= s <= 8000	clob
date	date
datetime	timestamp(3)
datetime2(0)	date
datetime2( <i>precision</i> ), 1 <= p <= 7	timestamp( <i>precision</i> ), 1 <= p <= 7
datetimeoffset( <i>precision</i> ), 0 <= p <= 7	timestamp( <i>precision</i> ) with time zone, 0 <= p <= 7
decimal(p,s), 1 <= p <= 38, 0 <= s <= 38	number(p,s), 1 <= p <= 38, 0 <= s <= 38
float	binary_double

Microsoft SQL Server Source Data Type	Oracle Target Data Type
geography	blob
geometry	blob
hierarchyid	blob
image	blob
int	number(10)
money	number(19,4)
nchar(size), 1 <= s <= 1000	nchar(s char), 1 <= s <= 1000
nchar(size), 1001 <= s <= 2000	nvarchar2(s char), 1001 <= s <= 2000
nchar(size), 2001 <= s <= 4000	nclob
ntext	nclob
numeric(p,s), 1 <= p <= 38, 0 <= s <= 38	number(p,s), 1 <= p <= 38, 0 <= s <= 38
nvarchar(size), 1 <= s <= 2000	nvarchar2(s char), 1 <= s <= 2000
nvarchar(size), 2001 <= s <= 4000	nclob
real	binary_float
smalldatetime	date
smallint	number(5)
smallmoney	number(10,4)
sql_variant	blob
text	clob
time(precision), 0 <= p <= 7	timestamp(precision), 0 <= p <= 7
timestamp(8)	raw(8)
tinyint	number(3)
uniqueidentifier	char(36 char)
varbinary(size), 1 <= s <= 2000	raw(size), 1 <= s <= 2000
varbinary(size), 2001 <= s <= 8000	blob
varchar(size), 1 <= s <= 4000	varchar2(s char), 1 <= s <= 4000

Microsoft SQL Server Source Data Type	Oracle Target Data Type
varchar( <i>size</i> ), 4001 <= s <= 8000	clob
xml	clob

## Microsoft SQL Server or Azure SQL Database Source and Snowflake Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Microsoft SQL Server or Azure SQL Database source and a Snowflake target:

Microsoft SQL Server Source Data Type	Snowflake Target Data Type
bigint	integer
binary( <i>size</i> ), 1 <= s <= 8000	binary( <i>size</i> ), 1 <= s <= 8000
bit	boolean
char( <i>size</i> ), 1 <= s <= 8000	varchar( <i>size</i> ), 1 <= s <= 8000
date	date
datetime	datetime(3)
datetime2( <i>precision</i> ), 0 <= p <= 7	datetime( <i>precision</i> ), 0 <= p <= 7
datetimeoffset( <i>precision</i> ), 0 <= p <= 7	timestamp_tz( <i>precision</i> ), 0 <= p <= 7
decimal(38,38)	char(41)
decimal(p,s), 1 <= p <= 38, 0 <= s <= 37	number(p,s), 1 <= p <= 38, 0 <= s <= 37
float	float
geography	binary
geometry	binary
hierarchyid	binary(892)
image	binary
int	integer
money	number(19,4)
nchar( <i>size</i> ), 1 <= s <= 4000	varchar( <i>size</i> ), 4 <= s <= 16000
ntext	varchar
numeric(38,38)	char(41)



Microsoft SQL Server Source Data Type	Snowflake Target Data Type
numeric(p,s), 1 <= p <= 38, 0 <= s <= 37	number(p,s), 1 <= p <= 38, 0 <= s <= 37
nvarchar(size), 1 <= s <= 4000	varchar(size), 4 <= s <= 16000
real	float
smalldatetime	datetime(0)
smallint	integer
smallmoney	number(10,4)
sql_variant	binary(8016)
text	varchar
time(precision), 0 <= p <= 7	time(precision), 0 <= p <= 7
timestamp(8)	binary(8)
tinyint	integer
uniqueidentifier	char(36)
varbinary(size) 1 <= s <= 8000	binary(size), 1 <= s <= 8000
varchar(size), 1 <= s <= 8000	varchar(size), 1 <= s <= 8000
xml	varchar

## MySQL Source and Amazon Redshift Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a MySQL source and an Amazon Redshift target:

MySQL Source Data Type	Amazon Redshift Target Data Type
bigint	bigint
bigint unsigned	numeric(20,0)
binary(size), 1 <= s <= 255	character varying(size), 2 <= s <= 510
bit(precision), 1 <= p <= 64	character varying(size), 2 <= s <= 16
blob	character varying(65535)
char(size), 1 <= s <= 255	character varying(size), 4 <= s <= 1020
date	date
datetime	timestamp without time zone

MySQL Source Data Type	Amazon Redshift Target Data Type
decimal(p,s), 1 <= p <= 38, 0 <= s <= 29	numeric(p,s), 1 <= p <= 38, 0 <= s <= 29
decimal(p,s), 39 <= p <= 65, 0 <= s <= 29	character varying(size), 40 <= s <= 67
double	double precision
float	real
geomcollection	character varying(65535)
geometry	character varying(65535)
geometrycollection	character varying(65535)
int	integer
int unsigned	bigint
json	character varying(65535)
linestring	character varying(65535)
longblob	character varying(65535)
longtext	character varying(65535)
mediumblob	character varying(65535)
mediumint	integer
mediumint unsigned	integer
mediumtext	character varying(65535)
multilinestring	character varying(65535)
multipoint	character varying(65535)
multipolygon	character varying(65535)
numeric	numeric(10,0)
point	character varying(65535)
polygon	character varying(65535)
smallint	smallint
smallint unsigned	integer
text	character varying(65535)
time(precision), 0 <= p <= 6	character varying(size), 10 <= s <= 17

MySQL Source Data Type	Amazon Redshift Target Data Type
timestamp( <i>precision</i> ), 0 <= p <= 6	timestamp without time zone
tinyblob	character varying(510)
tinyint	smallint
tinyint unsigned	smallint
tinytext	character varying(256)
varbinary( <i>size</i> ), 1 <= s <= 65535	character varying( <i>size</i> ), 2 <= s <= 65535
varchar( <i>size</i> ), 1 <= s <= 21844	character varying( <i>size</i> ), 4 <= s <= 65535
year	smallint

## MySQL Source and Databricks Delta Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a MySQL source and a Databricks Delta target:

MySQL Source Data Type	Databricks Delta Target Data Type
bigint	long
bigint unsigned	decimal(20)
binary( <i>size</i> ), 1 <= s <= 255	binary
bit( <i>precision</i> ), 1 <= p <= 64	binary
blob	binary
char( <i>size</i> ), 1 <= s <= 255	string
date	string
datetime	timestamp
decimal(p,s), 1 <= p <= 38, 0 <= s <= 29	decimal(p,s), 1 <= p <= 38, 0 <= s <= 29
decimal(p,s), 39 <= p <= 65, 0 <= s <= 29	string
double	double
float	float
geomcollection	binary
geometry	binary
geometrycollection	binary

MySQL Source Data Type	Databricks Delta Target Data Type
int	integer
int unsigned	long
json	string
linestring	binary
longblob	binary
longtext	string
mediumblob	binary
mediumint	integer
mediumint unsigned	integer
mediumtext	string
multilinestring	binary
multipoint	binary
multipolygon	binary
numeric	decimal
point	binary
polygon	binary
smallint	integer
smallint unsigned	integer
text	string
time( <i>precision</i> ), 0 <= p <= 6	string
timestamp( <i>precision</i> ), 0 <= p <= 6	timestamp
tinyblob	binary
tinyint	integer
tinyint unsigned	integer
tinytext	string
varbinary( <i>size</i> ), 1 <= s <= 65535	binary

MySQL Source Data Type	Databricks Delta Target Data Type
<code>varchar(size), 1 &lt;= s &lt;= 21844</code>	string
year	integer

## MySQL Source and Google BigQuery Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a MySQL source and a Google BigQuery target:

MySQL Source Data Type	Google BigQuery Target Data Type
bigint	int64
bigint unsigned	bignumeric
<code>binary(size), 1 &lt;= s &lt;= 255</code>	bytes
<code>bit(precision), 1 &lt;= p &lt;= 64</code>	bytes
blob	bytes
<code>char(size), 1 &lt;= s &lt;= 255</code>	string
date	date
datetime	datetime
decimal	int64
decimal(1)	int64
decimal(1,1)	numeric
decimal(10,10)	bignumeric
decimal(11)	int64
decimal(11,s), 10 <= s <= 11	bignumeric
decimal(12)	int64
decimal(12,s), 10 <= s <= 12	bignumeric
decimal(13)	int64
decimal(13,s), 10 <= s <= 13	bignumeric
decimal(14)	int64
decimal(14,s), 10 <= s <= 14	bignumeric
decimal(15)	int64

MySQL Source Data Type	Google BigQuery Target Data Type
decimal(15,s), 10 <= s <= 15	bignumeric
decimal(16)	int64
decimal(16,s), 10 <= s <= 16	bignumeric
decimal(17)	int64
decimal(17,s), 10 <= s <= 17	bignumeric
decimal(18)	int64
decimal(18,s), 10 <= s <= 18	bignumeric
decimal(19,s), 10 <= s <= 19	bignumeric
decimal(2)	int64
decimal(2,s), 1 <= s <= 2	numeric
decimal(20,s), 10 <= s <= 20	bignumeric
decimal(21,s), 10 <= s <= 21	bignumeric
decimal(22,s), 10 <= s <= 22	bignumeric
decimal(23,s), 10 <= s <= 23	bignumeric
decimal(24,s), 10 <= s <= 24	bignumeric
decimal(25,s), 10 <= s <= 25	bignumeric
decimal(26,s), 10 <= s <= 26	bignumeric
decimal(27,s), 10 <= s <= 27	bignumeric
decimal(28,s), 10 <= s <= 28	bignumeric
decimal(3)	int64
decimal(3,s), 1 <= s <= 3	numeric
decimal(38,9)	numeric
decimal(4)	int64
decimal(4,s), 1 <= s <= 4	numeric
decimal(40)	string
decimal(40,s), 1 <= s <= 21	bignumeric
decimal(41,s), 0 <= s <= 1	string

MySQL Source Data Type	Google BigQuery Target Data Type
decimal(41,s), 2 <= s <= 22	bignumeric
decimal(42,s), 0 <= s <= 2	string
decimal(42,s), 3 <= s <= 23	bignumeric
decimal(43,s), 0 <= s <= 3	string
decimal(43,s), 4 <= s <= 24	bignumeric
decimal(44,s), 0 <= s <= 4	string
decimal(44,s), 5 <= s <= 25	bignumeric
decimal(45,s), 0 <= s <= 5	string
decimal(45,s), 6 <= s <= 26	bignumeric
decimal(46,s), 0 <= s <= 6	string
decimal(46,s), 7 <= s <= 27	bignumeric
decimal(47,s), 0 <= s <= 7	string
decimal(47,s), 8 <= s <= 28	bignumeric
decimal(48,s), 0 <= s <= 8	string
decimal(49,s), 0 <= s <= 9	string
decimal(5)	int64
decimal(5,s), 1 <= s <= 5	numeric
decimal(50,s), 0 <= s <= 10	string
decimal(51,s), 0 <= s <= 11	string
decimal(52,s), 0 <= s <= 12	string
decimal(53,s), 0 <= s <= 13	string
decimal(54,s), 0 <= s <= 14	string
decimal(55,s), 0 <= s <= 15	string
decimal(56,s), 0 <= s <= 16	string
decimal(57,s), 0 <= s <= 17	string
decimal(58,s), 0 <= s <= 18	string
decimal(59,s), 0 <= s <= 19	string

MySQL Source Data Type	Google BigQuery Target Data Type
decimal(6)	int64
decimal(6,s), 1 <= s <= 6	numeric
decimal(60,s), 0 <= s <= 20	string
decimal(61,s), 0 <= s <= 21	string
decimal(62,s), 0 <= s <= 22	string
decimal(63,s), 0 <= s <= 23	string
decimal(64,s), 0 <= s <= 24	string
decimal(64,s), 25 <= s <= 29	bignumeric
decimal(65,s), 0 <= s <= 25	string
decimal(7)	int64
decimal(7,s), 1 <= s <= 7	numeric
decimal(8)	int64
decimal(8,s), 1 <= s <= 8	numeric
decimal(9)	int64
decimal(p,s), 10 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 11 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 12 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 13 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 14 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 15 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 16 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 17 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 18 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 19 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 20 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 21 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 22 <= p <= 38, 0 <= s <= 9	numeric



MySQL Source Data Type	Google BigQuery Target Data Type
decimal(p,s), 23 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 24 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 25 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 26 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 27 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 28 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 29 <= p <= 38, 0 <= s <= 9	numeric
decimal(p,s), 29 <= p <= 64, 10 <= s <= 29	bignumeric
decimal(p,s), 30 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 30 <= p <= 64, 1 <= s <= 29	bignumeric
decimal(p,s), 31 <= p <= 38, 2 <= s <= 9	numeric
decimal(p,s), 31 <= p <= 64, 1 <= s <= 29	bignumeric
decimal(p,s), 32 <= p <= 38, 3 <= s <= 9	numeric
decimal(p,s), 32 <= p <= 64, 1 <= s <= 29	bignumeric
decimal(p,s), 33 <= p <= 38, 4 <= s <= 9	numeric
decimal(p,s), 33 <= p <= 64, 1 <= s <= 29	bignumeric
decimal(p,s), 34 <= p <= 38, 5 <= s <= 9	numeric
decimal(p,s), 34 <= p <= 64, 1 <= s <= 29	bignumeric
decimal(p,s), 35 <= p <= 38, 6 <= s <= 9	numeric
decimal(p,s), 35 <= p <= 64, 1 <= s <= 29	bignumeric
decimal(p,s), 36 <= p <= 38, 7 <= s <= 9	numeric
decimal(p,s), 36 <= p <= 64, 1 <= s <= 29	bignumeric
decimal(p,s), 37 <= p <= 38, 8 <= s <= 9	numeric
decimal(p,s), 37 <= p <= 64, 1 <= s <= 29	bignumeric
decimal(p,s), 38 <= p <= 64, 1 <= s <= 29	bignumeric
decimal(p,s), 39 <= p <= 65, 21 <= s <= 29	bignumeric
decimal(p,s), 48 <= p <= 64, 9 <= s <= 29	bignumeric

MySQL Source Data Type	Google BigQuery Target Data Type
decimal(p,s), 49 <= p <= 64, 10 <= s <= 29	bignumeric
decimal(p,s), 50 <= p <= 64, 11 <= s <= 29	bignumeric
decimal(p,s), 51 <= p <= 64, 12 <= s <= 29	bignumeric
decimal(p,s), 52 <= p <= 64, 13 <= s <= 29	bignumeric
decimal(p,s), 53 <= p <= 64, 14 <= s <= 29	bignumeric
decimal(p,s), 54 <= p <= 64, 15 <= s <= 29	bignumeric
decimal(p,s), 55 <= p <= 64, 16 <= s <= 29	bignumeric
decimal(p,s), 56 <= p <= 64, 17 <= s <= 29	bignumeric
decimal(p,s), 57 <= p <= 64, 18 <= s <= 29	bignumeric
decimal(p,s), 58 <= p <= 64, 19 <= s <= 29	bignumeric
decimal(p,s), 59 <= p <= 64, 20 <= s <= 29	bignumeric
decimal(p,s), 60 <= p <= 64, 21 <= s <= 29	bignumeric
decimal(p,s), 61 <= p <= 64, 22 <= s <= 29	bignumeric
decimal(p,s), 62 <= p <= 64, 23 <= s <= 29	bignumeric
decimal(p,s), 63 <= p <= 64, 24 <= s <= 29	bignumeric
decimal(p,s), 9 <= p <= 38, 1 <= s <= 9	numeric
double	float64
float	float64
geomcollection	bytes
geometry	bytes
geometrycollection	bytes
int	int64
int unsigned	int64
json	string
linestring	bytes
longblob	bytes
longtext	string

MySQL Source Data Type	Google BigQuery Target Data Type
mediumblob	bytes
mediumint	int64
mediumint unsigned	int64
mediumtext	string
multilinestring	bytes
multipoint	bytes
multipolygon	bytes
numeric	int64
point	bytes
polygon	bytes
smallint	int64
smallint unsigned	int64
text	string
time( <i>precision</i> ), 0 <= p <= 6	string
timestamp( <i>precision</i> ), 0 <= p <= 6	datetime
tinyblob	bytes
tinyint	int64
tinyint unsigned	int64
tinytext	string
varbinary( <i>size</i> ), 1 <= s <= 65535	bytes
varchar( <i>size</i> ), 1 <= s <= 21844	string
year	int64

## MySQL Source and Microsoft Azure Synapse Analytics Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a MySQL source and a Microsoft Azure Synapse Analytics target:

MySQL Source Data Type	Synapse Analytics Target Data Type
bigint	bigint
bigint unsigned	decimal(20)
binary(size), 1 <= s <= 255	binary(size), 1 <= s <= 255
bit	bit
bit(precision), 1 <= p <= 64	binary(size), 1 <= s <= 8
blob	varbinary(max)
char(size), 1 <= s <= 255	varchar(size), 4 <= s <= 1020
date	date
datetime	datetime2
decimal(p,s), 1 <= p <= 38, 0 <= s <= 29	decimal(p,s), 1 <= p <= 38, 0 <= s <= 29
decimal(p,s), 39 <= p <= 65, 0 <= s <= 29	char(size), 40 <= s <= 67
double	float
float	real
geomcollection	varbinary(max)
geometry	varbinary(max)
geometrycollection	varbinary(max)
int	int
int unsigned	bigint
json	varchar(max)
linestring	varbinary(max)
longblob	varbinary(max)
longtext	varchar(max)
mediumblob	varbinary(max)
mediumint	int
mediumint unsigned	int

MySQL Source Data Type	Synapse Analytics Target Data Type
mediumtext	varchar(max)
multilinestring	varbinary(max)
multipoint	varbinary(max)
multipolygon	varbinary(max)
numeric	decimal(10)
point	varbinary(max)
polygon	varbinary(max)
smallint	smallint
smallint unsigned	int
text	varchar(max)
time( <i>precision</i> ), 0 <= p <= 6	varchar( <i>size</i> ), 10 <= s <= 17
timestamp( <i>precision</i> ), 0 <= p <= 3	datetime2
timestamp( <i>precision</i> ), 4 <= p <= 6	datetime2( <i>precision</i> ), 4 <= p <= 6
tinyblob	binary(255)
tinyint	smallint
tinyint unsigned	smallint
tinytext	char(256)
varbinary( <i>size</i> ), 1 <= s <= 65535	varbinary( <i>size</i> ), 1 <= s <= max
varchar( <i>size</i> ), 1 <= s <= 21844	varchar( <i>size</i> ), 4 <= s <= max
year	smallint

## MySQL Source and Snowflake Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a MySQL source and a Snowflake target:

MySQL Source Data Type	Snowflake Target Data Type
bigint	number(19)
bigint unsigned	number(20)
binary( <i>size</i> ), 1 <= s <= 255	binary( <i>size</i> ), 1 <= s <= 255

MySQL Source Data Type	Snowflake Target Data Type
bit( <i>precision</i> ), 1 <= p <= 64	binary( <i>size</i> ), 1 <= s <= 8
blob	binary(65535)
char( <i>size</i> ), 1 <= s <= 255	varchar( <i>size</i> ), 4 <= s <= 1020
date	date
datetime	datetime(0)
decimal(p,s), 1 <= p <= 38, 0 <= s <= 29	number(p,s), 1 <= p <= 38, 0 <= s <= 29
decimal(p,s), 39 <= p <= 65, 0 <= s <= 29	char( <i>size</i> ), 40 <= s <= 67
double	float
float	float
geomcollection	binary
geometry	binary
geometrycollection	binary
int	number(10)
int unsigned	number(10)
json	varchar
linestring	binary
longblob	binary
longtext	varchar
mediumblob	binary
mediumint	number(7)
mediumint unsigned	number(8)
mediumtext	varchar
multilinestring	binary
multipoint	binary
multipolygon	binary
numeric	integer
point	binary

MySQL Source Data Type	Snowflake Target Data Type
polygon	binary
smallint	number(5)
smallint unsigned	number(5)
text	varchar(65536)
time( <i>precision</i> ), 0 <= p <= 6	varchar( <i>size</i> ), 10 <= s <= 17
timestamp( <i>precision</i> ), 0 <= p <= 6	timestamp_ntz( <i>precision</i> ), 0 <= p <= 6
tinyblob	binary(255)
tinyint	number(3)
tinyint unsigned	number(3)
tinytext	varchar(256)
varbinary( <i>size</i> ), 1 <= s <= 65535	binary( <i>size</i> ), 1 <= s <= 65535
varchar( <i>size</i> ), 1 <= s <= 21844	varchar( <i>size</i> ), 4 <= s <= 87376
year	number(4)

## Netezza Source and Amazon Redshift Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Netezza source and an Amazon Redshift target:

Netezza Source Data Type	Amazon Redshift Target Data Type
bigint	bigint
boolean	boolean
byteint	smallint
char( <i>size</i> ), 1 <= s <= 4096	character( <i>size</i> ), 1 <= s <= 4096
char( <i>size</i> ), 4097 <= s <= 64000	character varying( <i>size</i> ), 4097 <= s <= 64000
date	date
double	double precision
integer	integer
interval	character varying(55)
nchar( <i>size</i> ), 1 <= s <= 16000	character varying( <i>size</i> ), 4 <= s <= 64000

Netezza Source Data Type	Amazon Redshift Target Data Type
numeric(38,38)	character varying(41)
numeric( <i>precision</i> ), 1 <= p <= 38	numeric(p,1), 38 <= p <= 0
numeric(p,s), 1 <= p <= 38, 1 <= s <= 37	numeric(p,s), 1 <= p <= 38, 1 <= s <= 37
nvarchar( <i>size</i> ), 1 <= s <= 16000	character varying( <i>size</i> ), 4 <= s <= 64000
real	real
smallint	smallint
st_geometry( <i>precision</i> ), 1 <= p <= 63001	character varying( <i>size</i> ), 2 <= s <= 65535
time	time without time zone
timestamp	timestamp without time zone
timetz	time with time zone
varbinary( <i>size</i> ), 1 <= s <= 64000	character varying( <i>size</i> ), 2 <= s <= 65535
varchar( <i>size</i> ), 1 <= s <= 64000	character varying( <i>size</i> ), 1 <= s <= 64000

## Netezza Source and Databricks Delta Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Netezza source and a Databricks Delta target:

Netezza Source Data Type	Databricks Delta Target Data Type
bigint	long
boolean	boolean
byteint	integer
char( <i>size</i> ), 1 <= s <= 64000	string
date	string
double	double
integer	integer
interval	string
nchar( <i>size</i> ), 1 <= s <= 16000	string
numeric(p,s), 1 <= p <= 38, 1 <= s <= 38	decimal(p,s), 1 <= p <= 38, 1 <= s <= 38
nvarchar( <i>size</i> ), 1 <= s <= 16000	string



Netezza Source Data Type	Databricks Delta Target Data Type
real	float
smallint	integer
st_geometry( <i>precision</i> ), 1 <= p <= 63001	binary
time	string
timestamp	timestamp
timetz	string
varbinary( <i>size</i> ), 1 <= s <= 64000	binary
varchar( <i>size</i> ), 1 <= s <= 64000	string

## Netezza Source and Google BigQuery Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Netezza source and a Google BigQuery target:

Netezza Source Data Type	Google BigQuery Target Data Type
bigint	int64
boolean	bool
byteint	int64
char( <i>size</i> ), 1 <= s <= 64000	string
date	date
double	float64
integer	int64
interval	string
nchar( <i>size</i> ), 1 <= s <= 16000	string
numeric(10,10)	bignumeric
numeric(11,s), 10 <= s <= 11	bignumeric
numeric(12,s), 10 <= s <= 12	bignumeric
numeric(13,s), 10 <= s <= 13	bignumeric
numeric(14,s), 10 <= s <= 14	bignumeric
numeric(15,s), 10 <= s <= 15	bignumeric

Netezza Source Data Type	Google BigQuery Target Data Type
numeric(16,s), 10 <= s <= 16	bignumeric
numeric(17,s), 10 <= s <= 17	bignumeric
numeric(18,s), 10 <= s <= 18	bignumeric
numeric(19,s), 10 <= s <= 19	bignumeric
numeric(20,s), 10 <= s <= 20	bignumeric
numeric(21,s), 10 <= s <= 21	bignumeric
numeric(22,s), 10 <= s <= 22	bignumeric
numeric(23,s), 10 <= s <= 23	bignumeric
numeric(24,s), 10 <= s <= 24	bignumeric
numeric(25,s), 10 <= s <= 25	bignumeric
numeric(26,s), 10 <= s <= 26	bignumeric
numeric(27,s), 10 <= s <= 27	bignumeric
numeric(28,s), 10 <= s <= 28	bignumeric
numeric(29,s), 10 <= s <= 29	bignumeric
numeric(38,9)	numeric
numeric(38,s), 10 <= s <= 38	bignumeric
numeric(precision), 1 <= p <= 17	int64
numeric(precision), 19 <= p <= 29	numeric
numeric(precision), 30 <= p <= 38	bignumeric
numeric(p,s), 1 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 11 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 12 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 13 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 14 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 15 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 16 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 17 <= p <= 38, 1 <= s <= 9	numeric

Netezza Source Data Type	Google BigQuery Target Data Type
numeric(p,s), 18 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 19 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 20 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 21 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 22 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 23 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 24 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 25 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 26 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 27 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 28 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 29 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 30 <= p <= 31, 1 <= s <= 30	bignumeric
numeric(p,s), 30 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 31 <= p <= 32, 1 <= s <= 31	bignumeric
numeric(p,s), 31 <= p <= 38, 2 <= s <= 9	numeric
numeric(p,s), 32 <= p <= 33, 1 <= s <= 32	bignumeric
numeric(p,s), 32 <= p <= 38, 3 <= s <= 9	numeric
numeric(p,s), 33 <= p <= 34, 1 <= s <= 33	bignumeric
numeric(p,s), 33 <= p <= 38, 4 <= s <= 9	numeric
numeric(p,s), 34 <= p <= 35, 1 <= s <= 34	bignumeric
numeric(p,s), 34 <= p <= 38, 5 <= s <= 9	numeric
numeric(p,s), 35 <= p <= 36, 1 <= s <= 35	bignumeric
numeric(p,s), 35 <= p <= 38, 6 <= s <= 9	numeric
numeric(p,s), 36 <= p <= 37, 1 <= s <= 36	bignumeric
numeric(p,s), 36 <= p <= 38, 7 <= s <= 9	numeric
numeric(p,s), 37 <= p <= 38, 1 <= s <= 37	bignumeric

Netezza Source Data Type	Google BigQuery Target Data Type
numeric(p,s), 37 <= p <= 38, 8 <= s <= 9	numeric
nvarchar(size), 1 <= s <= 16000	string
real	float64
smallint	int64
st_geometry(precision), 1 <= p <= 63001	bytes
time	time
timestamp	datetime
timetz	timestamp
varbinary(size), 1 <= s <= 64000	bytes
varchar(size), 1 <= s <= 64000	string

## Netezza Source and Microsoft Azure Synapse Analytics Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Netezza source and a Microsoft Azure Synapse Analytics target:

Netezza Source Data Type	Synapse Analytics Target Data Type
bigint	bigint
boolean	bit
byteint	smallint
char(size), 1 <= s <= 8000	char(size), 1 <= s <= 8000
char(size), 9000 <= s <= 64000	nvarchar(max)
date	date
double	float
integer	int
interval	varchar(55)
nchar(size), 1 <= s <= 3100	nchar(size), 1 <= s <= 3100
nchar(size), 4100 <= s <= 16000	nvarchar(max)
numeric(p,s), 1 <= p <= 38, 1 <= s <= 38	decimal(p,s), 1 <= p <= 38, 1 <= s <= 38
nvarchar(size), 1 <= s <= 16000	nvarchar(size), 1 <= s <= max

Netezza Source Data Type	Synapse Analytics Target Data Type
real	real
smallint	smallint
st_geometry( <i>precision</i> ), 1 <= p <= 63001	varbinary( <i>size</i> ), 1 <= s <= max
time	time(6)
timestamp	datetime2(6)
timetz	datetimeoffset(6)
varbinary( <i>size</i> ), 1 <= s <= 64000	varbinary( <i>size</i> ), 1 <= s <= max
varchar( <i>size</i> ), 1 <= s <= 64000	varchar( <i>size</i> ), 1 <= s <= max

## Netezza Source and Snowflake Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Netezza source and a Snowflake target:

Netezza Source Data Type	Snowflake Target Data Type
bigint	integer
boolean	boolean
byteint	integer
char( <i>size</i> ), 1 <= s <= 64000	char( <i>size</i> ), 1 <= s <= 64000
date	date
double	float
integer	integer
interval	varchar(55)
nchar( <i>size</i> ), 1 <= s <= 16000	char( <i>size</i> ), 4 <= s <= 64000
numeric(38,38)	char(41)
numeric( <i>precision</i> ), 1 <= p <= 38	integer
numeric(p,s), 1 <= p <= 38, 1 <= s <= 37	number(p,s), 1 <= p <= 38, 1 <= s <= 37
nvarchar( <i>size</i> ), 1 <= s <= 16000	varchar( <i>size</i> ), 4 <= s <= 64000
real	float
smallint	integer

Netezza Source Data Type	Snowflake Target Data Type
st_geometry( <i>precision</i> ), 1 <= p <= 63001	binary( <i>size</i> ), 1 <= s <= 63001
time	time(6)
timestamp	timestamp_ntz(6)
timetz	timestamp_tz(6)
varbinary( <i>size</i> ), 1 <= s <= 64000	binary( <i>size</i> ), 1 <= s <= 64000
varchar( <i>size</i> ), 1 <= s <= 64000	varchar( <i>size</i> ), 1 <= s <= 64000

## Oracle Source and Amazon Redshift Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with an Oracle source and an Amazon Redshift target:

Oracle Source Data Type	Amazon Redshift Target Data Type
binary_double	double precision
binary_float	real
char(s byte), 1 <= s <= 2000	character varying( <i>size</i> ), 4 <= s <= 2000
char(s char), 1 <= s <= 2000	character varying( <i>size</i> ), 4 <= s <= 8000
date	timestamp without time zone
float( <i>precision</i> ), 1 <= p <= 126	character varying(255)
integer	character varying(255)
long raw	character varying(65535)
long(2147483648 byte)	character varying(65535)
nchar(s char), 1 <= s <= 2000	character varying( <i>size</i> ), 4 <= s <= 8000
number	character varying(255)
number(*,s), -84 <= s <= 127	character varying(255)
number(38,s), 0 <= s <= 37	numeric(38,s), 0 <= s <= 37
number(p,s), 1 <= p <= 38, -37 <= s <= 37	numeric(p,s), 1 <= p <= 38, 0 <= s <= 37
number(p,s), 1 <= p <= 38, -84 <= s <= 127	character varying( <i>size</i> ), 40 <= s <= 130
number(p,s), 10 <= p <= 38, -28 <= s <= 37	numeric(p,s), 10 <= p <= 38, 0 <= s <= 37
number(p,s), 11 <= p <= 38, -27 <= s <= 37	numeric(p,s), 11 <= p <= 38, 0 <= s <= 37

Oracle Source Data Type	Amazon Redshift Target Data Type
number(p,s), 12 <= p <= 38, -26 <= s <= 37	numeric(p,s), 12 <= p <= 38, 0 <= s <= 37
number(p,s), 13 <= p <= 38, -25 <= s <= 37	numeric(p,s), 13 <= p <= 38, 0 <= s <= 37
number(p,s), 14 <= p <= 38, -24 <= s <= 37	numeric(p,s), 14 <= p <= 38, 0 <= s <= 37
number(p,s), 15 <= p <= 38, -23 <= s <= 37	numeric(p,s), 15 <= p <= 38, 0 <= s <= 37
number(p,s), 16 <= p <= 38, -22 <= s <= 37	numeric(p,s), 16 <= p <= 38, 0 <= s <= 37
number(p,s), 17 <= p <= 38, -21 <= s <= 37	numeric(p,s), 17 <= p <= 38, 0 <= s <= 37
number(p,s), 18 <= p <= 38, -20 <= s <= 37	numeric(p,s), 18 <= p <= 38, 0 <= s <= 37
number(p,s), 19 <= p <= 38, -19 <= s <= 37	numeric(p,s), 20 <= p <= 38, 0 <= s <= 37
number(p,s), 2 <= p <= 38, -36 <= s <= 37	numeric(p,s), 2 <= p <= 38, 0 <= s <= 37
number(p,s), 21 <= p <= 38, -17 <= s <= 37	numeric(p,s), 21 <= p <= 38, 0 <= s <= 37
number(p,s), 22 <= p <= 38, -16 <= s <= 37	numeric(p,s), 22 <= p <= 38, 0 <= s <= 37
number(p,s), 23 <= p <= 38, -15 <= s <= 37	numeric(p,s), 23 <= p <= 38, 0 <= s <= 37
number(p,s), 24 <= p <= 38, -14 <= s <= 37	numeric(p,s), 24 <= p <= 38, 0 <= s <= 37
number(p,s), 25 <= p <= 38, -13 <= s <= 37	numeric(p,s), 25 <= p <= 38, 0 <= s <= 37
number(p,s), 26 <= p <= 38, -12 <= s <= 37	numeric(p,s), 26 <= p <= 38, 0 <= s <= 37
number(p,s), 27 <= p <= 38, -11 <= s <= 37	numeric(p,s), 27 <= p <= 38, 0 <= s <= 37
number(p,s), 28 <= p <= 38, -10 <= s <= 37	numeric(p,s), 28 <= p <= 38, 0 <= s <= 37
number(p,s), 29 <= p <= 38, -9 <= s <= 37	numeric(p,s), 29 <= p <= 38, 0 <= s <= 37
number(p,s), 3 <= p <= 38, -35 <= s <= 37	numeric(p,s), 3 <= p <= 38, 0 <= s <= 37
number(p,s), 30 <= p <= 38, -8 <= s <= 37	numeric(p,s), 30 <= p <= 38, 0 <= s <= 37
number(p,s), 31 <= p <= 38, -7 <= s <= 37	numeric(p,s), 31 <= p <= 38, 0 <= s <= 37
number(p,s), 32 <= p <= 38, -6 <= s <= 37	numeric(p,s), 32 <= p <= 38, 0 <= s <= 37
number(p,s), 33 <= p <= 38, -5 <= s <= 37	numeric(p,s), 33 <= p <= 38, 0 <= s <= 37
number(p,s), 34 <= p <= 38, -4 <= s <= 37	numeric(p,s), 34 <= p <= 38, 0 <= s <= 37
number(p,s), 35 <= p <= 38, -3 <= s <= 37	numeric(p,s), 35 <= p <= 38, 0 <= s <= 37
number(p,s), 36 <= p <= 38, -2 <= s <= 37	numeric(p,s), 36 <= p <= 38, 0 <= s <= 37
number(p,s), 37 <= p <= 38, -1 <= s <= 37	numeric(p,s), 37 <= p <= 38, 0 <= s <= 37

Oracle Source Data Type	Amazon Redshift Target Data Type
number(p,s), 4 <= p <= 38, -34 <= s <= 37	numeric(p,s), 4 <= p <= 38, 0 <= s <= 37
number(p,s), 5 <= p <= 38, -33 <= s <= 37	numeric(p,s), 5 <= p <= 38, 0 <= s <= 37
number(p,s), 6 <= p <= 38, -32 <= s <= 37	numeric(p,s), 6 <= p <= 38, 0 <= s <= 37
number(p,s), 7 <= p <= 38, -31 <= s <= 37	numeric(p,s), 7 <= p <= 38, 0 <= s <= 37
number(p,s), 8 <= p <= 38, -30 <= s <= 37	numeric(p,s), 8 <= p <= 38, 0 <= s <= 37
number(p,s), 9 <= p <= 38, -29 <= s <= 37	numeric(p,s), 9 <= p <= 38, 0 <= s <= 37
nvarchar2(s char), 1 <= s <= 4000	character varying(size), 4 <= s <= 16000
raw(size), 1 <= s <= 4000	character varying(size), 2 <= s <= 8000
rowid	character varying(18)
timestamp(precision) with time zone, 0 <= p <= 6	timestamp with time zone
timestamp(precision) with time zone, 7 <= p <= 9	character varying(size), 67 <= s <= 69
timestamp(precision), 1 <= p <= 6	timestamp without time zone
timestamp(precision), 7 <= p <= 9	character varying(size), 27 <= s <= 29
varchar2(s byte), 1 <= s <= 4000	character varying(size), 4 <= s <= 4000
varchar2(s char), 1 <= s <= 4000	character varying(size), 4 <= s <= 16000

## Oracle Source and Databricks Delta Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Oracle source and a Databricks Delta target:

Oracle Source Data Type	Databricks Delta Target Data Type
binary_double	double
binary_float	float
char(s byte), 1 <= s <= 2000	string
char(s char), 1 <= s <= 2000	string
date	timestamp
float(precision), 1 <= p <= 126	string
integer	string



Oracle Source Data Type	Databricks Delta Target Data Type
long raw	binary
long(2147483648 byte)	string
nchar(s char), 1 <= s <= 2000	string
number	string
number(*,s), -84 <= s <= 127	string
number(p,s), 1 <= p <= 38, -37 <= s <= 38	decimal(p,s), 1 <= p <= 38, 0 <= s <= 38
number(p,s), 1 <= p <= 38, -84 <= s <= 127	string
nvarchar2(s char), 1 <= s <= 4000	string
raw(size), 1 <= s <= 4000	binary
rowid	string
timestamp(precision) with time zone, 0 <= p <= 6	timestamp
timestamp(precision) with time zone, 7 <= p <= 9	string
timestamp(precision), 1 <= p <= 6	timestamp
timestamp(precision), 7 <= p <= 9	string
varchar2(s byte), 1 <= s <= 4000	string
varchar2(s char), 1 <= s <= 4000	string

## Oracle Source and Google BigQuery Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Oracle source and a Google BigQuery target:

Oracle Source Data Type	Google BigQuery Target Data Type
binary_double	float64
binary_float	float64
char(s byte), 1 <= s <= 2000	string
char(s char), 1 <= s <= 2000	string
date	datetime
float(precision), 1 <= p <= 126	string

Oracle Source Data Type	Google BigQuery Target Data Type
integer	string
long raw	bytes
long(2147483648 byte)	string
nchar(s char), 1 <= s <= 2000	string
number	string
number(*,s), -84 <= s <= 127	string
number(1,-38)	bignumeric
number(1,s), -37 <= s <= -29	bignumeric
number(1,s), -84 <= s <= -39	string
number(10,-29)	bignumeric
number(11,-28)	bignumeric
number(12,-27)	bignumeric
number(13,-26)	bignumeric
number(14,-25)	bignumeric
number(15,-24)	bignumeric
number(16,-23)	bignumeric
number(17,-22)	bignumeric
number(18,-21)	bignumeric
number(19,-20)	bignumeric
number(2,-37)	bignumeric
number(20,-19)	bignumeric
number(21,-18)	bignumeric
number(22,-17)	bignumeric
number(23,-16)	bignumeric
number(24,-15)	bignumeric
number(25,-14)	bignumeric
number(26,-13)	bignumeric

Oracle Source Data Type	Google BigQuery Target Data Type
number(27,-12)	bignumeric
number(28,-11)	bignumeric
number(29,-10)	bignumeric
number(3,-36)	bignumeric
number(30,-9)	bignumeric
number(31,-8)	bignumeric
number(32,-7)	bignumeric
number(33,-6)	bignumeric
number(34,-5)	bignumeric
number(35,-4)	bignumeric
number(36,-3)	bignumeric
number(37,-2)	bignumeric
number(38,-1)	bignumeric
number(38,9)	numeric
number(38,s), 10 <= s <= 38	bignumeric
number(4,-35)	bignumeric
number(5,-34)	bignumeric
number(6,-33)	bignumeric
number(7,-32)	bignumeric
number(8,-31)	bignumeric
number(9,-30)	bignumeric
number(p,s), 1 <= p <= 38, -28 <= s <= 9	numeric
number(p,s), 1 <= p <= 38, -36 <= s <= 38	bignumeric
number(p,s), 1 <= p <= 38, -84 <= s <= 127	string
number(p,s), 10 <= p <= 38, -19 <= s <= 9	numeric
number(p,s), 10 <= p <= 38, -27 <= s <= 38	bignumeric
number(p,s), 10 <= p <= 38, -84 <= s <= 127	string

Oracle Source Data Type	Google BigQuery Target Data Type
number(p,s), 11 <= p <= 38, -18 <= s <= 9	numeric
number(p,s), 11 <= p <= 38, -26 <= s <= 38	bignumeric
number(p,s), 11 <= p <= 38, -84 <= s <= 127	string
number(p,s), 12 <= p <= 38, -17 <= s <= 9	numeric
number(p,s), 12 <= p <= 38, -25 <= s <= 38	bignumeric
number(p,s), 12 <= p <= 38, -84 <= s <= 127	string
number(p,s), 13 <= p <= 38, -16 <= s <= 9	numeric
number(p,s), 13 <= p <= 38, -24 <= s <= 38	bignumeric
number(p,s), 13 <= p <= 38, -84 <= s <= 127	string
number(p,s), 14 <= p <= 38, -15 <= s <= 9	numeric
number(p,s), 14 <= p <= 38, -23 <= s <= 38	bignumeric
number(p,s), 14 <= p <= 38, -84 <= s <= 127	string
number(p,s), 15 <= p <= 38, -14 <= s <= 9	numeric
number(p,s), 15 <= p <= 38, -22 <= s <= 38	bignumeric
number(p,s), 15 <= p <= 38, -84 <= s <= 127	string
number(p,s), 16 <= p <= 38, -13 <= s <= 9	numeric
number(p,s), 16 <= p <= 38, -21 <= s <= 38	bignumeric
number(p,s), 16 <= p <= 38, -84 <= s <= 127	string
number(p,s), 17 <= p <= 38, -12 <= s <= 9	numeric
number(p,s), 17 <= p <= 38, -20 <= s <= 38	bignumeric
number(p,s), 17 <= p <= 38, -84 <= s <= 127	string
number(p,s), 18 <= p <= 38, -11 <= s <= 9	numeric
number(p,s), 18 <= p <= 38, -19 <= s <= 38	bignumeric
number(p,s), 18 <= p <= 38, -84 <= s <= 127	string
number(p,s), 19 <= p <= 38, -10 <= s <= 9	numeric
number(p,s), 19 <= p <= 38, -18 <= s <= 38	bignumeric
number(p,s), 19 <= p <= 38, -84 <= s <= 127	string

Oracle Source Data Type	Google BigQuery Target Data Type
number(p,s), 2 <= p <= 38, -27 <= s <= 9	numeric
number(p,s), 2 <= p <= 38, -35 <= s <= 38	bignumeric
number(p,s), 2 <= p <= 38, -84 <= s <= 127	string
number(p,s), 20 <= p <= 38, -17 <= s <= 38	bignumeric
number(p,s), 20 <= p <= 38, -84 <= s <= 127	string
number(p,s), 20 <= p <= 38, -9 <= s <= 9	numeric
number(p,s), 21 <= p <= 38, -16 <= s <= 38	bignumeric
number(p,s), 21 <= p <= 38, -8 <= s <= 9	numeric
number(p,s), 21 <= p <= 38, -84 <= s <= 127	string
number(p,s), 22 <= p <= 38, -15 <= s <= 38	bignumeric
number(p,s), 22 <= p <= 38, -7 <= s <= 9	numeric
number(p,s), 22 <= p <= 38, -84 <= s <= 127	string
number(p,s), 23 <= p <= 38, -14 <= s <= 38	bignumeric
number(p,s), 23 <= p <= 38, -6 <= s <= 9	numeric
number(p,s), 23 <= p <= 38, -84 <= s <= 127	string
number(p,s), 24 <= p <= 38, -13 <= s <= 38	bignumeric
number(p,s), 24 <= p <= 38, -5 <= s <= 9	numeric
number(p,s), 24 <= p <= 38, -84 <= s <= 127	string
number(p,s), 25 <= p <= 38, -12 <= s <= 38	bignumeric
number(p,s), 25 <= p <= 38, -4 <= s <= 9	numeric
number(p,s), 25 <= p <= 38, -84 <= s <= 127	string
number(p,s), 26 <= p <= 38, -11 <= s <= 38	bignumeric
number(p,s), 26 <= p <= 38, -3 <= s <= 9	numeric
number(p,s), 26 <= p <= 38, -84 <= s <= 127	string
number(p,s), 27 <= p <= 38, -10 <= s <= 38	bignumeric
number(p,s), 27 <= p <= 38, -2 <= s <= 9	numeric
number(p,s), 27 <= p <= 38, -84 <= s <= 127	string

Oracle Source Data Type	Google BigQuery Target Data Type
number(p,s), 28 <= p <= 38, -1 <= s <= 9	numeric
number(p,s), 28 <= p <= 38, -84 <= s <= 127	string
number(p,s), 28 <= p <= 38, -9 <= s <= 38	bignumeric
number(p,s), 29 <= p <= 38, -8 <= s <= 38	bignumeric
number(p,s), 29 <= p <= 38, -84 <= s <= 127	string
number(p,s), 29 <= p <= 38, 0 <= s <= 9	numeric
number(p,s), 3 <= p <= 38, -26 <= s <= 9	numeric
number(p,s), 3 <= p <= 38, -34 <= s <= 38	bignumeric
number(p,s), 3 <= p <= 38, -84 <= s <= 127	string
number(p,s), 30 <= p <= 38, -7 <= s <= 38	bignumeric
number(p,s), 30 <= p <= 38, -84 <= s <= 127	string
number(p,s), 30 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 31 <= p <= 38, -6 <= s <= 38	bignumeric
number(p,s), 31 <= p <= 38, -84 <= s <= 127	string
number(p,s), 31 <= p <= 38, 2 <= s <= 9	numeric
number(p,s), 32 <= p <= 38, -5 <= s <= 38	bignumeric
number(p,s), 32 <= p <= 38, -84 <= s <= 127	string
number(p,s), 32 <= p <= 38, 3 <= s <= 9	numeric
number(p,s), 33 <= p <= 38, -4 <= s <= 38	bignumeric
number(p,s), 33 <= p <= 38, -84 <= s <= 127	string
number(p,s), 33 <= p <= 38, 4 <= s <= 9	numeric
number(p,s), 34 <= p <= 38, -3 <= s <= 38	bignumeric
number(p,s), 34 <= p <= 38, -84 <= s <= 127	string
number(p,s), 34 <= p <= 38, 5 <= s <= 9	numeric
number(p,s), 35 <= p <= 38, -2 <= s <= 38	bignumeric
number(p,s), 35 <= p <= 38, -84 <= s <= 127	string
number(p,s), 35 <= p <= 38, 6 <= s <= 9	numeric

Oracle Source Data Type	Google BigQuery Target Data Type
number(p,s), 36 <= p <= 37, -1 <= s <= 38	bignumeric
number(p,s), 36 <= p <= 38, -84 <= s <= 127	string
number(p,s), 36 <= p <= 38, 7 <= s <= 9	numeric
number(p,s), 37 <= p <= 38, -84 <= s <= 127	string
number(p,s), 37 <= p <= 38, 1 <= s <= 38	bignumeric
number(p,s), 37 <= p <= 38, 8 <= s <= 9	numeric
number(p,s), 4 <= p <= 38, -25 <= s <= 9	numeric
number(p,s), 4 <= p <= 38, -33 <= s <= 38	bignumeric
number(p,s), 4 <= p <= 38, -84 <= s <= 127	string
number(p,s), 5 <= p <= 38, -24 <= s <= 9	numeric
number(p,s), 5 <= p <= 38, -32 <= s <= 38	bignumeric
number(p,s), 5 <= p <= 38, -84 <= s <= 127	string
number(p,s), 6 <= p <= 38, -23 <= s <= 9	numeric
number(p,s), 6 <= p <= 38, -31 <= s <= 38	bignumeric
number(p,s), 6 <= p <= 38, -84 <= s <= 127	string
number(p,s), 7 <= p <= 38, -22 <= s <= 9	numeric
number(p,s), 7 <= p <= 38, -30 <= s <= 38	bignumeric
number(p,s), 7 <= p <= 38, -84 <= s <= 127	string
number(p,s), 8 <= p <= 38, -21 <= s <= 9	numeric
number(p,s), 8 <= p <= 38, -29 <= s <= 38	bignumeric
number(p,s), 8 <= p <= 38, -84 <= s <= 127	string
number(p,s), 9 <= p <= 38, -20 <= s <= 9	numeric
number(p,s), 9 <= p <= 38, -28 <= s <= 38	bignumeric
number(p,s), 9 <= p <= 38, -84 <= s <= 127	string
nvarchar2(s char), 1 <= s <= 4000	string
raw(size), 1 <= s <= 4000	bytes
rowid	string

Oracle Source Data Type	Google BigQuery Target Data Type
timestamp( <i>precision</i> ) with time zone, 0 <= p <= 6	timestamp
timestamp( <i>precision</i> ) with time zone, 7 <= p <= 9	string
timestamp( <i>precision</i> ), 1 <= p <= 6	datetime
timestamp( <i>precision</i> ), 7 <= p <= 9	string
varchar2(s byte), 1 <= s <= 4000	string
varchar2(s char), 1 <= s <= 4000	string

## Oracle Source and Microsoft Azure Synapse Analytics Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with an Oracle source and a Microsoft Azure Synapse Analytics target:

Oracle Source Data Type	Synapse Analytics Target Data Type
binary_double	float
binary_float	real
char(s byte), 1 <= s <= 2000	varchar( <i>size</i> ), 4 <= s <= 2000
char(s char), 1 <= s <= 2000	varchar( <i>size</i> ), 4 <= s <= 8000
date	datetime2(0)
float( <i>precision</i> ), 1 <= p <= 126	varchar(255)
integer	varchar(255)
long raw	varbinary(max)
long(2147483648 byte)	varchar(max)
nchar(s char), 1 <= s <= 2000	nchar( <i>size</i> ), 1 <= s <= 2000
number	varchar(255)
number(*,s), -84 <= s <= 127	varchar(255)
number(p,s), 1 <= p <= 38, -37 <= s <= 38	decimal(p,s), 1 <= p <= 38, 0 <= s <= 38
number(p,s), 1 <= p <= 38, -84 <= s <= 127	varchar( <i>size</i> ), 40 <= s <= 130
nvarchar2(s char), 1 <= s <= 4000	nvarchar( <i>size</i> ), 1 <= s <= 4000
raw( <i>size</i> ), 1 <= s <= 4000	varbinary( <i>size</i> ), 1 <= s <= 4000



Oracle Source Data Type	Synapse Analytics Target Data Type
rowid	varchar(18)
timestamp( <i>precision</i> ) with time zone, 0 <= p <= 7	datetimeoffset( <i>precision</i> ), 0 <= p <= 7
timestamp( <i>precision</i> ) with time zone, 8 <= p <= 9	char( <i>size</i> ), 68 <= s <= 69
timestamp( <i>precision</i> ), 1 <= p <= 7	datetime2( <i>precision</i> ), 1 <= p <= 7
timestamp( <i>precision</i> ), 8 <= p <= 9	char( <i>size</i> ), 28 <= s <= 29
varchar2(s byte), 1 <= s <= 4000	varchar( <i>size</i> ), 4 <= s <= 4000
varchar2(s char), 1 <= s <= 4000	varchar( <i>size</i> ), 4 <= s <= max

## Oracle Source and Oracle Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with an Oracle source and an Oracle target:

Oracle Source Data Type	Oracle Target Data Type
binary_double	binary_double
binary_float	binary_float
blob	blob
char(s byte), 1 <= s <= 2000	char(s byte), 1 <= s <= 2000
char(s char), 1 <= s <= 2000	char(s char), 1 <= s <= 2000
clob	clob
date	date
float( <i>precision</i> ), 1 <= p <= 126	float( <i>precision</i> ), 1 <= p <= 126
integer	integer
long raw	long raw
long(2147483648 byte)	long(2147483648 byte)
nchar(s char), 1 <= s <= 2000	nchar(s char), 1 <= s <= 2000
nclob	nclob
number	number
number(*,s), -84 <= s <= 127	number(*,s), -84 <= s <= 127

Oracle Source Data Type	Oracle Target Data Type
number(p,s), 1 <= p <= 38, -84 <= s <= 127	number(p,s), 1 <= p <= 38, -84 <= s <= 127
nvarchar2(s char), 1 <= s <= 4000	nvarchar2(s char), 1 <= s <= 4000
raw(size), 1 <= s <= 4000	raw(size), 1 <= s <= 4000
rowid	rowid
timestamp(precision) with time zone, 0 <= p <= 9	timestamp(precision) with time zone, 0 <= p <= 9
timestamp(precision), 1 <= p <= 9	timestamp(precision), 1 <= p <= 9
varchar2(s byte), 1 <= s <= 4000	varchar2(s byte), 1 <= s <= 4000
varchar2(s char), 1 <= s <= 4000	varchar2(s char), 1 <= s <= 4000

## Oracle Source and Snowflake Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with an Oracle source and a Snowflake target:

Oracle Source Data Type	Snowflake Target Data Type
binary_double	float
binary_float	float
char(s byte), 1 <= s <= 2000	char(size), 4 <= s <= 2000
char(s char), 1 <= s <= 2000	char(size), 4 <= s <= 8000
date	timestamp_ntz(0)
float(precision), 1 <= p <= 126	char(255)
integer	char(255)
long raw	binary
long(2147483648 byte)	char(16777216)
nchar(s char), 1 <= s <= 2000	char(size), 4 <= s <= 8000
number	char(255)
number(*,s), -84 <= s <= 127	char(255)
number(p,s), 1 <= p <= 38, -37 <= s <= 37	number(p,s), 1 <= p <= 38, 0 <= s <= 37
number(p,s), 1 <= p <= 38, -84 <= s <= 127	char(size), 40 <= s <= 130

Oracle Source Data Type	Snowflake Target Data Type
nvarchar2(s char), 1 <= s <= 4000	char(size), 4 <= s <= 16000
raw(size), 1 <= s <= 4000	binary(size), 1 <= s <= 4000
rowid	varchar(18)
timestamp(precision) with time zone, 0 <= p <= 9	timestamp_tz(precision), 0 <= p <= 9
timestamp(precision), 1 <= p <= 9	timestamp_ntz(precision), 1 <= p <= 9
varchar2(s byte), 1 <= s <= 4000	varchar(size), 4 <= s <= 4000
varchar2(s char), 1 <= s <= 4000	varchar(size), 4 <= s <= 16000

## PostgreSQL Source and Amazon Redshift Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a PostgreSQL source and an Amazon Redshift target:

PostgreSQL Source Data Type	Amazon Redshift Target Data Type
bigint	bigint
bit varying(precision), 1 <= p <= 83700001	character varying(size), 2 <= s <= 65535
bit(precision), 1 <= p <= 83886080	character varying(size), 2 <= s <= 65535
boolean	boolean
box	geometry
character varying(size), 1 <= s <= 10485760	character varying(size), 4 <= s <= 65535
character(size), 1 <= s <= 10485760	character varying(size), 4 <= s <= 65535
cidr	character varying(45)
circle	geometry
date	date
daterange	character varying(29)
double precision	double precision
inet	character varying(45)
int4range	character varying(25)
int8range	character varying(43)

PostgreSQL Source Data Type	Amazon Redshift Target Data Type
integer	integer
line	geometry
lseg	geometry
macaddr	character varying(17)
macaddr8	character varying(23)
money	numeric(20,2)
numeric(p,s), 1 <= p <= 38, 0 <= s <= 37	numeric(p,s), 1 <= p <= 38, 0 <= s <= 37
numeric(p,s), 38 <= p <= 999, 38 <= s <= 998	character varying(size), 41 <= s <= 1001
numrange	character varying(65535)
path	geometry
point	geometry
polygon	geometry
real	real
smallint	smallint
time(precision) with time zone, 0 <= p <= 6	time with time zone
time(precision) without time zone, 0 <= p <= 6	time without time zone
timestamp(precision) with time zone, 0 <= p <= 6	timestamp with time zone
timestamp(precision) without time zone, 0 <= p <= 6	timestamp without time zone
tsrange	character varying(63)
tstzrange	character varying(75)
uuid	character(36)

For incremental load jobs, Mass Ingestion Databases does not support the following PostgreSQL data types, in addition to those not supported for initial load jobs:

- BYTEA
- MONEY
- Spatial types
  - Box
  - Circle

- Line
- LSeg
- Path
- Point
- Polygon
- TEXT
- Unbounded varying types
- XML

## PostgreSQL Source and Databricks Delta Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a PostgreSQL source and a Databricks Delta target:

PostgreSQL Source Data Type	Databricks Delta Target Data Type
bigint	long
bit varying( <i>precision</i> ), 1 <= p <= 83700001	binary
bit( <i>precision</i> ), 1 <= p <= 83886080	binary
boolean	boolean
box	binary
character varying( <i>size</i> ), 1 <= s <= 10485760	string
character( <i>size</i> ), 1 <= s <= 10485760	string
cidr	string
circle	binary
date	string
daterange	string
double precision	double
inet	string
int4range	string
int8range	string
integer	integer
line	binary
lseg	binary

PostgreSQL Source Data Type	Databricks Delta Target Data Type
macaddr	string
macaddr8	string
money	decimal(19,2)
numeric(p,s), 1 <= p <= 38, 0 <= s <= 38	decimal(p,s), 1 <= p <= 38, 0 <= s <= 38
numeric(p,s), 39 <= p <= 999, 39 <= s <= 998	string
numrange	string
path	binary
point	binary
polygon	binary
real	float
smallint	integer
time(precision) with time zone, 0 <= p <= 6	string
time(precision) without time zone, 0 <= p <= 6	string
timestamp(precision) with time zone, 0 <= p <= 6	timestamp
timestamp(precision) without time zone, 0 <= p <= 6	timestamp
tsrange	string
tstzrange	string
uuid	string

## PostgreSQL Source and Google BigQuery Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a PostgreSQL source and a Google BigQuery target:

PostgreSQL Source Data Type	Google BigQuery Target Data Type
bigint	int64
bit varying(precision), 1 <= p <= 83700001	bytes
bit(precision), 1 <= p <= 83886080	bytes
boolean	bool

PostgreSQL Source Data Type	Google BigQuery Target Data Type
box	bytes
character varying(size), 1 <= s <= 10485760	string
character(size), 1 <= s <= 10485760	string
cidr	string
circle	bytes
date	date
daterange	string
double precision	float64
inet	string
int4range	string
int8range	string
integer	int64
line	bytes
lseg	bytes
macaddr	string
macaddr8	string
money	numeric
numeric(1)	int64
numeric(1,1)	numeric
numeric(10)	int64
numeric(10,10)	bignumeric
numeric(11)	int64
numeric(11,s), 10 <= s <= 11	bignumeric
numeric(12)	int64
numeric(12,s), 10 <= s <= 12	bignumeric
numeric(13)	int64
numeric(13,s), 10 <= s <= 13	bignumeric

PostgreSQL Source Data Type	Google BigQuery Target Data Type
numeric(14)	int64
numeric(14,s), 10 <= s <= 14	bignumeric
numeric(15)	int64
numeric(15,s), 10 <= s <= 15	bignumeric
numeric(16)	int64
numeric(16,s), 10 <= s <= 16	bignumeric
numeric(17)	int64
numeric(17,s), 10 <= s <= 17	bignumeric
numeric(18)	int64
numeric(18,s), 10 <= s <= 18	bignumeric
numeric(19,s), 10 <= s <= 19	bignumeric
numeric(2)	int64
numeric(2,s), 1 <= s <= 2	numeric
numeric(20,s), 10 <= s <= 20	bignumeric
numeric(21,s), 10 <= s <= 21	bignumeric
numeric(22,s), 10 <= s <= 22	bignumeric
numeric(23,s), 10 <= s <= 23	bignumeric
numeric(24,s), 10 <= s <= 24	bignumeric
numeric(25,s), 10 <= s <= 25	bignumeric
numeric(26,s), 10 <= s <= 26	bignumeric
numeric(27,s), 10 <= s <= 27	bignumeric
numeric(28,s), 10 <= s <= 28	bignumeric
numeric(3)	int64
numeric(3,s), 1 <= s <= 3	numeric
numeric(38,9)	numeric
numeric(38,s), 10 <= s <= 38	bignumeric
numeric(4)	int64



PostgreSQL Source Data Type	Google BigQuery Target Data Type
numeric(4,s), 1 <= s <= 4	numeric
numeric(5)	int64
numeric(5,s), 1 <= s <= 5	numeric
numeric(6)	int64
numeric(6,s), 1 <= s <= 6	numeric
numeric(7)	int64
numeric(7,s), 1 <= s <= 7	numeric
numeric(8)	int64
numeric(8,s), 1 <= s <= 8	numeric
numeric(9)	int64
numeric(p,s), 10 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 11 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 12 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 13 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 14 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 15 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 16 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 17 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 18 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 19 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 20 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 21 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 22 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 23 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 24 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 25 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 26 <= p <= 38, 0 <= s <= 9	numeric

PostgreSQL Source Data Type	Google BigQuery Target Data Type
numeric(p,s), 27 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 28 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 29 <= p <= 30, 10 <= s <= 29	bignumeric
numeric(p,s), 29 <= p <= 38, 0 <= s <= 9	numeric
numeric(p,s), 30 <= p <= 31, 1 <= s <= 30	bignumeric
numeric(p,s), 30 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 31 <= p <= 32, 1 <= s <= 31	bignumeric
numeric(p,s), 31 <= p <= 38, 2 <= s <= 9	numeric
numeric(p,s), 32 <= p <= 33, 1 <= s <= 32	bignumeric
numeric(p,s), 32 <= p <= 38, 3 <= s <= 9	numeric
numeric(p,s), 33 <= p <= 34, 1 <= s <= 33	bignumeric
numeric(p,s), 33 <= p <= 38, 4 <= s <= 9	numeric
numeric(p,s), 34 <= p <= 35, 1 <= s <= 34	bignumeric
numeric(p,s), 34 <= p <= 38, 5 <= s <= 9	numeric
numeric(p,s), 35 <= p <= 36, 1 <= s <= 35	bignumeric
numeric(p,s), 35 <= p <= 38, 6 <= s <= 9	numeric
numeric(p,s), 36 <= p <= 37, 1 <= s <= 36	bignumeric
numeric(p,s), 36 <= p <= 38, 7 <= s <= 9	numeric
numeric(p,s), 37 <= p <= 38, 1 <= s <= 37	bignumeric
numeric(p,s), 37 <= p <= 38, 8 <= s <= 9	numeric
numeric(p,s), 39 <= p <= 999, 39 <= s <= 998	string
numeric(p,s), 9 <= p <= 38, 1 <= s <= 9	numeric
numrange	string
path	bytes
point	bytes
polygon	bytes
real	float64

PostgreSQL Source Data Type	Google BigQuery Target Data Type
smallint	int64
time( <i>precision</i> ) with time zone, 0 <= p <= 6	string
time( <i>precision</i> ) without time zone, 0 <= p <= 6	time
timestamp( <i>precision</i> ) with time zone, 0 <= p <= 6	timestamp
timestamp( <i>precision</i> ) without time zone, 0 <= p <= 6	datetime
tsrange	string
tstzrange	string
uuid	string

## PostgreSQL Source and Microsoft Azure Synapse Analytics Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a PostgreSQL source and a Microsoft Azure Synapse Analytics target:

PostgreSQL Source Data Type	Synapse Analytics Target Data Type
bigint	bigint
bit varying(1)	bit
bit varying( <i>precision</i> ), 2 <= p <= 83700001	varbinary( <i>size</i> ), 1 <= s <= max
bit(1)	bit
bit( <i>precision</i> ), 2 <= p <= 63501	binary( <i>size</i> ), 1 <= s <= 7938
bit( <i>precision</i> ), 64001 <= p <= 83886080	varbinary(max)
boolean	bit
box	varbinary(115)
character varying( <i>size</i> ), 1 <= s <= 10485760	varchar( <i>size</i> ), 4 <= s <= max
character( <i>size</i> ), 1 <= s <= 10485760	varchar( <i>size</i> ), 4 <= s <= max
cidr	varchar(45)
circle	varbinary(87)
date	date
daterange	varchar(29)

PostgreSQL Source Data Type	Synapse Analytics Target Data Type
double precision	float
inet	varchar(45)
int4range	varchar(25)
int8range	varchar(43)
integer	int
line	varbinary(85)
lseg	varbinary(117)
macaddr	varchar(17)
macaddr8	varchar(23)
money	decimal(19,2)
numeric(p,s), 1 <= p <= 38, 0 <= s <= 38	decimal(p,s), 1 <= p <= 38, 0 <= s <= 38
numeric(p,s), 39 <= p <= 999, 39 <= s <= 998	varchar(size), 42 <= s <= max
numrange	varchar(max)
path	varbinary(max)
point	varbinary(57)
polygon	varbinary(max)
real	real
smallint	smallint
time(precision) with time zone, 0 <= p <= 6	datetimeoffset(precision), 0 <= p <= 6
time(precision) without time zone, 0 <= p <= 6	time(precision), 0 <= p <= 6
timestamp(precision) with time zone, 0 <= p <= 6	datetimeoffset(precision), 0 <= p <= 6
timestamp(precision) without time zone, 0 <= p <= 6	datetime2(precision), 0 <= p <= 6
tsrange	varchar(63)
tstzrange	varchar(75)
uuid	uniqueidentifier

For incremental load jobs, Mass Ingestion Databases does not support the following PostgreSQL data types, in addition to those not supported for initial load jobs:

- BYTEA
- MONEY
- Spatial types
  - Box
  - Circle
  - Line
  - LSeg
  - Path
  - Point
  - Polygon
- TEXT
- Unbounded varying types
- XML

## PostgreSQL Source and Snowflake Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a PostgreSQL source and a Snowflake target:

PostgreSQL Source Data Type	Snowflake Target Data Type
bigint	integer
bit varying( <i>precision</i> ), 1 <= p <= 83700001	binary( <i>size</i> ), 1 <= s <= 8387501
bit( <i>precision</i> ), 1 <= p <= 83886080	binary( <i>size</i> ), 1 <= s <= 8388563
boolean	boolean
box	binary(115)
bytea	binary(8388608) for initial loads only
character varying( <i>size</i> ), 1 <= s <= 10485760	varchar( <i>size</i> ), 4 <= s <= 16776004
character( <i>size</i> ), 1 <= s <= 10485760	char( <i>size</i> ), 4 <= s <= 16777216
cidr	varchar(45)
circle	binary(87)
date	date
daterange	varchar(29)
double precision	float

PostgreSQL Source Data Type	Snowflake Target Data Type
inet	varchar(45)
int4range	varchar(25)
int8range	varchar(43)
integer	integer
line	binary(85)
lseg	binary(117)
macaddr	varchar(17)
macaddr8	varchar(23)
money	number(19,2)
numeric(p,s), 1 <= p <= 38, 0 <= s <= 37	number(p,s), 1 <= p <= 38, 0 <= s <= 37
numeric(p,s), 38 <= p <= 999, 38 <= s <= 998	varchar(size), 41 <= s <= 131074
numrange	varchar(294917)
path	binary
point	binary(57)
polygon	binary
real	float
smallint	integer
text	varchar(16777216) for initial loads only
time(precision) with time zone, 0 <= p <= 6	timestamp_tz(precision), 0 <= p <= 6
time(precision) without time zone, 0 <= p <= 6	time(precision), 0 <= p <= 6
timestamp(precision) with time zone, 0 <= p <= 6	timestamp_tz(precision), 0 <= p <= 6
timestamp(precision) without time zone, 0 <= p <= 6	timestamp_ntz(precision), 0 <= p <= 6
tsrange	varchar(63)
tstzrange	varchar(75)
uuid	char(36)
xml	varchar(16777216) for initial loads only

For incremental load jobs, Mass Ingestion Databases does not support the following PostgreSQL data types, in addition to those not supported for initial load jobs:

- BYTEA
- MONEY
- Spatial types
  - Box
  - Circle
  - Line
  - LSeg
  - Path
  - Point
  - Polygon
- TEXT
- Unbounded varying types
- XML

## SAP HANA Source and Amazon Redshift Target

The following table identifies the default data-type mappings for Mass Ingestion Databases configurations with a SAP HANA source and an Amazon Redshift target:

SAP HANA Source Data Type	Amazon Redshift Target Data Type
alphanum( <i>precision</i> ), 1 <= p <= 127	character( <i>size</i> ), 1 <= s <= 127
array <sup>1</sup>	character varying(65535)
bigint	bigint
binary( <i>size</i> ), 1 <= s <= 2000	character varying( <i>size</i> ), 2 <= s <= 4000
bintext <sup>1</sup>	
blob <sup>1</sup>	
boolean	boolean
char( <i>size</i> ), 1 <= s <= 2000	character( <i>size</i> ), 1 <= s <= 2000
clob <sup>1</sup>	
date	date
decimal	character varying(255)
decimal(38,38)	character varying(41)
decimal(p,s), 1 <= p <= 38, 0 <= s <= 37	numeric(p,s), 1 <= p <= 38, 0 <= s <= 37

SAP HANA Source Data Type	Amazon Redshift Target Data Type
double	double precision
float	double precision
integer	integer
nchar( <i>size</i> ), 1 <= s <= 2000	character varying( <i>size</i> ), 4 <= s <= 8000
nclob <sup>1</sup>	
nvarchar( <i>size</i> ), 1 <= s <= 5000	character varying( <i>size</i> ), 4 <= s <= 20000
real	real
real	real
seconddate	timestamp without time zone
shorttext( <i>precision</i> ), 1 <= p <= 5000	character varying( <i>size</i> ), 1 <= s <= 5000
smalldecimal	character varying(255)
smallint	smallint
st_geometry <sup>1</sup>	character varying(65535)
st_point <sup>1</sup>	character varying(65535)
text <sup>1</sup>	
time	time without time zone
timestamp	character varying(27)
timestamp	timestamp without time zone
tinyint	smallint
varbinary( <i>size</i> ), 1 <= s <= 5000	character varying( <i>size</i> ), 2 <= s <= 10000
varchar( <i>size</i> ), 1 <= s <= 5000	character varying( <i>size</i> ), 1 <= s <= 5000
1. Unsupported source data type. Mass Ingestion Databases replicates only nulls for columns that have these unsupported data types, even though they are mapped to the default target data types.	



## SAP HANA Source and Databricks Delta Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a SAP HANA source and a Databricks Delta target:

SAP HANA Source Data Type	Databricks Delta Target Data Type
alphanum( <i>precision</i> ), 1 <= p <= 127	string
array	binary
bigint	long
binary( <i>size</i> ), 1 <= s <= 2000	binary
boolean	boolean
char( <i>size</i> ), 1 <= s <= 2000	string
date	string
decimal	string
decimal(p,s), 1 <= p <= 38, 0 <= s <= 38	decimal(p,s), 1 <= p <= 38, 0 <= s <= 38
double	double
float	double
integer	integer
nchar( <i>size</i> ), 1 <= s <= 2000	string
nvarchar( <i>size</i> ), 1 <= s <= 5000	string
real	float
real	float
seconddate	timestamp
shorttext( <i>precision</i> ), 1 <= p <= 5000	string
smalldecimal	string
smallint	integer
st_geometry	binary
st_point	binary
time	string
timestamp	string
timestamp	timestamp

SAP HANA Source Data Type	Databricks Delta Target Data Type
tinyint	integer
varbinary(size), 1 <= s <= 5000	binary
varchar(size), 1 <= s <= 5000	string

## SAP HANA Source and Google BigQuery Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a SAP HANA source and a Google BigQuery target:

SAP HANA Source Data Type	Google BigQuery Target Data Type
alphanum(precision), 1 <= p <= 127	string
array	bytes
bigint	int64
binary(size), 1 <= s <= 2000	bytes
boolean	bool
char(size), 1 <= s <= 2000	string
date	date
decimal	string
decimal(1,1)	numeric
decimal(10,10)	bignumeric
decimal(11,s), 10 <= s <= 11	bignumeric
decimal(12,s), 10 <= s <= 12	bignumeric
decimal(13,s), 10 <= s <= 13	bignumeric
decimal(14,s), 10 <= s <= 14	bignumeric
decimal(15,s), 10 <= s <= 15	bignumeric
decimal(16,s), 10 <= s <= 16	bignumeric
decimal(17,s), 10 <= s <= 17	bignumeric
decimal(18,s), 10 <= s <= 18	bignumeric
decimal(19,s), 10 <= s <= 19	bignumeric
decimal(2,s), 1 <= s <= 2	numeric

SAP HANA Source Data Type	Google BigQuery Target Data Type
decimal(20,s), 10 <= s <= 20	bignumeric
decimal(21,s), 10 <= s <= 21	bignumeric
decimal(22,s), 10 <= s <= 22	bignumeric
decimal(23,s), 10 <= s <= 23	bignumeric
decimal(24,s), 10 <= s <= 24	bignumeric
decimal(25,s), 10 <= s <= 25	bignumeric
decimal(26,s), 10 <= s <= 26	bignumeric
decimal(27,s), 10 <= s <= 27	bignumeric
decimal(28,s), 10 <= s <= 28	bignumeric
decimal(3,s), 1 <= s <= 3	numeric
decimal(38,s), 10 <= s <= 38	bignumeric
decimal(4,s), 1 <= s <= 4	numeric
decimal(5,s), 1 <= s <= 5	numeric
decimal(6,s), 1 <= s <= 6	numeric
decimal(7,s), 1 <= s <= 7	numeric
decimal(8,s), 1 <= s <= 8	numeric
decimal(p,1), 18 <= p <= 0	int64
decimal(p,s), 19 <= p <= 29, 0 <= s <= 9	numeric
decimal(p,s), 29 <= p <= 30, 0 <= s <= 29	bignumeric
decimal(p,s), 30 <= p <= 31, 0 <= s <= 30	bignumeric
decimal(p,s), 31 <= p <= 32, 0 <= s <= 31	bignumeric
decimal(p,s), 32 <= p <= 33, 0 <= s <= 32	bignumeric
decimal(p,s), 33 <= p <= 34, 0 <= s <= 33	bignumeric
decimal(p,s), 34 <= p <= 35, 0 <= s <= 34	bignumeric
decimal(p,s), 35 <= p <= 36, 0 <= s <= 35	bignumeric
decimal(p,s), 36 <= p <= 37, 0 <= s <= 36	bignumeric
decimal(p,s), 37 <= p <= 38, 0 <= s <= 37	bignumeric

SAP HANA Source Data Type	Google BigQuery Target Data Type
decimal(p,s), 9 <= p <= 38, 1 <= s <= 9	numeric
double	float64
float	float64
integer	int64
nchar(size), 1 <= s <= 2000	string
nvarchar(size), 1 <= s <= 5000	string
real	float64
real	float64
seconddate	datetime
shorttext(precision), 1 <= p <= 5000	string
smalldecimal	string
smallint	int64
st_geometry	bytes
st_point	bytes
time	time
timestamp	datetime
timestamp	string
tinyint	int64
varbinary(size), 1 <= s <= 5000	bytes
varchar(size), 1 <= s <= 5000	string

## SAP HANA Source and Microsoft Azure Synapse Analytics Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a SAP HANA source and a Microsoft Azure Synapse Analytics target:

SAP HANA Source Data Type	Synapse Analytics Target Data Type
alphanum(precision), 1 <= p <= 127	char(size), 1 <= s <= 127
array	varbinary(max)
bigint	bigint

SAP HANA Source Data Type	Synapse Analytics Target Data Type
binary(size), 1 <= s <= 2000	binary(size), 1 <= s <= 2000
boolean	bit
char(size), 1 <= s <= 2000	char(size), 1 <= s <= 2000
date	date
decimal	varchar(255)
decimal(p,s), 1 <= p <= 38, 0 <= s <= 38	decimal(p,s), 1 <= p <= 38, 0 <= s <= 38
double	float
float	float
integer	int
nchar(size), 1 <= s <= 2000	char(size), 4 <= s <= 8000
nvarchar(size), 1 <= s <= 5000	varchar(size), 4 <= s <= max
real	real
real	real
seconddate	datetime2(0)
shorttext(precision), 1 <= p <= 5000	varchar(size), 1 <= s <= 5000
smalldecimal	varchar(255)
smallint	smallint
st_geometry	varbinary(max)
st_point	varbinary(max)
time	time(0)
timestamp	datetime2(precision), 0 <= p <= 7
tinyint	tinyint
varbinary(size), 1 <= s <= 5000	varbinary(size), 1 <= s <= 5000
varchar(size), 1 <= s <= 5000	varchar(size), 1 <= s <= 5000

## SAP HANA Source and Snowflake Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a SAP HANA source and a Snowflake target:

SAP HANA Source Data Type	Snowflake Target Data Type
alphanum( <i>precision</i> ), 1 <= p <= 127	char( <i>size</i> ), 1 <= s <= 127
array	binary
bigint	integer
binary( <i>size</i> ), 1 <= s <= 2000	binary( <i>size</i> ), 1 <= s <= 2000
boolean	boolean
char( <i>size</i> ), 1 <= s <= 2000	char( <i>size</i> ), 1 <= s <= 2000
date	date
decimal	char(255)
decimal(38,38)	char(41)
decimal(p,s), 1 <= p <= 38, 0 <= s <= 37	number(p,s), 1 <= p <= 38, 0 <= s <= 37
double	float
float	float
integer	integer
nchar( <i>size</i> ), 1 <= s <= 2000	char( <i>size</i> ), 4 <= s <= 8000
nvarchar( <i>size</i> ), 1 <= s <= 5000	varchar( <i>size</i> ), 4 <= s <= 20000
real	float
real	float
seconddate	timestamp_ntz(0)
shorttext( <i>precision</i> ), 1 <= p <= 5000	varchar( <i>size</i> ), 1 <= s <= 5000
smalldecimal	char(255)
smallint	integer
st_geometry	binary
st_point	binary
time	time(0)
timestamp	timestamp_ntz( <i>precision</i> ), 0 <= p <= 7

SAP HANA Source Data Type	Snowflake Target Data Type
tinyint	integer
varbinary(size), 1 <= s <= 5000	binary(size), 1 <= s <= 5000
varchar(size), 1 <= s <= 5000	varchar(size), 1 <= s <= 5000

## Teradata Source and Amazon Redshift Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Teradata source and an Amazon Redshift target:

Teradata Source Data Type	Amazon Redshift Target Data Type
array	character varying(65535)
bigint	bigint
blob	character varying(65535)
byte(precision), 1 <= p <= 64000	character varying(size), 2 <= s <= 65535
byteint	smallint
char(size), 4 <= s <= 256000	character varying(size), 4 <= s <= 65535
clob	character varying(65535)
date	date
decimal(p,s), 1 <= p <= 38, 1 <= s <= 37	numeric(p,s), 1 <= p <= 38, 0 <= s <= 37
float	double precision
integer	integer
interval day(precision) to hour, 1 <= p <= 4	character varying(size), 5 <= s <= 8
interval day(precision) to minute, 1 <= p <= 4	character varying(size), 8 <= s <= 11
interval day(precision) to second (s), 1 <= p <= 4, 0 <= s <= 6	character varying(size), 12 <= s <= 21
interval day(precision), 1 <= p <= 4	character varying(size), 2 <= s <= 5
interval hour(precision) to minute, 1 <= p <= 4	character varying(size), 5 <= s <= 8
interval hour(precision) to second (s), 1 <= p <= 4, 0 <= s <= 6	character varying(size), 9 <= s <= 18
interval hour(precision), 1 <= p <= 4	character varying(size), 2 <= s <= 5

Teradata Source Data Type	Amazon Redshift Target Data Type
interval minute( <i>precision</i> ) to second (s), 1 <= p <= 4, 0 <= s <= 6	character varying( <i>size</i> ), 9 <= s <= 18
interval minute( <i>precision</i> ), 1 <= p <= 4	character varying( <i>size</i> ), 2 <= s <= 5
interval month( <i>precision</i> ), 1 <= p <= 4	character varying( <i>size</i> ), 2 <= s <= 5
interval second(p,s), 1 <= p <= 4, 0 <= s <= 6	character varying( <i>size</i> ), 3 <= s <= 12
interval year( <i>precision</i> ) to month, 1 <= p <= 4	character varying( <i>size</i> ), 5 <= s <= 8
interval year( <i>precision</i> ), 1 <= p <= 4	character varying( <i>size</i> ), 2 <= s <= 5
json	character varying(65535)
mbr	character varying(512)
number(*,s), 0 <= s <= 37	character varying(255)
number(p,s), 1 <= p <= 38, 1 <= s <= 37	numeric(p,s), 1 <= p <= 38, 0 <= s <= 37
numeric(p,s), 1 <= p <= 38, 1 <= s <= 37	numeric(p,s), 1 <= p <= 38, 0 <= s <= 37
period(date)	character varying(28)
period(time( <i>precision</i> ) with time zone), 0 <= p <= 6	character varying( <i>size</i> ), 36 <= s <= 50
period(time( <i>precision</i> )), 0 <= p <= 6	character varying( <i>size</i> ), 24 <= s <= 38
period(timestamp( <i>precision</i> ) with time zone), 0 <= p <= 6	character varying( <i>size</i> ), 58 <= s <= 72
period(timestamp( <i>precision</i> )), 0 <= p <= 6	character varying( <i>size</i> ), 46 <= s <= 60
smallint	smallint
time( <i>precision</i> ) with time zone, 0 <= p <= 6	time with time zone
time( <i>precision</i> ), 0 <= p <= 6	time without time zone
timestamp( <i>precision</i> ) with time zone, 0 <= p <= 6	timestamp with time zone
timestamp( <i>precision</i> ), 0 <= p <= 6	timestamp without time zone
varbyte( <i>precision</i> ), 1 <= p <= 64000	character varying( <i>size</i> ), 2 <= s <= 65535
varchar( <i>size</i> ), 4 <= s <= 256000	character varying( <i>size</i> ), 4 <= s <= 65535
varray	character varying(65535)
xml	character varying(65535)



## Teradata Source and Databricks Delta Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Teradata source and a Databricks Delta target:

Teradata Source Data Type	Databricks Delta Target Data Type
array	binary
bigint	long
blob	binary
byte( <i>precision</i> ), 1 <= p <= 64000	binary
byteint	integer
char( <i>size</i> ), 4 <= s <= 256000	string
clob	string
date	string
decimal(p,s), 1 <= p <= 38, 1 <= s <= 37	decimal(p,s), 1 <= p <= 38, 1 <= s <= 37
float	double
integer	integer
interval day( <i>precision</i> ) to hour, 1 <= p <= 4	string
interval day( <i>precision</i> ) to minute, 1 <= p <= 4	string
interval day( <i>precision</i> ) to second (s), 1 <= p <= 4, 0 <= s <= 6	string
interval day( <i>precision</i> ), 1 <= p <= 4	string
interval hour( <i>precision</i> ) to minute, 1 <= p <= 4	string
interval hour( <i>precision</i> ) to second (s), 1 <= p <= 4, 0 <= s <= 6	string
interval hour( <i>precision</i> ), 1 <= p <= 4	string
interval minute( <i>precision</i> ) to second (s), 1 <= p <= 4, 0 <= s <= 6	string
interval minute( <i>precision</i> ), 1 <= p <= 4	string
interval month( <i>precision</i> ), 1 <= p <= 4	string
interval second(p,s), 1 <= p <= 4, 0 <= s <= 6	string
interval year( <i>precision</i> ) to month, 1 <= p <= 4	string
interval year( <i>precision</i> ), 1 <= p <= 4	string

Teradata Source Data Type	Databricks Delta Target Data Type
json	string
mbr	binary
number(*,s), 0 <= s <= 37	string
number(p,s), 1 <= p <= 38, 1 <= s <= 37	decimal(p,s), 1 <= p <= 38, 1 <= s <= 37
numeric(p,s), 1 <= p <= 38, 1 <= s <= 37	decimal(p,s), 1 <= p <= 38, 1 <= s <= 37
period(date)	string
period(time( <i>precision</i> ) with time zone), 0 <= p <= 6	string
period(time( <i>precision</i> )), 0 <= p <= 6	string
period(timestamp( <i>precision</i> ) with time zone), 0 <= p <= 6	string
period(timestamp( <i>precision</i> )), 0 <= p <= 6	string
smallint	integer
time( <i>precision</i> ) with time zone, 0 <= p <= 6	string
time( <i>precision</i> ), 0 <= p <= 6	string
timestamp( <i>precision</i> ) with time zone, 0 <= p <= 6	timestamp
timestamp( <i>precision</i> ), 0 <= p <= 6	timestamp
varbyte( <i>precision</i> ), 1 <= p <= 64000	binary
varchar( <i>size</i> ), 4 <= s <= 256000	string
varray	binary
xml	string

## Teradata Source and Google BigQuery Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Teradata source and a Google BigQuery target:

Teradata Source Data Type	Google BigQuery Target Data Type
array	bytes
bigint	int64

Teradata Source Data Type	Google BigQuery Target Data Type
blob	bytes
byte( <i>precision</i> ), 1 <= p <= 64000	bytes
byteint	int64
char( <i>size</i> ), 4 <= s <= 256000	string
clob	string
date	date
decimal(10,10)	bignumeric
decimal(11,s), 10 <= s <= 11	bignumeric
decimal(12,s), 10 <= s <= 12	bignumeric
decimal(13,s), 10 <= s <= 13	bignumeric
decimal(14,s), 10 <= s <= 14	bignumeric
decimal(15,s), 10 <= s <= 15	bignumeric
decimal(16,s), 10 <= s <= 16	bignumeric
decimal(17,s), 10 <= s <= 17	bignumeric
decimal(18,s), 10 <= s <= 18	bignumeric
decimal(19,s), 10 <= s <= 19	bignumeric
decimal(20,s), 10 <= s <= 20	bignumeric
decimal(21,s), 10 <= s <= 21	bignumeric
decimal(22,s), 10 <= s <= 22	bignumeric
decimal(23,s), 10 <= s <= 23	bignumeric
decimal(24,s), 10 <= s <= 24	bignumeric
decimal(25,s), 10 <= s <= 25	bignumeric
decimal(26,s), 10 <= s <= 26	bignumeric
decimal(27,s), 10 <= s <= 27	bignumeric
decimal(28,s), 10 <= s <= 28	bignumeric
decimal(29,s), 10 <= s <= 29	bignumeric
decimal(38,9)	numeric

Teradata Source Data Type	Google BigQuery Target Data Type
decimal( <i>precision</i> ), 1 <= p <= 18	int64
decimal( <i>precision</i> ), 19 <= p <= 29	numeric
decimal( <i>precision</i> ), 30 <= p <= 38	bignumeric
decimal(p,s), 1 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 11 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 12 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 13 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 14 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 15 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 16 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 17 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 18 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 19 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 20 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 21 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 22 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 23 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 24 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 25 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 26 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 27 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 28 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 29 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 30 <= p <= 31, 1 <= s <= 30	bignumeric
decimal(p,s), 30 <= p <= 38, 1 <= s <= 9	numeric
decimal(p,s), 31 <= p <= 32, 1 <= s <= 31	bignumeric
decimal(p,s), 31 <= p <= 38, 2 <= s <= 9	numeric

Teradata Source Data Type	Google BigQuery Target Data Type
decimal(p,s), 32 <= p <= 33, 1 <= s <= 32	bignumeric
decimal(p,s), 32 <= p <= 38, 3 <= s <= 9	numeric
decimal(p,s), 33 <= p <= 34, 1 <= s <= 33	bignumeric
decimal(p,s), 33 <= p <= 38, 4 <= s <= 9	numeric
decimal(p,s), 34 <= p <= 35, 1 <= s <= 34	bignumeric
decimal(p,s), 34 <= p <= 38, 5 <= s <= 9	numeric
decimal(p,s), 35 <= p <= 36, 1 <= s <= 35	bignumeric
decimal(p,s), 35 <= p <= 38, 6 <= s <= 9	numeric
decimal(p,s), 36 <= p <= 37, 1 <= s <= 36	bignumeric
decimal(p,s), 36 <= p <= 38, 7 <= s <= 9	numeric
decimal(p,s), 37 <= p <= 38, 1 <= s <= 37	bignumeric
decimal(p,s), 37 <= p <= 38, 8 <= s <= 9	numeric
float	float64
integer	int64
interval day( <i>precision</i> ) to hour, 1 <= p <= 4	string
interval day( <i>precision</i> ) to minute, 1 <= p <= 4	string
interval day( <i>precision</i> ) to second (s), 1 <= p <= 4, 0 <= s <= 6	string
interval day( <i>precision</i> ), 1 <= p <= 4	string
interval hour( <i>precision</i> ) to minute, 1 <= p <= 4	string
interval hour( <i>precision</i> ) to second (s), 1 <= p <= 4, 0 <= s <= 6	string
interval hour( <i>precision</i> ), 1 <= p <= 4	string
interval minute( <i>precision</i> ) to second (s), 1 <= p <= 4, 0 <= s <= 6	string
interval minute( <i>precision</i> ), 1 <= p <= 4	string
interval month( <i>precision</i> ), 1 <= p <= 4	string
interval second(p,s), 1 <= p <= 4, 0 <= s <= 6	string
interval year( <i>precision</i> ) to month, 1 <= p <= 4	string

Teradata Source Data Type	Google BigQuery Target Data Type
interval year( <i>precision</i> ), 1 <= p <= 4	string
json	string
mbr	bytes
number(*,s), 0 <= s <= 37	string
number(10,10)	bignumeric
number(11,s), 10 <= s <= 11	bignumeric
number(12,s), 10 <= s <= 12	bignumeric
number(13,s), 10 <= s <= 13	bignumeric
number(14,s), 10 <= s <= 14	bignumeric
number(15,s), 10 <= s <= 15	bignumeric
number(16,s), 10 <= s <= 16	bignumeric
number(17,s), 10 <= s <= 17	bignumeric
number(18,s), 10 <= s <= 18	bignumeric
number(19,s), 10 <= s <= 19	bignumeric
number(20,s), 10 <= s <= 20	bignumeric
number(21,s), 10 <= s <= 21	bignumeric
number(22,s), 10 <= s <= 22	bignumeric
number(23,s), 10 <= s <= 23	bignumeric
number(24,s), 10 <= s <= 24	bignumeric
number(25,s), 10 <= s <= 25	bignumeric
number(26,s), 10 <= s <= 26	bignumeric
number(27,s), 10 <= s <= 27	bignumeric
number(28,s), 10 <= s <= 28	bignumeric
number(29,s), 10 <= s <= 29	bignumeric
number(38,9)	numeric
number( <i>precision</i> ), 1 <= p <= 18	int64
number( <i>precision</i> ), 19 <= p <= 29	numeric

Teradata Source Data Type	Google BigQuery Target Data Type
number( <i>precision</i> ), 30 <= p <= 36	bignumeric
number(p,s), 1 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 11 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 12 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 13 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 14 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 15 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 16 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 17 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 18 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 19 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 20 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 21 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 22 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 23 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 24 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 25 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 26 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 27 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 28 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 29 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 30 <= p <= 31, 1 <= s <= 30	bignumeric
number(p,s), 30 <= p <= 38, 1 <= s <= 9	numeric
number(p,s), 31 <= p <= 32, 1 <= s <= 31	bignumeric
number(p,s), 31 <= p <= 38, 2 <= s <= 9	numeric
number(p,s), 32 <= p <= 33, 1 <= s <= 32	bignumeric
number(p,s), 32 <= p <= 38, 3 <= s <= 9	numeric

Teradata Source Data Type	Google BigQuery Target Data Type
number(p,s), 33 <= p <= 34, 1 <= s <= 33	bignumeric
number(p,s), 33 <= p <= 38, 4 <= s <= 9	numeric
number(p,s), 34 <= p <= 35, 1 <= s <= 34	bignumeric
number(p,s), 34 <= p <= 38, 5 <= s <= 9	numeric
number(p,s), 35 <= p <= 36, 1 <= s <= 35	bignumeric
number(p,s), 35 <= p <= 38, 6 <= s <= 9	numeric
number(p,s), 36 <= p <= 37, 1 <= s <= 36	bignumeric
number(p,s), 36 <= p <= 38, 7 <= s <= 9	numeric
number(p,s), 37 <= p <= 38, 1 <= s <= 37	bignumeric
number(p,s), 37 <= p <= 38, 8 <= s <= 9	numeric
numeric(10,10)	bignumeric
numeric(11,s), 10 <= s <= 11	bignumeric
numeric(12,s), 10 <= s <= 12	bignumeric
numeric(13,s), 10 <= s <= 13	bignumeric
numeric(14,s), 10 <= s <= 14	bignumeric
numeric(15,s), 10 <= s <= 15	bignumeric
numeric(16,s), 10 <= s <= 16	bignumeric
numeric(17,s), 10 <= s <= 17	bignumeric
numeric(18,s), 10 <= s <= 18	bignumeric
numeric(19,s), 10 <= s <= 19	bignumeric
numeric(20,s), 10 <= s <= 20	bignumeric
numeric(21,s), 10 <= s <= 21	bignumeric
numeric(22,s), 10 <= s <= 22	bignumeric
numeric(23,s), 10 <= s <= 23	bignumeric
numeric(24,s), 10 <= s <= 24	bignumeric
numeric(25,s), 10 <= s <= 25	bignumeric
numeric(26,s), 10 <= s <= 26	bignumeric



Teradata Source Data Type	Google BigQuery Target Data Type
numeric(27,s), 10 <= s <= 27	bignumeric
numeric(28,s), 10 <= s <= 28	bignumeric
numeric(29,s), 10 <= s <= 29	bignumeric
numeric(38,9)	numeric
numeric(precision), 1 <= p <= 18	int64
numeric(precision), 19 <= p <= 29	numeric
numeric(precision), 30 <= p <= 36	bignumeric
numeric(p,s), 1 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 11 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 12 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 13 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 14 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 15 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 16 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 17 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 18 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 19 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 20 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 21 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 22 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 23 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 24 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 25 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 26 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 27 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 28 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 29 <= p <= 38, 1 <= s <= 9	numeric

Teradata Source Data Type	Google BigQuery Target Data Type
numeric(p,s), 30 <= p <= 31, 1 <= s <= 30	bignumeric
numeric(p,s), 30 <= p <= 38, 1 <= s <= 9	numeric
numeric(p,s), 31 <= p <= 32, 1 <= s <= 31	bignumeric
numeric(p,s), 31 <= p <= 38, 2 <= s <= 9	numeric
numeric(p,s), 32 <= p <= 33, 1 <= s <= 32	bignumeric
numeric(p,s), 32 <= p <= 38, 3 <= s <= 9	numeric
numeric(p,s), 33 <= p <= 34, 1 <= s <= 33	bignumeric
numeric(p,s), 33 <= p <= 38, 4 <= s <= 9	numeric
numeric(p,s), 34 <= p <= 35, 1 <= s <= 34	bignumeric
numeric(p,s), 34 <= p <= 38, 5 <= s <= 9	numeric
numeric(p,s), 35 <= p <= 36, 1 <= s <= 35	bignumeric
numeric(p,s), 35 <= p <= 38, 6 <= s <= 9	numeric
numeric(p,s), 36 <= p <= 37, 1 <= s <= 36	bignumeric
numeric(p,s), 36 <= p <= 38, 7 <= s <= 9	numeric
numeric(p,s), 37 <= p <= 38, 1 <= s <= 37	bignumeric
numeric(p,s), 37 <= p <= 38, 8 <= s <= 9	numeric
period(date)	string
period(time( <i>precision</i> ) with time zone), 0 <= p <= 6	string
period(time( <i>precision</i> )), 0 <= p <= 6	string
period(timestamp( <i>precision</i> ) with time zone), 0 <= p <= 6	string
period(timestamp( <i>precision</i> )), 0 <= p <= 6	string
smallint	int64
time( <i>precision</i> ) with time zone, 0 <= p <= 6	timestamp
time( <i>precision</i> ), 0 <= p <= 6	time
timestamp( <i>precision</i> ) with time zone, 0 <= p <= 6	timestamp
timestamp( <i>precision</i> ), 0 <= p <= 6	datetime

Teradata Source Data Type	Google BigQuery Target Data Type
varbyte( <i>precision</i> ), 1 <= p <= 64000	bytes
varchar( <i>size</i> ), 4 <= s <= 256000	string
varray	bytes
xml	string

## Teradata Source and Microsoft Azure Synapse Analytics Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Teradata source and a Microsoft Azure Synapse Analytics target:

Teradata Source Data Type	Synapse Analytics Target Data Type
array	varbinary(max)
bigint	bigint
blob	varbinary(max)
byte( <i>precision</i> ), 1 <= p <= 7501	binary( <i>size</i> ), 1 <= s <= 7501
byte( <i>precision</i> ), 8001 <= p <= 64000	varbinary(max)
byteint	smallint
char( <i>size</i> ), 128004 <= s <= 256000	nvarchar(max)
char( <i>size</i> ), 4 <= s <= 128000	char( <i>size</i> ), 4 <= s <= 32000
clob	varchar(max)
date	date
decimal(p,s), 1 <= p <= 38, 1 <= s <= 37	decimal(p,s), 1 <= p <= 38, 1 <= s <= 37
float	float
integer	int
interval day( <i>precision</i> ) to hour, 1 <= p <= 4	varchar( <i>size</i> ), 5 <= s <= 8
interval day( <i>precision</i> ) to minute, 1 <= p <= 4	varchar( <i>size</i> ), 8 <= s <= 11
interval day( <i>precision</i> ) to second (s), 1 <= p <= 4, 0 <= s <= 6	varchar( <i>size</i> ), 12 <= s <= 21
interval day( <i>precision</i> ), 1 <= p <= 4	varchar( <i>size</i> ), 2 <= s <= 5
interval hour( <i>precision</i> ) to minute, 1 <= p <= 4	varchar( <i>size</i> ), 5 <= s <= 8

Teradata Source Data Type	Synapse Analytics Target Data Type
interval hour( <i>precision</i> ) to second (s), 1 <= p <= 4, 0 <= s <= 6	varchar( <i>size</i> ), 9 <= s <= 18
interval hour( <i>precision</i> ), 1 <= p <= 4	varchar( <i>size</i> ), 2 <= s <= 5
interval minute( <i>precision</i> ) to second (s), 1 <= p <= 4, 0 <= s <= 6	varchar( <i>size</i> ), 9 <= s <= 18
interval minute( <i>precision</i> ), 1 <= p <= 4	varchar( <i>size</i> ), 2 <= s <= 5
interval month( <i>precision</i> ), 1 <= p <= 4	varchar( <i>size</i> ), 2 <= s <= 5
interval second(p,s), 1 <= p <= 4, 0 <= s <= 6	varchar( <i>size</i> ), 3 <= s <= 12
interval year( <i>precision</i> ) to month, 1 <= p <= 4	varchar( <i>size</i> ), 5 <= s <= 8
interval year( <i>precision</i> ), 1 <= p <= 4	varchar( <i>size</i> ), 2 <= s <= 5
json	varchar(max)
mbr	varbinary(256)
number(*,s), 0 <= s <= 37	varchar(255)
number(p,s), 1 <= p <= 38, 1 <= s <= 37	decimal(p,s), 1 <= p <= 38, 1 <= s <= 37
numeric(p,s), 1 <= p <= 38, 1 <= s <= 37	decimal(p,s), 1 <= p <= 38, 1 <= s <= 37
period(date)	varchar(28)
period(time( <i>precision</i> ) with time zone), 0 <= p <= 6	varchar( <i>size</i> ), 36 <= s <= 50
period(time( <i>precision</i> )), 0 <= p <= 6	varchar( <i>size</i> ), 24 <= s <= 38
period(timestamp( <i>precision</i> ) with time zone), 0 <= p <= 6	varchar( <i>size</i> ), 58 <= s <= 72
period(timestamp( <i>precision</i> )), 0 <= p <= 6	varchar( <i>size</i> ), 46 <= s <= 60
smallint	smallint
time( <i>precision</i> ) with time zone, 0 <= p <= 6	datetimeoffset( <i>precision</i> ), 0 <= p <= 6
time( <i>precision</i> ), 0 <= p <= 6	time( <i>precision</i> ), 0 <= p <= 6
timestamp( <i>precision</i> ) with time zone, 0 <= p <= 6	datetimeoffset( <i>precision</i> ), 0 <= p <= 6
timestamp( <i>precision</i> ), 0 <= p <= 6	datetime2( <i>precision</i> ), 0 <= p <= 6
varbyte( <i>precision</i> ), 1 <= p <= 64000	varbinary( <i>size</i> ), 1 <= s <= max
varchar( <i>size</i> ), 128004 <= s <= 256000	nvarchar(max)

Teradata Source Data Type	Synapse Analytics Target Data Type
<code>varchar(size), 4 &lt;= s &lt;= 126004</code>	<code>char(size), 4 &lt;= s &lt;= 32000</code>
<code>varray</code>	<code>varbinary(max)</code>
<code>xml</code>	<code>varchar(max)</code>

## Teradata Source and Snowflake Target

The following table identifies the recommended data-type mappings for Mass Ingestion Databases configurations with a Teradata source and a Snowflake target:

Teradata Source Data Type	Snowflake Target Data Type
<code>array</code>	<code>binary</code>
<code>bigint</code>	<code>integer</code>
<code>blob</code>	<code>binary</code>
<code>byte(precision), 1 &lt;= p &lt;= 64000</code>	<code>binary(size), 1 &lt;= s &lt;= 64000</code>
<code>byteint</code>	<code>integer</code>
<code>char(size), 4 &lt;= s &lt;= 256000</code>	<code>varchar(size), 4 &lt;= s &lt;= 256000</code>
<code>clob</code>	<code>varchar</code>
<code>date</code>	<code>date</code>
<code>decimal(precision), 1 &lt;= p &lt;= 38</code>	<code>integer</code>
<code>decimal(p,s), 1 &lt;= p &lt;= 38, 1 &lt;= s &lt;= 37</code>	<code>number(p,s), 1 &lt;= p &lt;= 38, 1 &lt;= s &lt;= 37</code>
<code>float</code>	<code>float</code>
<code>integer</code>	<code>integer</code>
<code>interval day(precision) to hour, 1 &lt;= p &lt;= 4</code>	<code>varchar(size), 5 &lt;= s &lt;= 8</code>
<code>interval day(precision) to minute, 1 &lt;= p &lt;= 4</code>	<code>varchar(size), 8 &lt;= s &lt;= 11</code>
<code>interval day(precision) to second (s), 1 &lt;= p &lt;= 4, 0 &lt;= s &lt;= 6</code>	<code>varchar(size), 12 &lt;= s &lt;= 21</code>
<code>interval day(precision), 1 &lt;= p &lt;= 4</code>	<code>varchar(size), 2 &lt;= s &lt;= 5</code>
<code>interval hour(precision) to minute, 1 &lt;= p &lt;= 4</code>	<code>varchar(size), 5 &lt;= s &lt;= 8</code>
<code>interval hour(precision) to second (s), 1 &lt;= p &lt;= 4, 0 &lt;= s &lt;= 6</code>	<code>varchar(size), 9 &lt;= s &lt;= 18</code>
<code>interval hour(precision), 1 &lt;= p &lt;= 4</code>	<code>varchar(size), 2 &lt;= s &lt;= 5</code>

Teradata Source Data Type	Snowflake Target Data Type
interval minute( <i>precision</i> ) to second (s), 1 <= p <= 4, 0 <= s <= 6	varchar( <i>size</i> ), 9 <= s <= 18
interval minute( <i>precision</i> ), 1 <= p <= 4	varchar( <i>size</i> ), 2 <= s <= 5
interval month( <i>precision</i> ), 1 <= p <= 4	varchar( <i>size</i> ), 2 <= s <= 5
interval second(p,s), 1 <= p <= 4, 0 <= s <= 6	varchar( <i>size</i> ), 3 <= s <= 12
interval year( <i>precision</i> ) to month, 1 <= p <= 4	varchar( <i>size</i> ), 5 <= s <= 8
interval year( <i>precision</i> ), 1 <= p <= 4	varchar( <i>size</i> ), 2 <= s <= 5
json	varchar
mbr	binary(256)
number(*,s), 0 <= s <= 37	char(255)
number( <i>precision</i> ), 1 <= p <= 36	integer
number(p,s), 1 <= p <= 38, 1 <= s <= 37	number(p,s), 1 <= p <= 38, 1 <= s <= 37
numeric( <i>precision</i> ), 1 <= p <= 36	integer
numeric(p,s), 1 <= p <= 38, 1 <= s <= 37	number(p,s), 1 <= p <= 38, 1 <= s <= 37
period(date)	varchar(28)
period(time( <i>precision</i> ) with time zone), 0 <= p <= 6	varchar( <i>size</i> ), 36 <= s <= 50
period(time( <i>precision</i> )), 0 <= p <= 6	varchar( <i>size</i> ), 24 <= s <= 38
period(timestamp( <i>precision</i> ) with time zone), 0 <= p <= 6	varchar( <i>size</i> ), 58 <= s <= 72
period(timestamp( <i>precision</i> )), 0 <= p <= 6	varchar( <i>size</i> ), 46 <= s <= 60
smallint	integer
time( <i>precision</i> ) with time zone, 0 <= p <= 6	timestamp_tz( <i>precision</i> ), 0 <= p <= 6
time( <i>precision</i> ), 0 <= p <= 6	time( <i>precision</i> ), 0 <= p <= 6
timestamp( <i>precision</i> ) with time zone, 0 <= p <= 6	timestamp_tz( <i>precision</i> ), 0 <= p <= 6
timestamp( <i>precision</i> ), 0 <= p <= 6	timestamp_ntz( <i>precision</i> ), 0 <= p <= 6
varbyte( <i>precision</i> ), 1 <= p <= 64000	binary( <i>size</i> ), 1 <= s <= 64000
varchar( <i>size</i> ), 4 <= s <= 256000	varchar( <i>size</i> ), 4 <= s <= 256000

Teradata Source Data Type	Snowflake Target Data Type
varray	binary
xml	varchar

**Note:** The Snowflake TIMESTAMP\_TZ data type includes a default date that the source Teradata TIME WITH TIME ZONE data type does not include. For example, a database ingestion job will replicate the source value "12:59:59" as "1900-01-01 12:59:59".

## CHAPTER 6

# Mass Ingestion Files

Use file ingestion tasks to transfer a large number of files of any file type between on-premises and cloud repositories and to track and monitor file transfers.

When you create a file ingestion task, you define the source and the target endpoints for file transfer. You can configure a file ingestion task to transfer multiple files in a batch to enhance performance.

You can define a schedule by which the task runs. You can also configure the task to perform actions, such as compression, decompression, encryption, or decryption of files.

## Use cases

Mass Ingestion Files can help you to migrate data from on-premises or cloud-based systems to cloud-based systems.

## Mass Ingestion Files source types

You can transfer files from on-premises and cloud sources that Mass Ingestion Files support to supported targets.

You can add a source to a file ingestion task in the following ways:

### Through data catalog discovery

If your organization administrator has configured Enterprise Data Catalog integration properties, you can perform data catalog discovery to find the source object in the catalog. You can discover Amazon S3 V2, Microsoft Azure Blob Storage V3, or Hadoop Files V2 objects to use as sources in new file ingestion tasks.

Search for the source object on the **Data Catalog** page, select the object in the search results, and then add it to a new file ingestion task.

### When you configure the task

Select the source connection and source object on the **Source** tab when you configure the file ingestion task.



Mass Ingestion Files supports the following source types:

- Source Connection— Supports the following sources:
  - Local folder
  - Advanced FTP V2
  - Advanced FTPS V2
  - Advanced SFTP V2
  - Amazon S3 V2
  - Google Cloud Storage V2
  - Hadoop Files V2
  - Microsoft Azure Blob Storage V3
  - Microsoft Azure Data Lake Store Gen2
  - Microsoft Azure Data Lake Store V3
  - Databricks Delta
- File Listener— Use a file listener component as a source.

To determine the connectors to use for these source types, see *Connectors and Connections > Mass Ingestion Files connectors*.

For more information, see ["Configuring the source" on page 397](#).

## Mass Ingestion Files target types

You can transfer files from any Mass Ingestion Files-supported sources to on-premises and cloud targets that Mass Ingestion Files supports.

Mass Ingestion Files supports the following targets:

- Local folder
- Advanced FTP V2
- Advanced FTPS V2
- Advanced SFTP V2
- Amazon S3 V2
- Amazon Redshift V2
- Google BigQuery V2
- Google Cloud Storage V2
- Hadoop Files V2
- Microsoft Azure Blob Storage V3
- Microsoft Azure Data Lake Store Gen2
- Microsoft Azure Data Lake Store V3
- Microsoft Azure Synapse SQL
- Snowflake Data Cloud
- Databricks Delta

To determine the connectors to use for these target types, see *Connectors and Connections > Mass Ingestion Files connectors*.

## Mass Ingestion Files actions

When you configure a file ingestion task, you can define file-processing actions, such as compress, decompress, encrypt, and decrypt to be performed before transferring the files. You can also scan files for viruses.

You can perform the following file-processing actions on the files that the file ingestion task transfers:

- **Compress.** Uses one of the following methods to compress files: Zip, Tar, or Gzip. The file ingestion job compresses files and flattens the file structure in the target directory.
- **Decompress.** Uses one of the following methods to decompress compressed files: Unzip, Untar, or Gunzip. The file ingestion job decompresses files and flattens the file structure in the target directory.
- **Encrypt.** Uses the Pretty Good Privacy (PGP) method to encrypt files. The file ingestion job encrypts files and flattens the file structure in the target directory. The Gnu Privacy Guard (GPG) method is compatible with the PGP method to encrypt files.  
For more information about securing the files that the file ingestion job transfers, refer to [“Mass Ingestion Files security” on page 395](#).
- **Decrypt.** Uses the PGP method to decrypt files. The file ingestion task decrypts files and flattens the file structure in the target directory. The GPG method is compatible with the PGP method to decrypt files.  
For more information about securing the files that the file ingestion job transfers, refer to [“Mass Ingestion Files security” on page 395](#).
- **Flatten file structure.** Moves the files from multiple folders to a single folder in the target directory. File-structure flattening might result in the loss of files if files with the same file name exist in different folders. The session log displays the overridden files.
- **Virus scan.** Identifies viruses and malware in the files that the file ingestion job transfers by using the Internet Content Adaptation Protocol (ICAP). The ICAP server scans the files and sends a response code of 200 when the scan does not identify any virus in the files. The file ingestion job fails when the scan detects a virus.

**Note:** The file ingestion task does not flatten the file structure if you do not configure any action.

The file ingestion job performs the file-processing actions in the order you configure them in the task definition.

For example, you want to compress and encrypt files prior to transferring them from a local repository to an FTP server.

In this scenario, add the following file processing actions:

1. Compress with an action type of Zip.
2. Encrypt with an action type of PGP.

You can add multiple file processing actions to a file ingestion task. You can drag and drop the order of the file processing action.

# Mass Ingestion Files runtime options

You can run a file ingestion task manually. You can also schedule the task to run at a specific time or when a file is ready.

You can choose to receive notifications if the task fails and if the task detects infected files.

A file ingestion task can have multiple jobs. You can run multiple jobs simultaneously to enhance the performance and scalability of a file ingestion task. You can configure a file ingestion task to run multiple jobs concurrently or use the job resource of the Mass Ingestion Files REST API to run multiple jobs concurrently.

**Warning:** Running concurrent jobs might cause unexpected results if the targets include duplicate files.

You can run a batch of files or multiple batches in parallel to reduce the duration of processing a large number of files. The maximum number of batches you can run in parallel depends on the `fmi-task-max-pool-size` and `mi-task-core-pool-size` properties that you configure for the runtime environment in Administrator. The `fmi-task-max-pool-size` property determines the maximum number of threads to execute a file ingestion task. A thread count close to the maximum value of `fmi-task-max-pool-size` might impact the performance of other jobs running on the same Secure Agent.

For more information about scheduling a file ingestion task and running parallel batch, see [“Configuring runtime options” on page 445](#).

For more information about running concurrent jobs using the job resource, see *REST API Reference*.

## Mass Ingestion Files security

Use the encryption and decryption methods to secure files that a file ingestion job transfers.

When you define the file ingestion task, you can specify the encryption and decryption methods to use. For more information, see [“Mass Ingestion Files actions” on page 394](#).

File ingestion jobs use the PGP method to encrypt and decrypt files. To encrypt files, you must provide a key ID. To decrypt files, you must provide a key passphrase.

If your user privileges allow you to update files in the agent location, you can use the key ring command line interface (CLI) to manage key IDs. For more information, see [“Key ring command reference” on page 446](#). If you do not have the privilege to access the agent location, ask your administrator for the key ID and key passphrase.

**Note:** GPG method is compatible with the PGP method to encrypt and decrypt files.

### Encryption

When you configure the encryption action for a file ingestion task, you provide a key ID. The key ID is the public key ID of the receiver who decrypts the file. You can also add your private key ID and key passphrase to sign the files.

### Decryption

When you configure the decryption action for a file ingestion task, you provide a key passphrase. The key passphrase is the private key passphrase of the receiver who decrypts the file.

# Configuring a file ingestion task

Use the task wizard to configure a file ingestion task.

In the wizard, perform the following actions:

1. Define the task.
2. Configure the source.
3. Configure the target.
4. Optionally, configure one or more file-processing actions.
5. Optionally, set the runtime options.

As you work through the task wizard, you can click **Save** to save your work at any time. When you have completed the wizard, click **Finish** to save the task and close the wizard.

Before you begin, verify that the prerequisites are met. For more information, see [“Before you begin” on page 396](#).

## Before you begin

Before you create file ingestion tasks, verify that the following conditions exist:

- Check that your organization has licenses for Mass Ingestion Files and the FMI packages.
- The Mass ingestion application is running on the Secure Agent.
- Source and target connections exist, based on the sources from where you want to transfer files and the targets to where you want to transfer files.

## Defining basic task information

To begin defining a file ingestion task, you must first enter some basic information about the task, such as task name and project or project folder location.

1. To define a file ingestion task, click **New > File Ingestion Task**.  
The **Definition** page of the File Ingestion Task wizard appears.
2. Configure the following properties:

Property	Description
Task Name	Name of the file ingestion task. The names of file ingestion tasks must be unique within the organization. Task names can contain alphanumeric characters, spaces, and underscores. Names must begin with an alphabetic character or underscore. Task names are not case sensitive.
Location	Project or folder in which the task will reside.
Description	Optional description of the task. Maximum length is 1024 characters.
Runtime Environment	Runtime environment that contains the Secure Agent used to run the task. The file ingestion application must run on the Secure Agent.

3. Click **Next**.

To edit a file ingestion task, on the **Explore** page, navigate to the task. In the row that contains the task, from the **Actions** menu, select **Edit**.

## Configuring the source

To configure the source, select a source type and a source connection from which to transfer files and then configure source options.

1. On the **Source** page, select the source type.
2. Select a source connection type and a source connection.

The file ingestion task supports the following source connection types:

- Local folder
- Advanced FTP V2
- Advanced FTPS V2
- Advanced SFTP V2
- Amazon S3 V2
- Google Cloud Storage V2
- Hadoop Files V2
- Microsoft Azure Blob Storage V3
- Microsoft Azure Data Lake Store Gen2
- Microsoft Azure Data Lake Store V3
- Databricks Delta

3. Based on the source connection that you select, enter the source options.

Options that appear on the **Source** tab of the task wizard vary based on the type of source connection that you select.

4. Click **Next**.

The **Target** tab appears.

## Advanced FTP V2 source properties

When you define a file ingestion task with an Advanced FTP V2 source, you must enter source options on the **Source** tab of the task wizard. The options vary based on the file pickup method that you select for the task.

**Note:** You can overwrite the file name pattern, folder, and table parameters, and define your own variable for sources by using the job resource of the Mass Ingestion Files REST API. For more information, see [Mass Ingestion Files REST API](#).

The following table describes the source options:

Option	Description
File Pickup	The file ingestion task supports the following file pickup methods: <ul style="list-style-type: none"> <li>- <b>By Pattern.</b> The file ingestion task picks up files by pattern.</li> <li>- <b>By File List.</b> The file ingestion task picks up files based on a file list.</li> </ul>
Source Directory	Directory from where files are transferred. The default value is the source directory specified in the connection. <p>You can enter a relative path to the source file system. To enter a relative path, start the path with a period, followed by a slash (/). The path is relative to the source directory specified in the connection.</p>
Add Parameters	Create an expression and add it as a <b>Source Directory</b> parameter. For more information, see <a href="#">"Source and target parameters" on page 420</a> .
File Pattern	This applies when <b>File Pickup</b> is <b>By Pattern</b> . File name pattern to use for selecting the files to transfer. The pattern can be a regular expression or a pattern with wildcard characters. <p>The following wildcard characters are allowed:</p> <ul style="list-style-type: none"> <li>- An asterisk (*) to represent any number of characters.</li> <li>- A question mark (?) to represent a single character.</li> </ul> <p>For example, you can specify the following regular expression:</p> <pre>([a-zA-Z0-9\s_\.\-\\(\):])+(\.doc \.docx \.pdf)\$</pre>
File Date	This applies when <b>File Pickup</b> is <b>By Pattern</b> . A date and time expression for filtering the files to transfer. <p>Select one of the following options:</p> <ul style="list-style-type: none"> <li>- <b>Greater than or Equal.</b> Filters files that are modified on or after the specified date and time. To specify a date, click the calendar. To specify a time, click the clock.</li> <li>- <b>Less than or Equal.</b> Filters files that are modified before or on the specified date and time.</li> <li>- <b>Equal.</b> Filters files that are modified on the specified date and time. Click the calendar to select the date and the clock to select the time.</li> <li>- <b>Days before today.</b> Filters files that are modified within the specified number of days until the current date (today). Enter the number of days. The current date calculation starts from 00:00 hours.</li> </ul> <p>For example, if you schedule the file ingestion task to run weekly and want to filter for the files that were modified in the previous week, set <b>Days before today</b> to 7. The task will pick up any file with a date between 7 days ago and the date on which it runs.</p>
Time Zone	This applies when <b>File Pickup</b> is <b>By Pattern</b> . If you selected a <b>File Date</b> option, enter the time zone of the location where the files are located.
File Size	This applies when <b>File Pickup</b> is <b>By Pattern</b> . Filters the files to transfer based on file size. Enter the file size, select the file size unit, and filter options. <p>Select one of the following filter options:</p> <ul style="list-style-type: none"> <li>- <b>Greater than or Equal.</b> Filters files that are greater than or equal to the specified size.</li> <li>- <b>Less than or Equal.</b> Filters files that are less than or equal to the specified size.</li> <li>- <b>Equal.</b> Filters files that have the specified size.</li> </ul>
The file path containing the list of files	This applies when <b>File Pickup</b> is <b>By File List</b> . Select this option to provide the path that contains the list of files to pick up and enter the file path.
File list	This applies when <b>File Pickup</b> is <b>By File List</b> . Select this option to provide the list of files to pick up and enter a comma-separated list of file names.

Option	Description
Skip Duplicate Files	Indicates whether to skip duplicate files. If you select this option, the file ingestion task does not transfer files that have the same name and creation date as another file. The file ingestion task marks these files as duplicate in the job log. If you do not select this option, the task transfers all files, even files with duplicate names and creation dates.
Check file stability	Indicates whether to verify that a file is stable before a file ingestion task attempts to pick it. The task skips unstable files it detects in the current run.
Stability check interval	<p>This applies when you enable the <b>Check file stability</b> option. Time in seconds that a file ingestion task waits to check the file stability.</p> <p>For example, if the stability time is 15 seconds, the file ingestion task detects all the files in the source folder that match the defined file pattern, it waits for 15 seconds, and then it processes only the stable files.</p> <p>The interval ranges between 10 seconds to 300 seconds. Default is 10 seconds.</p>
Batch Size	<p>The number of files a file ingestion task can transfer in a batch.</p> <p>Default is 5.</p> <p>The maximum value of the batch varies based on whether the files are transferred through an intermediate staging area.</p>
Transfer Mode	<p>File transfer mode. Select one of the following filter modes:</p> <ul style="list-style-type: none"> <li>- <b>Auto.</b> Mass Ingestion Files determines the transfer mode.</li> <li>- <b>ASCII.</b></li> <li>- <b>Binary.</b></li> </ul>
After File Pickup	<p>Determines what to do with the source files after the files are transferred.</p> <p>Select one of the following filter options:</p> <ul style="list-style-type: none"> <li>- Keep the files in the source directory.</li> <li>- Delete the files from the source directory.</li> <li>- Rename the files in the source directory. You must specify a file name suffix that file ingestion task adds to the file name when renaming the files.</li> <li>- Archive the files to a different location. You must specify an archive directory which is the absolute path or relative path from the source file system.</li> </ul>

## Advanced FTPS V2 source properties

When you define a file ingestion task with an Advanced FTPS V2 source, you must enter source properties on the **Source** tab of the task wizard. The options vary based on the file pickup method that you select for the task.

**Note:** You can overwrite the file name pattern, folder, and table parameters, and define your own variable for sources by using the job resource of the Mass Ingestion Files REST API. For more information, see [Mass Ingestion Files REST API](#).

The following table describes the source options:

Option	Description
File Pickup	The file ingestion task supports the following file pickup methods: <ul style="list-style-type: none"> <li>- <b>By Pattern.</b> The file ingestion task picks up files by pattern.</li> <li>- <b>By File List.</b> The file ingestion task picks up files based on a file list.</li> </ul>
Source Directory	Directory from where files are transferred. The default value is the source directory specified in the connection. You can enter a relative path to the source file system. To enter a relative path, start the path with a period, followed by a slash (. /). The path is relative to the source directory specified in the connection.
Add Parameters	Create an expression to add it as a <b>Source Directory</b> parameter. For more information, see <a href="#">“Source and target parameters” on page 420</a> .
Include files from sub-folders	This applies when <b>File Pickup</b> is <b>By Pattern</b> . Transfer files from all subfolders under the defined source directory.
File Pattern	This applies when <b>File Pickup</b> is <b>By Pattern</b> . File name pattern to use for selecting the files to transfer. The pattern can be a regular expression or a pattern with wildcard characters. The following wildcard characters are allowed: <ul style="list-style-type: none"> <li>- An asterisk (*) to represent any number of characters.</li> <li>- A question mark (?) to represent a single character.</li> </ul> For example, you can specify the following regular expression: <code>([a-zA-Z0-9\s_\.\-\(\)]+)(.doc .docx .pdf)\$</code>
File Date	This applies when <b>File Pickup</b> is <b>By Pattern</b> . A date and time expression for filtering the files to transfer. Select one of the following options: <ul style="list-style-type: none"> <li>- <b>Greater than or Equal.</b> Filters files that are modified on or after the specified date and time. To specify a date, click the calendar. To specify a time, click the clock.</li> <li>- <b>Less than or Equal.</b> Filters files that are modified before or on the specified date and time.</li> <li>- <b>Equal.</b> Filters files that are modified on the specified date and time. Click the calendar to select the date and the clock to select the time.</li> <li>- <b>Days before today.</b> Filters files that are modified within the specified number of days until the current date (today). Enter the number of days. The current date calculation starts from 00:00 hours.</li> </ul> For example, if you schedule the file ingestion task to run weekly and want to filter for the files that were modified in the previous week, set <b>Days before today</b> to 7. The task will pick up any file with a date between 7 days ago and the date on which it runs.
Time Zone	This applies when <b>File Pickup</b> is <b>By Pattern</b> . If you selected a <b>File Date</b> option, enter the time zone of the location where the files are located.
File Size	This applies when <b>File Pickup</b> is <b>By Pattern</b> . Filters the files to transfer based on file size. Enter the file size, select the file size unit, and filter options. Select one of the following filter options: <ul style="list-style-type: none"> <li>- <b>Greater than or Equal.</b> Filters files that are greater than or equal to the specified size.</li> <li>- <b>Less than or Equal.</b> Filters files that are less than or equal to the specified size.</li> <li>- <b>Equal.</b> Filters files that have the specified size.</li> </ul>
The file path containing the list of files	This applies when <b>File Pickup</b> is <b>By File List</b> . This applies when <b>File Pickup</b> is <b>By File List</b> . Select this option to provide the path that contains the list of files to pick up and enter the file path.



Option	Description
File list	This applies when <b>File Pickup</b> is <b>By File List</b> . This applies when <b>File Pickup</b> is <b>By File List</b> . Select this option to provide the list of files to pick up and enter a comma-separated list of file names.
Skip Duplicate Files	Indicates whether to skip duplicate files. If you select this option, the file ingestion task does not transfer files that have the same name and creation date as another file. The file ingestion task marks these files as duplicate in the job log. If you do not select this option, the task transfers all files, even files with duplicate names and creation dates.
Check file stability	Indicates whether to verify that a file is stable before a file ingestion task attempts to pick it. The task skips unstable files it detects in the current run.
Stability check interval	<p>This applies when you enable the <b>Check file stability</b> option. Time in seconds that a file ingestion task waits to check the file stability.</p> <p>For example, if the stability time is 15 seconds, the file ingestion task detects all the files in the source folder that match the defined file pattern, it waits for 15 seconds, and then it processes only the stable files.</p> <p>The interval ranges between 10 seconds to 300 seconds. Default is 10 seconds.</p>
Batch Size	<p>The number of files a file ingestion task can transfer in a batch.</p> <p>Default is 5.</p>
Transfer Mode	<p>File transfer mode. Select one of the following filter modes:</p> <ul style="list-style-type: none"> <li>- <b>Auto</b>. Mass Ingestion Files determines the transfer mode.</li> <li>- <b>ASCII</b>.</li> <li>- <b>Binary</b>.</li> </ul>
After File Pickup	<p>Determines what to do with the source files after the files are transferred.</p> <p>Select one of the following filter options:</p> <ul style="list-style-type: none"> <li>- Keep the files in the source directory.</li> <li>- Delete the files from the source directory.</li> <li>- Rename the files in the source directory. You must specify a file name suffix that file ingestion task adds to the file name when renaming the files.</li> <li>- Archive the files to a different location. You must specify an archive directory which is the absolute path or relative path from the source file system.</li> </ul>

## Advanced SFTP V2 source properties

When you define a file ingestion task with an Advanced SFTP V2 source, you must enter source options on the **Source** tab of the task wizard. The options vary based on the file pickup method that you select for the task.

**Note:** You can overwrite the file name pattern, folder, and table parameters, and define your own variable for sources by using the job resource of the Mass Ingestion Files REST API. For more information, see [Mass Ingestion Files REST API](#).

The following table describes the source options:

Option	Description
File Pickup	The file ingestion task supports the following file pickup methods: <ul style="list-style-type: none"> <li>- <b>By Pattern.</b> The file ingestion task picks up files by pattern.</li> <li>- <b>By File List.</b> The file ingestion task picks up files based on a file list.</li> </ul>
Source Directory	Directory from where files are transferred. The default value is the source directory specified in the connection. You can enter a relative path to the source file system. To enter a relative path, start the path with a period, followed by a slash (. /). The path is relative to the source directory specified in the connection.
Add Parameters	Create an expression to add it as a <b>Source Directory</b> . For more information, see <a href="#">“Source and target parameters” on page 420</a> .
Include files from sub-folders	This applies when <b>File Pickup</b> is <b>By Pattern</b> . Transfer files from all subfolders under the defined source directory.
File Pattern	This applies when <b>File Pickup</b> is <b>By Pattern</b> . File name pattern to use for selecting the files to transfer. The pattern can be a regular expression or a pattern with wildcard characters. The following wildcard characters are allowed: <ul style="list-style-type: none"> <li>- An asterisk (*) to represent any number of characters.</li> <li>- A question mark (?) to represent a single character.</li> </ul> For example, you can specify the following regular expression: <code>([a-zA-Z0-9\s_\.\-\(\)]+)(.doc .docx .pdf)\$</code>
File Date	This applies when <b>File Pickup</b> is <b>By Pattern</b> . A date and time expression for filtering the files to transfer. Select one of the following options: <ul style="list-style-type: none"> <li>- <b>Greater than or Equal.</b> Filters files that are modified on or after the specified date and time. To specify a date, click the calendar. To specify a time, click the clock.</li> <li>- <b>Less than or Equal.</b> Filters files that are modified before or on the specified date and time.</li> <li>- <b>Equal.</b> Filters files that are modified on the specified date and time. Click the calendar to select the date and the clock to select the time.</li> <li>- <b>Days before today.</b> Filters files that are modified within the specified number of days until the current date (today). Enter the number of days. The current date calculation starts from 00:00 hours.</li> </ul> For example, if you schedule the file ingestion task to run weekly and want to filter for the files that were modified in the previous week, set <b>Days before today</b> to 7. The task will pick up any file with a date between 7 days ago and the date on which it runs.
Time Zone	This applies when <b>File Pickup</b> is <b>By Pattern</b> . If you selected a <b>File Date</b> option, enter the time zone of the location where the files are located.
File Size	This applies when <b>File Pickup</b> is <b>By Pattern</b> . Filters the files to transfer based on file size. Enter the file size, select the file size unit, and filter options. Select one of the following filter options: <ul style="list-style-type: none"> <li>- <b>Greater than or Equal.</b> Filters files that are greater than or equal to the specified size.</li> <li>- <b>Less than or Equal.</b> Filters files that are less than or equal to the specified size.</li> <li>- <b>Equal.</b> Filters files that have the specified size</li> </ul>
The file path containing the list of files	This applies when <b>File Pickup</b> is <b>By File List</b> . Select this option to provide the path that contains the list of files to pick up and enter the file path.

Option	Description
File list	This applies when <b>File Pickup</b> is <b>By File List</b> . Select this option to provide the list of files to pick up and enter a comma-separated list of file names.
Skip Duplicate Files	Indicates whether to skip duplicate files. If you select this option, the file ingestion task does not transfer files that have the same name and creation date as another file. The file ingestion task marks these files as duplicate in the job log. If you do not select this option, the task transfers all files, even files with duplicate names and creation dates.
Check file stability	Indicates whether to verify that a file is stable before a file ingestion task attempts to pick it. The task skips unstable files it detects in the current run.
Stability check interval	<p>This applies when you enable the <b>Check file stability</b> option. Time in seconds that a file ingestion task waits to check the file stability.</p> <p>For example, if the stability time is 15 seconds, the file ingestion task detects all the files in the source folder that match the defined file pattern, it waits for 15 seconds, and then it processes only the stable files.</p> <p>The interval ranges between 10 seconds to 300 seconds. Default is 10 seconds.</p>
Batch Size	<p>The number of files a file ingestion task can transfer in a batch.</p> <p>Default is 5.</p>
After File Pickup	<p>Determines what to do with the source files after the files are transferred.</p> <p>Select one of the following filter options:</p> <ul style="list-style-type: none"> <li>- Keep the files in the source directory.</li> <li>- Delete the files from the source directory.</li> <li>- Rename the files in the source directory. You must specify a file name suffix that file ingestion task adds to the file name when renaming the files.</li> <li>- Archive the files to a different location. You must specify an archive directory which is the absolute path or relative path from the source file system.</li> </ul>

## Amazon S3 V2 source properties

When you define a file ingestion task with an Amazon S3 V2 source, you must enter source options on the **Source** tab of the task wizard. The options vary based on the file pickup method that you select for the task.

**Note:** You can overwrite the file name pattern, folder, and table parameters, and define your own variable for sources by using the job resource of the Mass Ingestion Files REST API. For more information, see [Mass Ingestion Files REST API](#).

The following table describes the source options:

Option	Description
File Pickup	<p>The file ingestion task supports the following file pickup methods:</p> <ul style="list-style-type: none"> <li>- <b>By Pattern</b>. The file ingestion task picks up files by pattern.</li> <li>- <b>By File List</b>. The file ingestion task picks up files based on a file list.</li> </ul>
Source Directory	<p>Amazon S3 folder path from where files are transferred, including bucket name. The default value is the Folder Path value specified in the connection properties.</p> <p>You can enter a relative path to the source file system. To enter a relative path, start the path with a period, followed by a slash (. /). The path is relative to the source directory specified in the connection.</p>

Option	Description
Add Parameters	Create an expression to add it as a <b>Folder Path</b> parameter. For more information, see <a href="#">“Source and target parameters” on page 420</a> .
Include files from sub-folders	This applies when <b>File Pickup</b> is <b>By Pattern</b> . Transfer files from all subfolders under the defined source directory.
File Pattern	This applies when <b>File Pickup</b> is <b>By Pattern</b> . File name pattern used to select the files to transfer. In the pattern, you can use the following wildcard characters: <ul style="list-style-type: none"> <li>- An asterisk (*) to represent any number of characters.</li> <li>- A question mark (?) to represent a single character.</li> </ul>
File Date	This applies when <b>File Pickup</b> is <b>By Pattern</b> . A date and time expression for filtering the files to transfer. <p>Select one of the following options:</p> <ul style="list-style-type: none"> <li>- <b>Greater than or Equal</b>. Filters files that are modified on or after the specified date and time. To specify a date, click the calendar. To specify a time, click the clock.</li> <li>- <b>Less than or Equal</b>. Filters files that are modified before or on the specified date and time.</li> <li>- <b>Equal</b>. Filters files that are modified on the specified date and time. Click the calendar to select the date and the clock to select the time.</li> <li>- <b>Days before today</b>. Filters files that are modified within the specified number of days until the current date. Enter the number of days. The current date calculation starts from 00:00 hours. For example, if you schedule the file ingestion task to run weekly and want to filter for the files that were modified in the previous week, set <b>Days before today</b> to 7. The task will pick up any file with a date between 7 days ago and the date on which it runs.</li> </ul>
Time Zone	This applies when <b>File Pickup</b> is <b>By Pattern</b> . If you selected a <b>File Date</b> option, enter the time zone of the location where the files are located.
File Size	This applies when <b>File Pickup</b> is <b>By Pattern</b> . Filters the files to transfer based on file size. Enter the file size, select the file size unit, and filter options. <p>Select one of the following filter options:</p> <ul style="list-style-type: none"> <li>- <b>Greater than or Equal</b>. Filters files that are greater than or equal to the specified size.</li> <li>- <b>Less than or Equal</b>. Filters files that are less than or equal to the specified size.</li> <li>- <b>Equal</b>. Filters files that have the specified size.</li> </ul>
The file path containing the list of files	This applies when <b>File Pickup</b> is <b>By File List</b> . Select this option to provide the path that contains the list of files to pick up and enter the file path.
File list	This applies when <b>File Pickup</b> is <b>By File List</b> . Select this option to provide the list of files to pick up and enter a comma-separated list of file names.
Skip Duplicate Files	Indicates whether to skip duplicate files. If you select this option, the file ingestion task does not transfer files that have the same name and creation date as another file. The file ingestion task marks these files as duplicate in the job log. If you do not select this option, the task transfers all files, even files with duplicate names and creation dates.
Check file stability	Indicates whether to verify that a file is stable before a file ingestion task attempts to pick it. The task skips unstable files it detects in the current run.

Option	Description
Stability check interval	<p>This applies when you enable the <b>Check file stability</b> option. Time in seconds that a file ingestion task waits to check the file stability.</p> <p>For example, if the stability time is 15 seconds, the file ingestion task detects all the files in the source folder that match the defined file pattern, it waits for 15 seconds, and then it processes only the stable files.</p> <p>The interval ranges between 10 seconds to 300 seconds. Default is 10 seconds.</p>
Batch Size	<p>The number of files a file ingestion task can transfer in a batch.</p> <p>Default is 5.</p> <p>The maximum value of the batch depends on whether the files transfer through an intermediate staging server.</p> <p>A file ingestion task does not transfer files through an intermediate staging server if the files are transferred from the following source to target endpoints:</p> <ul style="list-style-type: none"> <li>- Amazon S3 to Amazon Redshift, if you choose to transfer files without using intermediate staging.</li> <li>- Amazon S3 to Snowflake</li> </ul> <p><b>Note:</b> When you transfer files using a command line, the file ingestion task transfers files through an intermediate staging server.</p> <p>Consider the following guidelines when you define a batch size:</p> <ul style="list-style-type: none"> <li>- If files are transferred from the source to target without an intermediate staging server, the maximum number of files you can transfer in a batch is 8000.</li> <li>- If files pass through an intermediate staging server, the maximum number of files you can transfer in a batch is 20.</li> <li>- If you transfer files from any source to a Snowflake target, the maximum number of files you can transfer in a batch is 1000.</li> </ul>
File Encryption Type	<p>Type of Amazon S3 file encryption to use during file transfer.</p> <p>Select one of the following options:</p> <ul style="list-style-type: none"> <li>- <b>None.</b> Files are not encrypted during file transfer. Default is <b>None</b>.</li> <li>- <b>S3 server-side encryption.</b> Amazon S3 encrypts the file by using AWS-managed encryption keys.</li> <li>- <b>S3 client-side encryption.</b> Ensure that unrestricted policies are implemented for the AgentJVM, and that the master symmetric key for the connection is set.</li> </ul>
S3 Accelerated Transfer	<p>Select whether to use Amazon S3 Transfer Acceleration on the S3 bucket.</p> <p>To use Transfer Acceleration, accelerated transfer must be enabled for the bucket. The following options are available:</p> <ul style="list-style-type: none"> <li>- <b>Disabled.</b> Do not use Amazon S3 Transfer Acceleration.</li> <li>- <b>Accelerated.</b> Use Amazon S3 Transfer Acceleration.</li> <li>- <b>Dualstack Accelerated.</b> Use Amazon S3 Transfer Acceleration on a dual-stack endpoint.</li> </ul>
Minimum Download Part Size	<p>Minimum download part size in megabytes when downloading a large file as a set of multiple independent parts.</p>

Option	Description
Multipart Download Threshold	Multipart download minimum threshold in megabytes that is used to determine when to upload objects in multiple parts in parallel.
After File Pickup	<p>Determines what to do with the source files after the task streams them to the target.</p> <p>Select one of the following options:</p> <ul style="list-style-type: none"> <li>- Keep the files in the source directory.</li> <li>- Delete the files from the source directory.</li> <li>- Rename the files in the source directory. You must specify a file name suffix that file ingestion task adds to the file name when renaming the files. You can choose to suffix the new file name with a timestamp (\$timestamp), date (\$date), runID (\$runId), or time (\$time).</li> <li>- Archive the files to a different location. You must specify an archive directory which is the absolute path or relative path to the source file system.</li> </ul>

## Databricks Delta source properties

In a file ingestion task, you can configure Databricks Delta source properties to transfer tables from a Databricks Delta source to a Microsoft Azure Data Lake Store Gen2 target and an Amazon S3 V2 target. The tables from the Databricks Delta source are stored as Parquet files in the target.

**Note:** You can overwrite the file name pattern, folder, and table parameters, and define your own variable for sources by using the job resource of the Mass Ingestion Files REST API. For more information, see [Mass Ingestion Files REST API](#).

The following table describes the source options:

Option	Description
Database	Required. Name of the Databricks Delta Lake database that contains the source table(s).
Add Parameters	Create an expression to add it as <b>Database</b> and <b>Table Pattern</b> parameters. For more information, see <a href="#">"Source and target parameters" on page 420</a> .
Table Pattern Type	<p>Required. Type of pattern that determines how you select the tables for the transfer. Select <b>Wildcard</b> or <b>Regex</b>.</p> <p>Default is Wildcard.</p>
Table Pattern	<p>Required. Enter the table name pattern for the pattern type you specified:</p> <ul style="list-style-type: none"> <li>- For a wildcard pattern, enter a pattern with the following characters: <ul style="list-style-type: none"> <li>- An asterisk (*) to represent any number of characters.</li> <li>- A question mark (?) to represent a single character.</li> </ul> </li> <li>- For a Regex pattern, enter a regular expression.</li> </ul>
Batch Size	<p>Required. The maximum number of tables that a file ingestion task can transfer from a Databricks Delta source to a Microsoft Azure Data Lake Store Gen2 target in a batch.</p> <p>Default is 5. The maximum number of tables the task can transfer in a batch is 1000.</p> <p><b>Note:</b> The task transfers tables with no intermediate staging.</p>

**Note:** If a job fails with the following error, see the cluster logs for more information:

```
"[ERROR] Job execution failed. State : JOB_FAILED ; State Message :"
```

## File listener source properties

Configure a file listener as a source type when you use the file listener to trigger the file ingestion task.

To configure a file listener as a source, you must create a file listener by using the Components service. For more information about creating a file listener, see the *Components* help.

**Note:** You cannot run the file ingestion task with a file listener as a source from the file ingestion user interface. A file ingestion task with a file listener as a source runs automatically when the file listener starts.

The following table describes the source options:

Option	Description
File Pattern	<p>File name pattern to use for selecting the files to transfer. The pattern can be a regular expression or a pattern with wildcard characters.</p> <p>The following wildcard characters are allowed:</p> <ul style="list-style-type: none"><li>- An asterisk (*) to represent any number of characters.</li><li>- A question mark (?) to represent a single character.</li></ul> <p>For example, you can specify the following regular expression:</p> <pre>([a-zA-Z0-9\s_\.\-\(\)]+)(.doc .docx .pdf)\$</pre>
File Date	<p>A date and time expression for filtering the files to transfer.</p> <p>Select one of the following options:</p> <ul style="list-style-type: none"><li>- <b>Greater than or Equal.</b> Filters files that are modified on or after the specified date and time. To specify a date, click the calendar. To specify a time, click the clock.</li><li>- <b>Less than or Equal.</b> Filters files that are modified before or on the specified date and time.</li><li>- <b>Equal.</b> Filters files that are modified on the specified date and time. Click the calendar to select the date and the clock to select the time.</li><li>- <b>Days before today.</b> Filters files that are modified within the specified number of days until the current date (today). Enter the number of days. The current date calculation starts from 00:00 hours.</li></ul> <p>For example, if you schedule the file ingestion task to run weekly and want to filter for the files that were modified in the previous week, set <b>Days before today</b> to 7. The task will pick up any file with a date between 7 days ago and the date on which it runs.</p>
Time Zone	<p>If you selected a <b>File Date</b> option, enter the time zone of the location where the files are located.</p>
File Size	<p>Filters the files to transfer based on file size. Enter the file size, select the file size unit, and filter options.</p> <p>Select one of the following filter options:</p> <ul style="list-style-type: none"><li>- <b>Greater than or Equal.</b> Filters files that are greater than or equal to the specified size.</li><li>- <b>Less than or Equal.</b> Filters files that are less than or equal to the specified size.</li><li>- <b>Equal.</b> Filters files that have the specified size.</li></ul>
Skip Duplicate Files	<p>Indicates whether to skip duplicate files. If you select this option, the file ingestion task does not transfer files that have the same name and creation date as another file. The file ingestion task marks these files as duplicate in the job log. If you do not select this option, the task transfers all files, even files with duplicate names and creation dates.</p>

Option	Description
Batch Size	The number of files a file ingestion task can transfer in a batch. Default is 5.
After File Pickup	Determines what to do with the source files after the files are transferred. Select one of the following filter options: <ul style="list-style-type: none"> <li>- Keep the files in the source directory.</li> <li>- Delete the files from the source directory.</li> <li>- Rename the files in the source directory. You must specify a file name suffix that file ingestion task adds to the file name when renaming the files.</li> <li>- Archive the files to a different location. You must specify an archive directory.</li> </ul>

## Google Cloud Storage V2 source properties

When you define a file ingestion task with a Google Cloud Storage V2 source, you must enter source options on the **Source** tab of the task wizard.

**Note:** You can overwrite the file name pattern, folder, and table parameters, and define your own variable for sources by using the job resource of the Mass Ingestion Files REST API. For more information, see [Mass Ingestion Files REST API](#).

The following table describes the source options:

Option	Description
File Pickup	The file ingestion task supports the following file pickup methods: <ul style="list-style-type: none"> <li>- <b>By Pattern.</b> The file ingestion task picks up files by pattern.</li> <li>- <b>By File List.</b> The file ingestion task picks up files based on a file list.</li> </ul>
Source Directory	Directory from where files are transferred. You can enter a relative path to the source file system. To enter a relative path, start the path with a period, followed by a slash (.). The path is relative to the source directory specified in the connection.
Add Parameters	Create an expression to add it as a <b>Source Directory</b> parameter. For more information, see <a href="#">"Source and target parameters" on page 420</a> .
Include files from sub-folders	Transfer files from all subfolders under the defined source directory.
File Pattern	File name pattern to use for selecting the files to transfer. The pattern can be a regular expression or a pattern with wildcard characters. The following wildcard characters are allowed: <ul style="list-style-type: none"> <li>- An asterisk (*) to represent any number of characters.</li> <li>- A question mark (?) to represent a single character.</li> </ul> For example, you can specify the following regular expression: <code>([a-zA-Z0-9\s_\.\-\(\)]+)(.doc .docx .pdf)\$</code>



Option	Description
File Date	<p>A date and time expression for filtering the files to transfer.</p> <p>Select one of the following options:</p> <ul style="list-style-type: none"> <li>- <b>Greater than or Equal.</b> Filters files that are modified on or after the specified date and time. To specify a date, click the calendar. To specify a time, click the clock.</li> <li>- <b>Less than or Equal.</b> Filters files that are modified before or on the specified date and time.</li> <li>- <b>Equal.</b> Filters files that are modified on the specified date and time. Click the calendar to select the date and the clock to select the time.</li> <li>- <b>Days before today.</b> Filters files that are modified within the specified number of days until the current date (today). Enter the number of days. The current date calculation starts from 00:00 hours.</li> </ul> <p>For example, if you schedule the file ingestion task to run weekly and want to filter for the files that were modified in the previous week, set <b>Days before today</b> to 7. The task will pick up any file with a date between 7 days ago and the date on which it runs.</p>
Time Zone	If you selected a <b>File Date</b> option, enter the time zone of the location where the files are located.
File Size	<p>Filters the files to transfer based on file size. Enter the file size, select the file size unit, and filter options.</p> <p>Select one of the following filter options:</p> <ul style="list-style-type: none"> <li>- <b>Greater than or Equal.</b> Filters files that are greater than or equal to the specified size.</li> <li>- <b>Less than or Equal.</b> Filters files that are less than or equal to the specified size.</li> <li>- <b>Equal.</b> Filters files that have the specified size.</li> </ul>
Skip Duplicate Files	Indicates whether to skip duplicate files. If you select this option, the file ingestion task does not transfer files that have the same name and creation date as another file. The file ingestion task marks these files as duplicate in the job log. If you do not select this option, the task transfers all files, even files with duplicate names and creation dates.
Check file stability	Indicates whether to verify that a file is stable before a file ingestion task attempts to pick it. The task skips unstable files it detects in the current run.
Stability check interval	<p>This applies when you enable the <b>Check file stability</b> option. Time in seconds that a file ingestion task waits to check the file stability.</p> <p>For example, if the stability time is 15 seconds, the file ingestion task detects all the files in the source folder that match the defined file pattern, waits for 15 seconds, and processes only the stable files.</p> <p>The interval ranges between 10 seconds to 300 seconds. Default is 10 seconds.</p>
Batch Size	<p>The number of files a file ingestion task can transfer in a batch.</p> <p>Default is 5.</p> <p>The maximum batch size varies, based on the following conditions:</p> <ul style="list-style-type: none"> <li>- If files are transferred from the source to target without an intermediate staging server, the maximum number of files you can transfer in a batch is 8000.</li> <li>- If files pass through an intermediate staging server, the maximum number of files you can transfer in a batch is 20.</li> <li>- If you transfer files from any source to a Snowflake target, the maximum number of files you can transfer in a batch is 1000.</li> </ul> <p><b>Note:</b> If you transfer files from Google Cloud Storage to Google BigQuery, the task transfers files with no intermediate staging server.</p>

## Hadoop Files V2 source properties

When you define a file ingestion task with an Hadoop Files V2 source, you must enter source options on the **Source** tab of the task wizard. The options vary based on the file pickup method that you select for the task.

**Note:** You can overwrite the file name pattern, folder, and table parameters, and define your own variable for sources by using the job resource of the Mass Ingestion Files REST API. For more information, see [Mass Ingestion Files REST API](#).

The following table describes the source options:

Option	Description
File Pickup	The file ingestion task supports the following file pickup methods: <ul style="list-style-type: none"><li>- <b>By Pattern.</b> The file ingestion task picks up files by pattern.</li><li>- <b>By File List.</b> The file ingestion task picks up files based on a file list.</li></ul>
Source Directory	Directory from where files are transferred.
Add Parameters	Create an expression to add it as a <b>Source Directory</b> parameter. For more information, see <a href="#">“Source and target parameters” on page 420</a> .
Include files from sub folders	This applies when <b>File Pickup</b> is <b>By Pattern</b> . Transfer files from all subfolders under the defined source directory.
File Pattern	<p>This applies when <b>File Pickup</b> is <b>By Pattern</b>. File name pattern used to select the files to transfer. Based on the file pattern that you have selected, enter the file name patterns.</p> <p>Select one of the following file patterns:</p> <ul style="list-style-type: none"><li>- <b>Wildcard.</b> Use the following wildcard character filters:<ul style="list-style-type: none"><li>- An asterisk (*) to represent any number of characters.</li><li>- A question mark (?) to represent a single character.</li></ul></li><li>- <b>Regex.</b> Use regular expression to match the pattern type. Consider the following samples:<ul style="list-style-type: none"><li>- <code>^(?!.*(?:out baz foo)).*\$</code> all except Identifies all files except for files whose name contains out, foo, and baz.</li><li>- <code>([a-zA-Z0-9\s_\\.\-\\(\):])+(.doc .docx .pdf)\$</code> Identifies all files that have an extension of doc, docx, or pdf.</li><li>- <code>^(?!out).*\.txt\$</code> Identifies all text files except for files whose name contains out.txt.</li></ul></li></ul>

Option	Description
File Date	<p>This applies when <b>File Pickup</b> is <b>By Pattern</b>. A date and time expression for filtering the files to transfer.</p> <p>Select one of the following options:</p> <ul style="list-style-type: none"> <li>- <b>Greater than or Equal</b>. Filters files that are modified on or after the specified date and time. To specify a date, click the calendar. To specify a time, click the clock.</li> <li>- <b>Less than or Equal</b>. Filters files that are modified before or on the specified date and time.</li> <li>- <b>Equal</b>. Filters files that are modified on the specified date and time. Click the calendar to select the date and the clock to select the time.</li> <li>- <b>Days before today</b>. Filters files that are modified within the specified number of days until the current date (today). Enter the number of days. The current date calculation starts from 00:00 hours.</li> </ul> <p>For example, if you schedule the file ingestion task to run weekly and want to filter for the files that were modified in the previous week, set <b>Days before today</b> to 7. The task will pick up any file with a date between 7 days ago and the date on which it runs.</p>
Time Zone	<p>This applies when <b>File Pickup</b> is <b>By Pattern</b>. If you selected a <b>File Date</b> option, enter the time zone of the location where the files are located.</p>
File Size	<p>This applies when <b>File Pickup</b> is <b>By Pattern</b>. Filters the files to transfer based on file size. Enter the file size, select the file size unit, and filter options.</p> <p>Select one of the following filter options:</p> <ul style="list-style-type: none"> <li>- <b>Greater than or Equal</b>. Filters files that are greater than or equal to the specified size.</li> <li>- <b>Less than or Equal</b>. Filters files that are less than or equal to the specified size.</li> <li>- <b>Equal</b>. Filters files that have the specified size.</li> </ul>
The file path containing the list of files	<p>This applies when <b>File Pickup</b> is <b>By File List</b>. Select this option to provide the path that contains the list of files to pick up and enter the file path.</p>
File list	<p>This applies when <b>File Pickup</b> is <b>By File List</b>. Select this option to provide the list of files to pick up and enter a comma-separated list of file names.</p>
Skip Duplicate Files	<p>Indicates whether to skip duplicate files. If you select this option, the file ingestion task does not transfer files that have the same name and creation date as another file. The file ingestion task marks these files as duplicate in the job log. If you do not select this option, the task transfers all files, even files with duplicate names and creation dates.</p>
Check file stability	<p>Indicates whether to verify that a file is stable before a file ingestion task attempts to pick it. The task skips unstable files it detects in the current run.</p>
Stability check interval	<p>This applies when you enable the <b>Check file stability</b> option. Time in seconds that a file ingestion task waits to check the file stability.</p> <p>For example, if the stability time is 15 seconds, the file ingestion task detects all the files in the source folder that match the defined file pattern, waits for 15 seconds, and processes only the stable files.</p> <p>The interval ranges between 10 seconds to 300 seconds. Default is 10 seconds.</p>
Batch Size	<p>The number of files a file ingestion task can transfer in a batch.</p> <p>Default is 5.</p>

## Local folder source properties

When you define a file ingestion task with an local folder source, you must enter source properties on the **Source** tab of the task wizard. The options vary based on the file pickup method that you select for the task.

**Note:** You can overwrite the file name pattern, folder, and table parameters, and define your own variable for sources by using the job resource of the Mass Ingestion Files REST API. For more information, see [Mass Ingestion Files REST API](#).

The following table describes the source options:

Option	Description
File Pickup	The file ingestion task supports the following file pickup methods: <ul style="list-style-type: none"><li>- <b>By Pattern</b>. The file ingestion task picks up files by pattern.</li><li>- <b>By File List</b>. The file ingestion task picks up files based on a file list.</li></ul>
Source Directory	Directory from where files are transferred. The Secure Agent must be able to access the directory.  The use of slashes around the source folder path differs between connectors. Using slashes incorrectly will result in connection failures. For more information, see the Knowledge Base article <a href="#">625869</a> . <b>Note:</b> File listener can access files and directories on network shares with support for NFS and CIFS.
Add Parameters	Create an expression to add it as a <b>Source Directory</b> parameter. For more information, see <a href="#">"Source and target parameters" on page 420</a> .
Include files from sub-folders	This applies when <b>File Pickup</b> is <b>By Pattern</b> . Transfer files from all subfolders under the defined source directory.
File Pattern	This applies when <b>File Pickup</b> is <b>By Pattern</b> . File name pattern used to select the files to transfer. Based on the file pattern that you have selected, enter the file name patterns.  The following file patterns are available: <ul style="list-style-type: none"><li>- Wildcard. Use the following wildcard character filters:<ul style="list-style-type: none"><li>- An asterisk (*) matches any number of characters.</li><li>- A question mark (?) matches a single character.</li></ul></li><li>- Regex. Use regular expression to match the pattern type. Consider the following samples:<ul style="list-style-type: none"><li>- <code>^(?!.*(?:out baz foo)).*\$</code> all except Identifies all files except for files whose name contains out, foo, and baz.</li><li>- <code>([a-zA-Z0-9\s_\\.\-\\(\):])+(\.doc \.docx \.pdf)\$</code> Identifies all files that have an extension of doc, docx, or pdf.</li><li>- <code>^(?!out).*\.txt\$</code> Identifies all text files except for files whose name contains out.txt.</li></ul></li></ul>

Option	Description
File Date	<p>This applies when <b>File Pickup</b> is <b>By Pattern</b>. A date and time expression for filtering the files to transfer.</p> <p>Select one of the following options:</p> <ul style="list-style-type: none"> <li>- <b>Greater than or Equal</b>. Filters files that are modified on or after the specified date and time. To specify a date, click the calendar. To specify a time, click the clock.</li> <li>- <b>Less than or Equal</b>. Filters files that are modified before or on the specified date and time.</li> <li>- <b>Equal</b>. Filters files that are modified on the specified date and time. Click the calendar to select the date and the clock to select the time.</li> <li>- <b>Days before today</b>. Filters files that are modified within the specified number of days until the current date (today). Enter the number of days. The current date calculation starts from 00:00 hours.</li> </ul> <p>For example, if you schedule the file ingestion task to run weekly and want to filter for the files that were modified in the previous week, set <b>Days before today</b> to 7. The task will pick up any file with a date between 7 days ago and the date on which it runs.</p>
Time Zone	<p>This applies when <b>File Pickup</b> is <b>By Pattern</b>. If you selected a <b>File Date</b> option, enter the time zone of the location where the files are located.</p>
File Size	<p>This applies when <b>File Pickup</b> is <b>By Pattern</b>. Filters the files to transfer based on file size. Enter the file size, select the file size unit, and filter options.</p> <p>Select one of the following filter options:</p> <ul style="list-style-type: none"> <li>- <b>Greater than or Equal</b>. Filters files that are greater than or equal to the specified size.</li> <li>- <b>Less than or Equal</b>. Filters files that are less than or equal to the specified size.</li> <li>- <b>Equal</b>. Filters files that have the specified size.</li> </ul>
The file path containing the list of files	<p>This applies when <b>File Pickup</b> is <b>By File List</b>. Select this option to provide the path that contains the list of files to pick up and enter the file path.</p>
File list	<p>This applies when <b>File Pickup</b> is <b>By File List</b>. Select this option to provide the list of files to pick up and enter a comma-separated list of file names.</p>
Skip Duplicate Files	<p>Indicates whether to skip duplicate files. If you select this option, the file ingestion task does not transfer files that have the same name and creation date as another file. The file ingestion task marks these files as duplicate in the job log. If you do not select this option, the task transfers all files, even files with duplicate names and creation dates.</p>
Check file stability	<p>Indicates whether to verify that a file is stable before a file ingestion task attempts to pick it. The task skips unstable files it detects in the current run.</p>
Stability check interval	<p>This applies when you enable the <b>Check file stability</b> option. Time in seconds that a file ingestion task waits to check the file stability.</p> <p>For example, if the stability time is 15 seconds, the file ingestion task detects all the files in the source folder that match the defined file pattern, waits for 15 seconds, and processes only the stable files.</p> <p>The interval ranges between 10 seconds to 300 seconds. Default is 10 seconds.</p>

Option	Description
Batch Size	<p>The maximum number of files a file ingestion task transfers in a batch. Default is 5.</p> <p>The maximum batch size varies, based on the following conditions:</p> <ul style="list-style-type: none"> <li>- If the task transfers files from source to target with no intermediate staging, the maximum number of files the task can transfer in a batch is 8000.</li> <li>- If the task transfers files from source to target with intermediate staging, the maximum number of files the task can transfer in a batch is 20.</li> <li>- If the task transfers files from any source to a Snowflake target, the maximum number of files the task can transfer in a batch is 1000.</li> </ul> <p>Consider the following guidelines when you define the batch size:</p> <ul style="list-style-type: none"> <li>- The task transfers files with no intermediate staging in the following scenarios: <ul style="list-style-type: none"> <li>- File transfers from Amazon S3 to Amazon Redshift when Amazon Redshift Connector is configured to upload files with no intermediate staging</li> <li>- File transfers from Google Cloud Storage to Google BigQuery</li> <li>- File transfers from Azure Blob to Microsoft Azure Data Warehouse</li> <li>- File transfers from Amazon S3 and from Azure Blob to Snowflake</li> </ul> </li> <li>- When you use a command line to transfer files, the task transfers files with intermediate staging.</li> </ul>
After File Pickup	<p>Determines what to do with source files after the files transfer.</p> <p>The following options are available:</p> <ul style="list-style-type: none"> <li>- Keep files in the source directory.</li> <li>- Delete files from the source directory.</li> <li>- Rename files in the source directory. You must specify a file name suffix that File ingestion adds to the file name when renaming the files.</li> <li>- Archive the files to a different location. You must specify an archive directory.</li> </ul>

## Microsoft Azure Blob Storage V3 source properties

When you define a file ingestion task with an Microsoft Azure Blob Storage source, you must enter source options on the **Source** tab of the task wizard. The options vary based on the file pickup method that you select for the task.

**Note:** You can overwrite the file name pattern, folder, and table parameters, and define your own variable for sources by using the job resource of the Mass Ingestion Files REST API. For more information, see [Mass Ingestion Files REST API](#).

The following table describes the source options:

Advance Source Property	Description
File Pickup	<p>The file ingestion task supports the following file pickup methods:</p> <ul style="list-style-type: none"> <li>- <b>By Pattern.</b> The file ingestion task picks up files by pattern.</li> <li>- <b>By File List.</b> The file ingestion task picks up files based on a file list.</li> </ul>
Source Directory	<p>Microsoft Azure Blob Storage directory from where files are transferred, including the container name. The default value is the container path specified in the connection.</p> <p>You can enter a relative path to the source file system. To enter a relative path, start the path with a period, followed by a slash (./). The path is relative to the source directory specified in the connection.</p>

Advance Source Property	Description
Add Parameters	Create an expression to add it as a <b>Folder Path</b> parameter. For more information, see <a href="#">"Source and target parameters" on page 420</a> .
Include files from sub-folders	This applies when <b>File Pickup</b> is <b>By Pattern</b> . Transfer files from sub-folders present in the folder path.
File Pattern	<p>This applies when <b>File Pickup</b> is <b>By Pattern</b>. File name pattern used to select the files to transfer. You can use a regular expression or wildcard characters.</p> <p>The following wildcard characters are allowed:</p> <ul style="list-style-type: none"> <li>- An asterisk (*) to represent any number of characters.</li> <li>- A question mark (?) to represent a single character.</li> </ul> <p>For example, you can specify the following regular expression:</p> <pre>([a-zA-Z0-9\s_\\.\-\\(\)] : ) + (.doc   .docx   .pdf) \$</pre>
File Date	<p>This applies when <b>File Pickup</b> is <b>By Pattern</b>. A date and time expression for filtering the files to transfer.</p> <p>Select one of the following options:</p> <ul style="list-style-type: none"> <li>- <b>Greater than or Equal</b>. Filters files that are modified on or after the specified date and time. To specify a date, click the calendar. To specify a time, click the clock.</li> <li>- <b>Less than or Equal</b>. Filters files that are modified before or on the specified date and time.</li> <li>- <b>Equal</b>. Filters files that are modified on the specified date and time. Click the calendar to select the date and the clock to select the time.</li> <li>- <b>Days before today</b>. Filters files that are modified within the specified number of days until the current date (today). Enter the number of days. The current date calculation starts from 00:00 hours.</li> </ul> <p>For example, if you schedule the file ingestion task to run weekly and want to filter for the files that were modified in the previous week, set <b>Days before today</b> to 7. The task will pick up any file with a date between 7 days ago and the date on which it runs.</p>
Time Zone	This applies when <b>File Pickup</b> is <b>By Pattern</b> . If you selected a <b>File Date</b> option, enter the time zone of the location where the files are located.
File Size	<p>This applies when <b>File Pickup</b> is <b>By Pattern</b>. Filters the files to transfer based on file size. Enter the file size, select the file size unit, and filter options.</p> <p>Select one of the following filter options:</p> <ul style="list-style-type: none"> <li>- <b>Greater than or Equal</b>. Filters files that are greater than or equal to the specified size.</li> <li>- <b>Less than or Equal</b>. Filters files that are less than or equal to the specified size.</li> <li>- <b>Equal</b>. Filters files that have the specified size.</li> </ul>
The file path containing the list of files	This applies when <b>File Pickup</b> is <b>By File List</b> . Select this option to provide the path that contains the list of files to pick up and enter the file path.
File list	This applies when <b>File Pickup</b> is <b>By File List</b> . Select this option to provide the list of files to pick up and enter a comma-separated list of file names.
Skip Duplicate Files	Do not transfer duplicate files. If files with the same name and creation date were transferred by the same file ingestion task, the task does not transfer them again, and the files are marked as duplicate in the job log. If this option is not selected the task transfers all files.

Advance Source Property	Description
Check file stability	Indicates whether to verify that a file is stable before a file ingestion task attempts to pick it. The task skips unstable files it detects in the current run.
Stability check interval	<p>This applies when you enable the <b>Check file stability</b> option. Time in seconds that a file ingestion task waits to check the file stability.</p> <p>For example, if the stability time is 15 seconds, the file ingestion task detects all the files in the source folder that match the defined file pattern, it waits for 15 seconds, and then it processes only the stable files.</p> <p>The interval ranges between 10 seconds to 300 seconds. Default is 10 seconds.</p>
Batch Size	<p>The number of files a file ingestion task can transfer in a batch.</p> <p>Default is 5.</p> <p>The maximum batch size varies, based on the following conditions:</p> <ul style="list-style-type: none"> <li>- If files are transferred from the source to target without an intermediate staging server, the maximum number of files the task can transfer in a batch is 8000.</li> <li>- If files pass through an intermediate staging server, the maximum number of files the task can transfer in a batch is 20.</li> <li>- If the task transfers files from any source to a Snowflake target, the maximum number of files the task can transfer in a batch is 1000.</li> </ul> <p><b>Note:</b> If you transfer files from Azure Blob Storage to Azure SQL Data Warehouse and Snowflake, the task transfers files with no intermediate staging.</p>

## Microsoft Azure Data Lake Storage Gen2 source properties

In a file ingestion task, you can configure the Microsoft Azure Data Lake Storage Gen2 source properties to transfer files from a Microsoft Azure Data Lake Storage Gen2 source to a Microsoft Azure Data Lake Storage Gen2 target or any target that a file ingestion task supports. The source options vary based on the file pickup method that you select for the task.

When the task transfers files from a Microsoft Azure Data Lake Storage Gen2 source to a Databricks Delta target, the files must be of Parquet format and must have the same schema as the Databricks Delta target.

**Note:** You can overwrite the file name pattern, folder, and table parameters, and define your own variable for sources by using the job resource of the Mass Ingestion Files REST API. For more information, see [Mass Ingestion Files REST API](#).

The following table describes the source options:

Advance Source Property	Description
File Pickup	<p>The file ingestion task supports the following file pickup methods:</p> <ul style="list-style-type: none"> <li>- <b>By Pattern.</b> The file ingestion task picks up files by pattern.</li> <li>- <b>By File List.</b> The file ingestion task picks up files based on a file list.</li> </ul>
Source Directory	<p>Microsoft Azure Data Lake Storage Gen2 folder path from where files are transferred. The default value is the container path specified in the connection.</p> <p>You can enter a relative path to the source file system. To enter a relative path, start the path with a period, followed by a slash (. /). The path is relative to the source directory specified in the connection.</p>



Advance Source Property	Description
Add Parameters	Create an expression to add it as a <b>Source Directory</b> parameter. For more information, see <a href="#">"Source and target parameters" on page 420</a> .
Include files from sub-folders	This applies when <b>File Pickup</b> is <b>By Pattern</b> . Transfer files from sub-folders present in the folder path.
File Pattern	<p>This applies when <b>File Pickup</b> is <b>By Pattern</b>. File name pattern used to select the files to transfer. You can use a regular expression or wildcard characters.</p> <p>The following wildcard characters are allowed:</p> <ul style="list-style-type: none"> <li>- An asterisk (*) to represent any number of characters.</li> <li>- A question mark (?) to represent a single character.</li> </ul> <p>For example, you can specify the following regular expression:</p> <pre>([a-zA-Z0-9\s_\.\-\(\)]+)(.doc .docx .pdf)\$</pre>
File Date	<p>This applies when <b>File Pickup</b> is <b>By Pattern</b>. A date and time expression for filtering the files to transfer.</p> <p>Select one of the following options:</p> <ul style="list-style-type: none"> <li>- <b>Greater than or Equal</b>. Filters files that are modified on or after the specified date and time. To specify a date, click the calendar. To specify a time, click the clock.</li> <li>- <b>Less than or Equal</b>. Filters files that are modified before or on the specified date and time.</li> <li>- <b>Equal</b>. Filters files that are modified on the specified date and time. Click the calendar to select the date and the clock to select the time.</li> <li>- <b>Days before today</b>. Filters files that are modified within the specified number of days until the current date (today). Enter the number of days. The current date calculation starts from 00:00 hours.</li> </ul> <p>For example, if you schedule the file ingestion task to run weekly and want to filter for the files that were modified in the previous week, set <b>Days before today</b> to 7. The task will pick up any file with a date between 7 days ago and the date on which it runs.</p>
Time Zone	This applies when <b>File Pickup</b> is <b>By Pattern</b> . If you selected a <b>File Date</b> option, enter the time zone of the location where the files are located.
File Size	<p>This applies when <b>File Pickup</b> is <b>By Pattern</b>. Filters the files to transfer based on file size. Enter the file size, select the file size unit, and filter options.</p> <p>Select one of the following filter options:</p> <ul style="list-style-type: none"> <li>- <b>Greater than or Equal</b>. Filters files that are greater than or equal to the specified size.</li> <li>- <b>Less than or Equal</b>. Filters files that are less than or equal to the specified size.</li> <li>- <b>Equal</b>. Filters files that have the specified size.</li> </ul>
The file path containing the list of files	This applies when <b>File Pickup</b> is <b>By File List</b> . Select this option to provide the path that contains the list of files to pick up and enter the file path.
File list	This applies when <b>File Pickup</b> is <b>By File List</b> . Select this option to provide the list of files to pick up and enter a comma-separated list of file names.
Skip duplicate files	Do not transfer duplicate files. If files with the same name and creation date were transferred by the same file ingestion task, the task does not transfer them again, and the files are marked as duplicate in the job log. If this option is not selected the task transfers all files.

Advance Source Property	Description
Check file stability	Indicates whether to verify that a file is stable before a file ingestion task attempts to pick it. The task skips unstable files it detects in the current run.
Stability check interval	<p>This applies when you enable the <b>Check file stability</b> option. Time in seconds that a file ingestion task waits to check the file stability.</p> <p>For example, if the stability time is 15 seconds, the file ingestion task detects all the files in the source folder that match the defined file pattern, it waits for 15 seconds, and then it processes only the stable files.</p> <p>The interval ranges between 10 seconds to 300 seconds. Default is 10 seconds.</p>
Batch Size	<p>The number of files a file ingestion task can transfer in a batch.</p> <p>Default is 5.</p> <p>The maximum batch size varies, based on the following conditions:</p> <ul style="list-style-type: none"> <li>- If the task transfers files from source to target with no intermediate staging, the maximum number of files the task can transfer in a batch is 8000.</li> <li>- If the task transfers files from source to target with intermediate staging, the maximum number of files the task can transfer in a batch is 20.</li> <li>- If the task transfers files from any source to a Snowflake or Databricks Delta target, the maximum number of files the task can transfer in a batch is 1000.</li> </ul> <p><b>Note:</b> If you transfer files from Microsoft Azure Data Lake Storage Gen2 to Azure SQL Data Warehouse, the task transfers files with no intermediate staging.</p>
Block Size (Bytes)	<p>Divides a large file into smaller specified block size. When you read a large file, divide the file into smaller parts and configure concurrent connections to spawn the required number of threads to process data in parallel.</p> <p>Default is 8388608 bytes (8 MB).</p>
After File Pickup	<p>Determines what to do with source files after the files transfer. The following options are available:</p> <ul style="list-style-type: none"> <li>- Keep files in the source directory.</li> <li>- Delete files from the source directory.</li> <li>- Rename files in the source directory. You must specify a file name suffix that File Ingestion adds to the file name when renaming the files.</li> <li>- Archive the files to a different location. You must specify an archive directory which is the absolute path or relative path from the source file system.</li> </ul>

## Microsoft Azure Data Lake Store Gen1 V3 source properties

When you define a file ingestion task with an Microsoft Azure Data Lake Store Gen1 V3 source, you must enter source options on the **Source** tab of the task wizard. The options vary based on the file pickup method that you select for the task.

**Note:** You can overwrite the file name pattern, folder, and table parameters, and define your own variable for sources by using the job resource of the Mass Ingestion Files REST API. For more information, see [Mass Ingestion Files REST API](#).

The following table describes the source options:

Advance Source Property	Description
File Pickup	The file ingestion task supports the following file pickup methods: <ul style="list-style-type: none"> <li>- <b>By Pattern.</b> The file ingestion task picks up files by pattern.</li> <li>- <b>By File List.</b> The file ingestion task picks up files based on a file list.</li> </ul>
Source Directory	Microsoft Azure Data Lake Store directory from where files are transferred. The default value is the container path specified in the connection.  You can enter a relative path to the source file system. To enter a relative path, start the path with a period, followed by a slash (./). The path is relative to the source directory specified in the connection.
Add Parameters	Create an expression to add it as a <b>Source Directory</b> parameter. For more information, see <a href="#">“Source and target parameters” on page 420</a> .
Include files from sub-folders	This applies when <b>File Pickup</b> is <b>By Pattern</b> . Transfer files from sub-folders present in the folder path.
File Pattern	This applies when <b>File Pickup</b> is <b>By Pattern</b> . File name pattern to use for selecting the files to transfer. The pattern can be a regular expression or a pattern with wildcard characters.  The following wildcard characters are allowed: <ul style="list-style-type: none"> <li>- An asterisk (*) to represent any number of characters.</li> <li>- A question mark (?) to represent a single character.</li> </ul> For example, you can specify the following regular expression: <code>([a-zA-Z0-9\s_\.\\-\\(\):])+(.doc .docx .pdf)\$</code>
File Date	This applies when <b>File Pickup</b> is <b>By Pattern</b> . A date and time expression for filtering the files to transfer.  Select one of the following options: <ul style="list-style-type: none"> <li>- <b>Greater than or Equal.</b> Filters files that are modified on or after the specified date and time. To specify a date, click the calendar. To specify a time, click the clock.</li> <li>- <b>Less than or Equal.</b> Filters files that are modified before or on the specified date and time.</li> <li>- <b>Equal.</b> Filters files that are modified on the specified date and time. Click the calendar to select the date and the clock to select the time.</li> <li>- <b>Days before today.</b> Filters files that are modified within the specified number of days until the current date (today). Enter the number of days. The current date calculation starts from 00:00 hours.</li> </ul> For example, if you schedule the file ingestion task to run weekly and want to filter for the files that were modified in the previous week, set <b>Days before today</b> to 7. The task will pick up any file with a date between 7 days ago and the date on which it runs.
Time Zone	This applies when <b>File Pickup</b> is <b>By Pattern</b> . If you selected a <b>File Date</b> option, enter the time zone of the location where the files are located.
File Size	This applies when <b>File Pickup</b> is <b>By Pattern</b> . Filters the files to transfer based on file size. Enter the file size, select the file size unit, and filter options.  Select one of the following filter options: <ul style="list-style-type: none"> <li>- <b>Greater than or Equal.</b> Filters files that are greater than or equal to the specified size.</li> <li>- <b>Less than or Equal.</b> Filters files that are less than or equal to the specified size.</li> <li>- <b>Equal.</b> Filters files that have the specified size.</li> </ul>

Advance Source Property	Description
The file path containing the list of files	This applies when <b>File Pickup</b> is <b>By File List</b> . Select this option to provide the path that contains the list of files to pick up and enter the file path.
File list	This applies when <b>File Pickup</b> is <b>By File List</b> . Select this option to provide the list of files to pick up and enter a comma-separated list of file names.
Skip Duplicate Files	Indicates whether to skip duplicate files. If you select this option, the file ingestion task does not transfer files that have the same name and creation date as another file. The file ingestion task marks these files as duplicate in the job log. If you do not select this option, the task transfers all files, even files with duplicate names and creation dates.
Check file stability	Indicates whether to verify that a file is stable before a file ingestion task attempts to pick it. The task skips unstable files it detects in the current run.
Stability check interval	<p>This applies when you enable the <b>Check file stability</b> option. Time in seconds that a file ingestion task waits to check the file stability.</p> <p>For example, if the stability time is 15 seconds, the file ingestion task detects all the files in the source folder that match the defined file pattern, it waits for 15 seconds, and then it processes only the stable files.</p> <p>The interval ranges between 10 seconds to 300 seconds. Default is 10 seconds.</p>
Batch Size	<p>The number of files a file ingestion task can transfer in a batch.</p> <p>Default is 5.</p> <p>The maximum batch size varies, based on the following conditions:</p> <ul style="list-style-type: none"> <li>- If files are transferred from source to target with no intermediate staging server, the maximum number of files the task can transfer in a batch is 8000.</li> <li>- If files are transferred from source to target with intermediate staging server, the maximum number of files the task can transfer in a batch is 20.</li> <li>- If files are transferred from any source to a Snowflake target, the maximum number of files the task can transfer in a batch is 1000.</li> </ul> <p><b>Note:</b> If you transfer files from Azure Blob Storage to Azure SQL Data Warehouse and Snowflake, the task transfers files with no intermediate staging server.</p> <p>When you use a command line to transfer files, the task transfers files with intermediate staging server.</p>

## Source and target parameters

You can configure the file name pattern, folder, and table parameters for sources and targets that a file ingestion task reads from or writes to.

You can use one of the following types of variables to configure a parameter:

- System variables
- User-defined variables

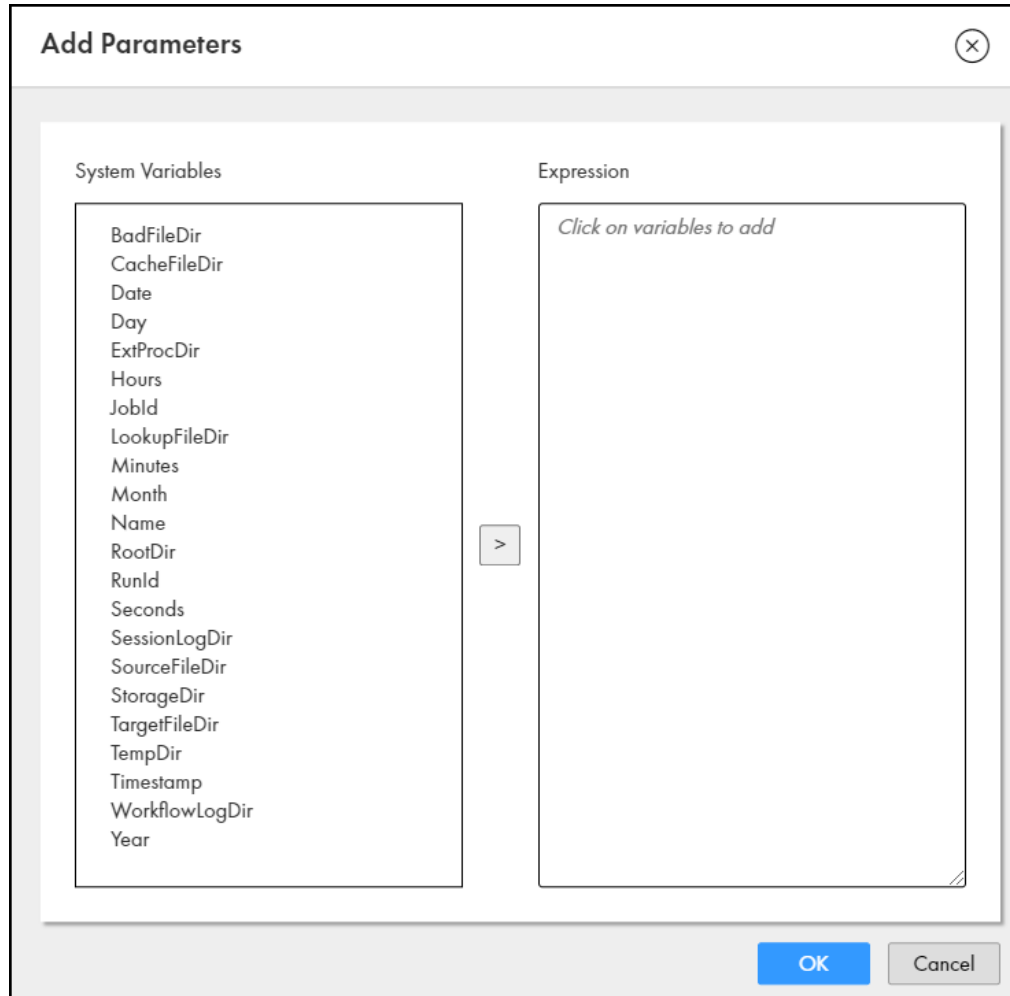
**Note:** You cannot run a task with user-defined variable from the user interface. The value of the user-defined variable must be passed using the job resource of the Mass Ingestion Files REST API. For more information, see [job](#).


## Using system variables to add source and target parameters

Use system variables to add parameters to task sources and targets.

1. Click **Add Parameter** next to the input field, such as **Source Directory** or **Target Directory** on the **Source** or **Target** tab of the task wizard. .

The **Add Parameters** window appears.



2. Select the required variable from the **System Variables** column and click . The selected system variable appears on the **Expression** column. Repeat this procedure to select multiple system variables.

**Note:** When using a system variable within a task, it should be formatted as `${systemvariablename}`.

The following table describes the system variables:

System Variables	Description	Expression
BadFileDir *	Directory for reject files. It cannot include the following special characters: * ? < > "   ,	\${PMBadFileDir}
CacheFileDir *	The location for the cache file.	\${PMCacheDir}
Date **	The current date in ISO (yyyy-MM-dd) format.	\${system.date}
Day **	The day of week	\${system.day}
ExtProcDir *	Directory for external procedures. It cannot include the following special characters: * ? < > "   ,	\${PMExtProcDir}
Hours **	Hours	\${system.hours}
JobId	The id (or job number) of the current job.	\${system.jobid}
LookupFileDir *	Directory for lookup files. It cannot include the following special characters: * ? < > "   ,	\${PMLookupFileDir}
Minutes **	Minutes	\${system.minutes}
Month **	Numerical month	\${system.month}
Name	The name of the current Project.	\${system.name}
RootDir *	Root directory accessible by the node. This is the root directory for other service process variables. It cannot include the following special characters: * ? < > "   ,	\${PMRootDir}
RunId	The id when a job is run.	\${system.runid}
Seconds **	Seconds	\${system.seconds}

System Variables	Description	Expression
SessionLogDir *	Directory for session logs. It cannot include the following special characters: * ? < > "   ,	\${PMSessionLogDir}
SourceFileDir *	Directory for source files. It cannot include the following special characters: * ? < > "   ,	\${PMSourceFileDir}
StorageDir *	Directory for run-time files. Workflow recovery files save to the \$PMStorageDir configured in the PowerCenter Integration Service properties. Session recovery files save to the \$PMStorageDir configured in the operating system profile. It cannot include the following special characters: * ? < > "   ,	\${PMStorageDir}
TargetFileDir *	Directory for target files. It cannot include the following special characters: * ? < > "   ,	\${PMTargetFileDir}
TempDir *	Directory for temporary files. It cannot include the following special characters: * ? < > "   ,	\${PMTempDir}
Timestamp **	The current date and time in ISO (yyyy-MM-dd HH:mm:ss) format.	\${system.timestamp}
WorkflowLogDir *	The location for the workflow log file.	\${PMWorkflowLogDir}
Year **	Year	\${system.year}
* Values are fetched from the Data Integration Server. ** Time zone is the Secure Agent time zone.		

3. Click **OK**.

The expression appears in the input field.

## Using user-defined variables to add source and target parameters

Use user-defined variables to add parameters to add sources and targets.

1. Click an input field, such as **Source Directory** or **Target Directory** on the **Source** or **Target** tab of the task wizard and enter the variable. The variable should be formatted as `${systemvariablename}`.



2. Click **OK**.  
The expression appears in the input field.

## Configuring the target

To configure the target, select a connection type and a connection to which to transfer files and then configure target options.

1. On the **Target** page, select a connection type.

The file ingestion task supports the following target connection types:

- Local folder
- Advanced FTP V2
- Advanced FTPS V2
- Advanced SFTP V2
- Amazon S3 V2
- Amazon Redshift V2
- Google BigQuery V2
- Google Cloud Storage V2
- Hadoop Files V2
- Microsoft Azure Blob Storage V3
- Microsoft Azure Data Lake Store Gen2
- Microsoft Azure Data Lake Store V3
- Microsoft Azure Synapse SQL
- Snowflake Data Cloud
- Databricks Delta

2. Select a connection.
3. Based on the target connection that you select, enter the target options.

Options that appear on the **Target** tab of the task wizard vary based on the type of target connection that you select.

4. Click **Next**.

The **Schedule** tab appears.



## Advanced FTP V2 target properties

When you define a file ingestion task with an Advanced FTP V2 target, you must enter target options on the **Target** tab of the task wizard.

The following table describes the target options:

Option	Description
Target Directory	Directory to where files are transferred. The default value is the target directory specified in the connection. You can enter a relative path. To enter a relative path, start the path with a period, followed by a slash (. /). The path is relative to the target directory specified in the connection.
Add Parameters	Create an expression to add it as a <b>Target Directory</b> parameter. For more information, see <a href="#">"Source and target parameters" on page 420</a> .
If File Exists	Determines what to do with a file if a file with the same name already exists in the target directory. Select one of the following filter options: <ul style="list-style-type: none"><li>- <b>Overwrite</b>. Overwrites the existing file.</li><li>- <b>Append Timestamp</b>. Retains the existing file and appends a timestamp to the name of file being transferred.</li></ul>
Transfer Mode	File transfer mode. Select one of the following filter options: <ul style="list-style-type: none"><li>- <b>Auto</b>. File ingestion determines the transfer mode.</li><li>- <b>ASCII</b>.</li><li>- <b>Binary</b>.</li></ul>

## Advanced FTPS V2 target properties

When you define a file ingestion task with an Advanced FTPS V2 target, you must enter target options on the **Target** tab of the task wizard.

The following table describes the target options:

Option	Description
Target Directory	Directory to where files are transferred. The default value is the target directory specified in the connection. You can enter a relative path. To enter a relative path, start the path with a period, followed by a slash (. /). The path is relative to the target directory specified in the connection.
Add Parameters	Create an expression to add it as a <b>Target Directory</b> parameter. For more information, see <a href="#">"Source and target parameters" on page 420</a> .
If File Exists	Determines what to do with a file if a file with the same name already exists in the target directory. Select one of the following filter options: <ul style="list-style-type: none"><li>- <b>Overwrite</b>. Overwrites the existing file.</li><li>- <b>Append Timestamp</b>. Retains the existing file and appends a timestamp to the name of file being transferred.</li></ul>
Transfer Mode	File transfer mode. Select one of the following filter options: <ul style="list-style-type: none"><li>- <b>Auto</b>. File ingestion determines the transfer mode.</li><li>- <b>ASCII</b>.</li><li>- <b>Binary</b>.</li></ul>

## Advanced SFTP V2 target properties

When you define a file ingestion task with an Advanced FTPS V2 target, you must enter target options on the **Target** tab of the task wizard.

The following table describes the target options:

Option	Description
Target Directory	Directory to where files are transferred. The default value is the target directory specified in the connection. You can enter a relative path. To enter a relative path, start the path with a period, followed by a slash (. /). The path is relative to the target directory specified in the connection.
Add Parameters	Create an expression to add it as a <b>Target Directory</b> parameter. For more information, see <a href="#">"Source and target parameters" on page 420</a> .
If File Exists	Determines what to do with a file if a file with the same name already exists in the target directory. Select one of the following filter options: <ul style="list-style-type: none"><li>- <b>Overwrite</b>. Overwrites the existing file.</li><li>- <b>Append Timestamp</b>. Retains the existing file and appends a timestamp to the name of file being transferred.</li></ul>

## Amazon Redshift V2 target properties

When you define a file ingestion task with an Amazon Redshift V2 target, you must enter target options on the **Target** tab of the task wizard.

Amazon Redshift V2 connection provides the following options that you must select to perform the copy command method:

- **Define Redshift Copy Command Properties**. Select this option to define the Amazon Redshift copy command properties.
- **Enter Custom Redshift Copy Command**. Select this option to provide a custom Amazon Redshift copy command that the file ingestion task uses.

The following table describes the advanced target options that you can configure in a file ingestion task if you select the **Define Redshift Copy Command Properties** option:

Option	Description
Target Table Name	Name of the table in Amazon Redshift to which the files are loaded.
Schema	The Amazon Redshift schema name. Default is the schema that is used while establishing the target connection.
Add Parameters	Create an expression to add it as <b>Schema</b> and <b>Target Table Name</b> parameters. For more information, see <a href="#">"Source and target parameters" on page 420</a> .
Truncate Target Table	Truncate the target table before loading data to the table.
Analyze Target Table	The analyze command collects statistics about the contents of tables in the database to help determine the most efficient execution plans for queries.

Option	Description
Vacuum Target Table	<p>You can select to vacuum the target table to recover disk space and sorts rows in a specified table.</p> <p>Select one of the following recovery options:</p> <ul style="list-style-type: none"> <li>- <b>Full.</b> Sorts the specified table and recovers disk space occupied by rows marked for deletion by previous update and delete operations.</li> <li>- <b>Sort.</b> Sorts the specified table without recovering space freed by deleted rows.</li> <li>- <b>Delete.</b> Recovers disk space occupied by rows marked for deletion by previous update and delete operations, and compresses the table to free up used space.</li> </ul>
File Format and Copy Options	<p>Select the format with which to copy data. Select one of the following options:</p> <ul style="list-style-type: none"> <li>- <b>DELIMITER.</b> A single ASCII character to separate fields in the input file. You can use characters such as pipe ( ), tilde (~), or a tab (\t). The delimiter you specify cannot be a part of the data.</li> <li>- <b>QUOTE.</b> Specifies the quote character used to identify nvarchar characters and skip them.</li> <li>- <b>COMPUPDATE.</b> Overrides current compression encoding and applies compression to an empty table.</li> <li>- <b>AWS_IAM_ROLE.</b> Specify the Amazon Redshift Role Resource Name to run on an Amazon EC2 system.</li> <li>- <b>IGNOREHEADER.</b> Select to ignore headers. For example, if you specify <code>IGNOREHEADER 0</code>, the task processes data from row 0.</li> <li>- <b>DATEFORMAT.</b> Specify the format for date fields.</li> <li>- <b>TIMEFORMAT.</b> Specify the format for time fields.</li> </ul>

The following table describes the advanced target options that you can configure in a file ingestion task if you select the **Enter Custom Redshift Copy Command** option:

Property	Description
Copy Command	<p>Amazon Redshift COPY command appends the data to any existing rows in the table.</p> <p>If the Amazon S3 staging directory and the Amazon Redshift target belongs to different regions, you must specify the region in the COPY command.</p> <p>For example,</p> <pre>copy public.messages from '{{FROM-S3PATH}}' credentials 'aws_access_key_id={{ACCESS-KEY-ID}};aws_secret_access_key={{SECRET-ACCESS-KEY-ID}}' MAXERROR 0 REGION '' QUOTE ''' DELIMITER ',' NULL '' CSV;</pre> <p>Where <code>public</code> is the schema and <code>messages</code> is the table name.</p> <p>For more information about the COPY command, see the AWS documentation.</p>

The following table describes the Amazon Redshift advanced target options that you can configure in a file ingestion task after you select one of the copy command methods:

Property	Description
Pre SQL	SQL command to run before the file ingestion task runs the COPY command.
Post SQL	SQL command to run after the file ingestion task runs the COPY command.

Property	Description
S3 Staging Directory	Specify the Amazon S3 staging directory. You must specify the Amazon S3 staging directory in <bucket_name/folder_name> format. The staging directory is deleted after the file ingestion task runs.
Upload to Redshift with no Intermediate Staging	Upload files from Amazon S3 to Amazon Redshift directly from the Amazon S3 source directory with no additional, intermediate staging. If you select this option, ensure that the Amazon S3 bucket and the Amazon S3 staging directory belong to the same region. If you do not select this option, ensure that the Amazon S3 staging directory and Amazon Redshift target belong to the same region.
File Compression*	Determines whether or not files are compressed before they are transferred to the target directory. Select one of the following options: <ul style="list-style-type: none"> <li>- <b>None.</b> Files are not compressed.</li> <li>- <b>GZIP.</b> Files are compressed using GZIP compression.</li> </ul>
File Encryption Type*	Type of Amazon S3 file encryption to use during file transfer. Select one of the following options: <ul style="list-style-type: none"> <li>- <b>None.</b> Files are not encrypted during transfer.</li> <li>- <b>S3 server-side encryption.</b> Amazon S3 encrypts the file using AWS-managed encryption keys.</li> <li>- <b>S3 client-side encryption.</b> Ensure that unrestricted policies are implemented for the AgentJVM, and that the master symmetric key for the connection is set.</li> </ul> <b>Note:</b> Client-side encryption does not apply to tasks where Amazon S3 is the source.
S3 Accelerated Transfer*	Select whether to use Amazon S3 Transfer Acceleration on the S3 bucket. To use Transfer Acceleration, accelerated transfer must be enabled for the bucket. Select one of the following options: <ul style="list-style-type: none"> <li>- <b>Disabled.</b> Do not use Amazon S3 Transfer Acceleration.</li> <li>- <b>Accelerated.</b> Use Amazon S3 Transfer Acceleration.</li> <li>- <b>Dualstack Accelerated.</b> Use Amazon S3 Transfer Acceleration on a dual-stack endpoint.</li> </ul>
Minimum Upload Part Size*	Minimum upload part size in megabytes when uploading a large file as a set of multiple independent parts. Use this option to tune the file load to Amazon S3.
Multipart Upload Threshold*	Multipart download minimum threshold in megabytes to determine when to upload objects in multiple parts in parallel.
*Not applicable when you read data from Amazon S3 to Amazon Redshift V2.	

## Amazon S3 V2 target properties

When you define a file ingestion task with an Amazon S3 V2 target, you must enter target options on the **Target** tab of the task wizard.

The following table describes the target options:

Option	Description
Folder Path	Amazon S3 folder path to where files are transferred, including bucket name. The default value is the folder path specified in the connection.  You can enter a relative path. To enter a relative path, start the path with a period, followed by a slash (/). The path is relative to the target directory specified in the connection.
Add Parameters	Create an expression to add it as a <b>Folder Path</b> parameter. For more information, see <a href="#">“Source and target parameters” on page 420</a> .
File Compression*	Determines whether or not files are compressed before they are transferred to the target directory. Select one of the following options: <ul style="list-style-type: none"> <li>- <b>None</b>. Files are not compressed.</li> <li>- <b>GZIP</b>. Files are compressed using GZIP compression.</li> </ul>
File Encryption Type*	Type of Amazon S3 file encryption to use during file transfer. Select one of the following options: <ul style="list-style-type: none"> <li>- <b>None</b>. Files are not encrypted during transfer.</li> <li>- <b>S3 server-side encryption</b>. Amazon S3 encrypts the file using AWS-managed encryption keys.</li> <li>- <b>S3 client-side encryption</b>. Ensure that unrestricted policies are implemented for the AgentJVM, and that the master symmetric key for the connection is set.</li> </ul>
If File Exists*	Determines what to do with a file if a file with the same name already exists in the target directory. Select one of the following filter options: <ul style="list-style-type: none"> <li>- <b>Overwrite</b>. Overwrites the existing file.</li> <li>- <b>Append Timestamp</b>. Retains the existing file and appends a timestamp to the name of file being transferred.</li> </ul>
S3 Accelerated Transfer*	Select whether to use Amazon S3 Transfer Acceleration on the S3 bucket. To use Transfer Acceleration, accelerated transfer must be enabled for the bucket. Select one of the following options: <ul style="list-style-type: none"> <li>- <b>Disabled</b>. Do not use Amazon S3 Transfer Acceleration.</li> <li>- <b>Accelerated</b>. Use Amazon S3 Transfer Acceleration.</li> <li>- <b>Dualstack Accelerated</b>. Use Amazon S3 Transfer Acceleration on a dual-stack endpoint.</li> </ul>
Minimum Upload Part Size*	Minimum upload part size in megabytes when uploading a large file as a set of multiple independent parts. Use this option to tune the file load to Amazon S3.
Multipart Upload Threshold*	Multipart download minimum threshold in megabytes to determine when to upload objects in multiple parts in parallel.
<i>*Not applicable when you read data from Databricks Delta.</i>	

## Databricks Delta target properties

When you define a file ingestion task with a Databricks Delta target, you must enter target options on the **Target** tab of the task wizard.

**Note:** You can transfer only Parquet files from Amazon S3 V2 source and a Microsoft Azure Data Lake Store Gen2 source to a Databricks Delta target, and all the files must have the same metadata.

The following table describes the target options:

Option	Description
Database	Required. Name of the database in Databricks Delta Lake that contains the target table. Default value is the database name specified in the Databricks Delta connection.
Add Parameters	Create an expression to add it as <b>Database</b> and <b>Table Name</b> parameters. For more information, see <a href="#">"Source and target parameters" on page 420</a> .
Table Name	Required. Name of the table in Databricks Delta Lake. <b>Note:</b> If you specify the name of a table that does not exist in target database, the Secure Agent creates a new table with the specified name.
If Table Exists	Determines the action that the Secure Agent must take on a table if the table name matches the name of an existing table in the target database. Select one of the following filter options: <ul style="list-style-type: none"><li>- <b>Overwrite</b></li><li>- <b>Append</b></li></ul> Default is <b>Overwrite</b> .

**Note:** If a job fails with the following error, see the cluster logs for more information:

```
"[ERROR] Job execution failed. State : JOB_FAILED ; State Message :"
```

## Google BigQuery V2 target properties

When you define a file ingestion task with a Google BigQuery V2 target, you must enter target options on the **Target** tab of the task wizard.

**Note:** When you define a file ingestion task with a Google BigQuery V2 target, you can configure only Google Cloud Storage V2 as a source.

The following table describes the target options:

Option	Description
Target Table Name	Specify the Google BigQuery target table name.
Dataset ID	Specify the Google BigQuery dataset name.
Add Parameters	Create an expression to add it as <b>Target Table Name</b> and <b>Dataset ID</b> parameters. For more information, see <a href="#">"Source and target parameters" on page 420</a> .
Field Delimiter	Indicates whether Google BigQuery V2 Connector must allow field separators for the fields in a .csv file.
Quote Character	Specifies the quote character to skip when you write data to Google BigQuery. When you write data to Google BigQuery and the source table contains the specified quote character, the task fails. Change the quote character value to a value that does not exist in the source table.
Allow Quoted Newlines	Indicates whether Google BigQuery V2 Connector must allow the quoted data sections with newline character in a .csv file.
Allow Jagged Rows	Indicates whether Google BigQuery V2 Connector must accept the rows without trailing columns in a .csv file.

Option	Description
Skip Leading Rows	Specifies the number of top rows in the source file that Google BigQuery V2 Connector skips when loading the data. The default value is 0.
Data format of the File	Specifies the data format of the source file. You can select one of the following data formats: <ul style="list-style-type: none"> <li>- JSON (Newline Delimited)</li> <li>- CSV</li> <li>- Avro</li> <li>- Parquet</li> <li>- ORC</li> </ul>
Write Disposition	Specifies how Google BigQuery V2 Connector must write data in bulk mode if the target table already exists. You can select one of the following values: <ul style="list-style-type: none"> <li>- Write Append. If the target table exists, Google BigQuery V2 Connector appends the data to the existing data in the table.</li> <li>- Write Truncate. If the target table exists, Google BigQuery V2 Connector overwrites the existing data in the table.</li> <li>- Write Empty. If the target table exists and contains data, Google BigQuery V2 Connector displays an error and does not write the data to the target. Google BigQuery V2 Connector writes the data to the target only if the target table does not contain any data.</li> </ul>

## Google Cloud Storage V2 target properties

When you define a file ingestion task with a Google Cloud Storage V2 target, you must enter target options on the **Target** tab of the task wizard.

The following table describes the target options:

Option	Description
Folder Path	Path in Google Cloud Storage where files are transferred. You can either enter the bucket name or the bucket name and folder name. For example, enter <code>&lt;bucket name&gt;</code> or <code>&lt;bucket name&gt;/&lt;folder name&gt;</code> <b>Note:</b> Do not use a single slash (/) in the beginning of path. You can enter a relative path. To enter a relative path, start the path with a period, followed by a slash (./). The path is relative to the bucket specified in the connection.
Add Parameters	Create an expression to add it as a <b>Folder Path</b> parameter. For more information, see <a href="#">"Source and target parameters" on page 420</a> .
File Compression	Determines whether or not files are compressed before they are transferred to the target directory. Select one of the following options: <ul style="list-style-type: none"> <li>- <b>None</b>. Files are not compressed.</li> <li>- <b>GZIP</b>. Files are compressed using the GZIP compression format.</li> </ul>
If File Exists	Determines the action that the Secure Agent must take with a file if a file with the same name exists in the target directory. Select one of the following options: <ul style="list-style-type: none"> <li>- <b>Overwrite</b></li> <li>- <b>Append Timestamp</b></li> </ul>

## Hadoop Files V2 target properties

When you define a file ingestion task with a Hadoop Files V2 target, you must enter target options on the **Target** tab of the task wizard.

The following table describes the target option:

Option	Description
Target Directory	Directory to where files are transferred.
Add Parameters	Create an expression to add it as a <b>Target Directory</b> parameter. For more information, see <a href="#">"Source and target parameters" on page 420</a> .

## Local folder target properties

When you define a file ingestion task with a local folder target, you must enter target properties on the **Target** tab of the task wizard.

The following table describes the target options:

Option	Description
Target Directory	Directory to where files are transferred. The Secure Agent must be able to access the directory. <b>Note:</b> File listener can access files and directories on network shares with support for NFS and CIFS.
Add Parameters	Create an expression to add it as a <b>Target Directory</b> parameter. For more information, see <a href="#">"Source and target parameters" on page 420</a> .
If File Exists	Determines what to do with a file if a file with the same name exists in the target directory. The following options are available: <ul style="list-style-type: none"><li>- <b>Overwrite</b></li><li>- <b>Append Timestamp</b></li></ul>

## Microsoft Azure Blob Storage V3 target properties

When you define a file ingestion task with a Microsoft Azure Blob Storage target, you must enter target options on the **Target** tab of the task wizard.

The following table describes the target options:

Option	Description
Blob Container	Microsoft Azure Blob Storage container, including folder path and container name.
Add Parameters	Create an expression to add it as a <b>Blob Container</b> parameter. For more information, see <a href="#">"Source and target parameters" on page 420</a> .
Blob Type	Type of blob. Select one of the following options: <ul style="list-style-type: none"><li>- <b>Block Blob</b>. Ideal for storing text or binary files, such as documents and media files.</li><li>- <b>Append Blob</b>. Optimized for append operations, for example, logging scenarios.</li></ul>



Option	Description
File Compression	Determines whether or not files are compressed before they are transferred to the target directory. The following options are available: <ul style="list-style-type: none"> <li>- <b>None</b>. Files are not compressed.</li> <li>- <b>GZIP</b>. Files are compressed using GZIP compression.</li> </ul>
Number of Concurrent Connections to Blob Store	Number of concurrent connections to the Microsoft Azure Blob Store Storage container.

## Microsoft Azure Data Lake Storage Gen2 target properties

When you define a file ingestion task with a Microsoft Azure Data Lake Storage Gen2 target, you must enter target options on the **Target** tab of the task wizard.

The following table describes the target options:

Target Property	Description
Target Directory	Directory to where files are transferred. The directory is created at run time if it does not exist. The directory path specified at run time overrides the path specified while creating a connection. The default value is the target directory specified in the connection. You can enter a relative path. To enter a relative path, start the path with a period, followed by a slash (.). The path is relative to the target directory specified in the connection.
Add Parameters	Create an expression to add it as a <b>Target Directory</b> parameter. For more information, see <a href="#">"Source and target parameters" on page 420</a> .
File Compression*	Determines whether or not files are compressed before they are transferred to the target directory. The following options are available: <ul style="list-style-type: none"> <li>- <b>None</b>. Files are not compressed.</li> <li>- <b>GZIP</b>. Files are compressed using GZIP compression.</li> </ul>
If File Exists*	Determines what to do with a file if a file with the same name exists in the target directory. The following options are available: <ul style="list-style-type: none"> <li>- <b>Overwrite</b></li> <li>- <b>Append</b></li> <li>- <b>Fail</b></li> </ul>
Block Size (Bytes)*	Divides a large file into smaller specified block size. When you write a large file, divide the file into smaller parts and configure concurrent connections to spawn the required number of threads to process data in parallel. Default is 8388608 bytes (8 MB).
<i>*Not applicable when you read data from Databricks Delta.</i>	

## Microsoft Azure Data Lake Store Gen1 V3 target properties

When you define a file ingestion task with a Microsoft Azure Data Lake Store Gen1 V3 target, you must enter target options on the **Target** tab of the task wizard.

The following table describes the target options:

Target Property	Description
Target Directory	Directory to where files are transferred. The default value is the target directory specified in the connection. You can enter a relative path. To enter a relative path, start the path with a period, followed by a slash (.). The path is relative to the target directory specified in the connection.
Add Parameters	Create an expression to add it as a <b>Target Directory</b> parameter. For more information, see <a href="#">"Source and target parameters" on page 420</a> .
File Compression	Determines whether or not files are compressed before they are transferred to the target directory. The following options are available: <ul style="list-style-type: none"><li>- <b>None</b>. Files are not compressed.</li><li>- <b>GZIP</b>. Files are compressed using GZIP compression.</li></ul>
If File Exists	Determines what to do with a file if a file with the same name exists in the target directory. The following options are available: <ul style="list-style-type: none"><li>- <b>Overwrite</b></li><li>- <b>Append</b></li><li>- <b>Fail</b></li></ul>

## Microsoft Azure Synapse SQL target properties

When you define a file ingestion task with a Microsoft Azure Synapse SQL target, you must enter target options on the **Target** tab of the task wizard.

The following table describes the target options:

Property	Description
Ingestion Method	The ingestion method to load data to Microsoft Azure Synapse SQL. Select one of the following options: <ul style="list-style-type: none"><li>- <b>Polybase</b></li><li>- <b>COPY Command</b></li></ul>
Command Type	The command type for the ingestion method. Select one of the following options: <ul style="list-style-type: none"><li>- <b>Auto Generated</b>. Select this option to define the command properties.</li><li>- <b>Custom</b>. Select this option to provide a custom command that the file ingestion task uses.</li></ul>

The following table describes the Microsoft Azure Synapse SQL advanced target options when you select **Polybase** or **COPY Command** ingestion method and **Auto Generated** command type:

**Note:** The **Auto Generated** command type is applicable only for files in text and CSV formats.

Property	Description
Target Table Name	Name of the table in Microsoft Azure Synapse SQL to which the files are loaded.
Add Parameters	Create an expression to add it as <b>Target Table Name</b> and <b>Schema</b> parameters. For more information, see <a href="#">"Source and target parameters" on page 420</a> .

Property	Description
Schema	The Microsoft Azure Synapse SQL schema name.
Truncate Target Table	Truncate the target table before loading.
Pre SQL	SQL command to run before the file ingestion task runs the PolyBase or Copy command.
Post SQL	SQL command to run after the file ingestion task runs the PolyBase or Copy command.
Field Delimiter	Character used to separate fields in the file. Default is 0x1e. You can select the following field delimiters from the list: ~ `   . TAB 0x1e
Quote Character	Specifies the quote character to skip when you write data to Microsoft Azure Synapse SQL. When you write data to Microsoft Azure Synapse SQL and the source table contains the specified quote character, the task fails. Change the quote character value to a value that does not exist in the source table.
External Stage*	Specifies the external stage directory to use for loading files into Microsoft Azure Synapse SQL. You can stage the files in Microsoft Azure Blob Storage or Microsoft Azure Data Lake Storage Gen2.
File Compression*	Determines whether or not files are compressed before they are transferred to the target directory. The following options are available: - <b>None</b> . Files are not compressed. - <b>GZIP</b> . Files are compressed using GZIP compression.
Number of Concurrent Connections*	Number of concurrent connections to extract data from the Microsoft Azure Blob Storage or Microsoft Azure Data Lake Storage Gen2. When reading a large file or object, you can spawn multiple threads to process data. Configure <b>Blob Part Size</b> or <b>Block Size</b> to divide a large file into smaller parts. Default is 4. Maximum is 10.
<i>*Not applicable when you read data from Microsoft Azure Blob Storage or Microsoft Azure Data Lake Storage Gen2.</i>	

The following table describes the Microsoft Azure Synapse SQL advanced target properties when you select **Polybase** or **COPY Command** ingestion method and **Custom** command type:

Property	Description
File Format Definition	<p>Applies to Polybase ingestion method.</p> <p>Transact-SQL CREATE EXTERNAL FILE FORMAT statement. For example:</p> <pre>CREATE EXTERNAL FILE FORMAT {{fileFormatName}} WITH ( FORMAT_TYPE = DELIMITEDTEXT, FORMAT_OPTIONS (FIELD_TERMINATOR = ',', STRING_DELIMITER = '"') )</pre> <p>The following is an example to create an external file in parquet format:</p> <pre>CREATE EXTERNAL FILE FORMAT {{fileFormatName}} WITH (FORMAT_TYPE = PARQUET)</pre> <p>Similarly, you can create an external file in JSON, Avro, and ORC formats.</p> <p>For more information about the CREATE EXTERNAL FILE FORMAT statement, see the Microsoft documentation.</p>
External Table Definition	<p>Applies to Polybase ingestion method.</p> <p>Transact-SQL CREATE EXTERNAL TABLE statement. For example:</p> <pre>CREATE EXTERNAL TABLE {{externalTable}} ( id INT, name NVARCHAR ( 100 ) ) WITH (LOCATION = '{{blobLocation}}', DATA_SOURCE = {{dataSourceName}}, FILE_FORMAT = {{fileFormatName}})</pre> <p>The following is an example to create an external table in parquet format:</p> <pre>CREATE EXTERNAL TABLE {{externalTable}} (username VARCHAR(100),number int,colour VARCHAR(100))WITH (LOCATION='{{blobLocation}}',DATA_SOURCE={{dataSourceName}},FILE_FORMAT={{fileFormatName}})</pre> <p>Similarly, you can create an external table in JSON, Avro, and ORC formats.</p> <p>For more information about the CREATE EXTERNAL TABLE statement, see the Microsoft documentation.</p>
Insert SQL Definition	<p>Applies to Polybase ingestion method.</p> <p>Transact-SQL INSERT statement. For example:</p> <pre>INSERT INTO schema.table (id, name) SELECT id+5, name FROM {{externalTable}}</pre> <p>The following is an example for defining insert SQL in parquet format:</p> <pre>INSERT INTO testing.test_parq(username,number,colour) SELECT username, number,colour FROM {{externalTable}};</pre> <p>Similarly, you can define insert SQL in JSON, Avro, and ORC formats.</p> <p>For information about the INSERT statement, see the Microsoft documentation.</p>
Copy Command Definition	<p>Applies to COPY Command ingestion method.</p> <p>Transact-SQL COPY INTO statement. For example:</p> <pre>COPY INTO schema.table FROM EXTERNALLOCATION WITH (CREDENTIAL = (AZURECREDENTIALS), FIELDTERMINATOR = ',', FIELDQUOTE = '')</pre> <p>The following is an example for defining COPY Command in parquet format:</p> <pre>COPY INTO testing.test_parq FROM EXTERNALLOCATION WITH (CREDENTIAL = (AZURECREDENTIALS), FILE_TYPE = 'PARQUET')</pre> <p>Similarly, you can define COPY Command in JSON, Avro, and ORC formats.</p> <p>For more information about the COPY INTO statement, see the Microsoft documentation.</p>
Pre SQL	SQL command to run before the file ingestion task runs the PolyBase command.
Post SQL	SQL command to run after the file ingestion task runs the PolyBase command.
External Stage*	Specifies the external stage directory to use for loading files into Microsoft Azure Synapse SQL. You can stage the files in Microsoft Azure Blob Storage or Microsoft Azure Data Lake Storage Gen2.

Property	Description
Number of Concurrent Connections*	Number of concurrent connections to extract data from the Microsoft Azure Blob Storage or Microsoft Azure Data Lake Storage Gen2. When reading a large file or object, you can spawn multiple threads to process data. Configure <b>Blob Part Size</b> or <b>Block Size</b> to divide a large file into smaller parts. Default is 4. Maximum is 10.
<i>*Not applicable when you read data from Microsoft Azure Blob Storage or Microsoft Azure Data Lake Storage Gen2.</i>	

## Snowflake Data Cloud target properties

When you define a file ingestion task with a Snowflake Data Cloud target, you must enter target options on the **Target** tab of the task wizard.

The following table describes the target options:

Property	Description
Warehouse	Overrides the name specified in the Snowflake Data Cloud connection.
Add Parameters	Create an expression to add it as <b>Warehouse</b> , <b>Database</b> , <b>Schema</b> , and <b>Target Table Name</b> parameters. For more information, see <a href="#">"Source and target parameters" on page 420</a> .
Database	The database name of Snowflake Data Cloud.
Schema	The schema name in Snowflake Data Cloud.
Target Table Name	The table name of the Snowflake Data Cloud target table. The target table name is case-sensitive.
Role	Overrides the Snowflake Data Cloud user role specified in the connection.
Pre SQL	SQL statement to run on the target before the start of write operations.
Post SQL	SQL statement to run on the target table after a write operation completes.
Truncate Target Table	Truncates the database target table before inserting new rows. Select one of the following options: <ul style="list-style-type: none"> <li>- <b>True</b>. Truncates the target table before inserting all rows.</li> <li>- <b>False</b>. Inserts new rows without truncating the target table</li> </ul> Default is false.

Property	Description
File Format and Copy Options	<p>The copy option and the file format to load the data to Snowflake Data Cloud.</p> <p>The copy option specifies the action that the task performs when an error is encountered while loading data from a file:</p> <p>You can specify the following copy option to abort the COPY statement if any error is encountered:</p> <pre>ON_ERROR = ABORT_STATEMENT</pre> <p>When you load files, you can specify the file format and define the rules for the data files. The task uses the specified file format and rules while bulk loading data into Snowflake Data Cloud tables.</p> <p>The following formats are supported:</p> <ul style="list-style-type: none"> <li>- CSV</li> <li>- JSON</li> <li>- Avro</li> <li>- ORC</li> <li>- Parquet</li> </ul>
External Stage	<p>Specifies the external stage directory to use for loading files into Snowflake Data Cloud tables.</p> <p>Ensure that the source folder path you specify is the same as the folder path provided in the URL of the external stage for the specific connection type in Snowflake Data Cloud.</p> <p>Applicable when the source for file ingestion is Microsoft Azure Blob Storage and Amazon S3. The external stage is mandatory when you use the connection type Microsoft Azure Blob Storage V3, but is optional for Amazon S3 V2. If you do not specify an external stage for Amazon S3 V2, Snowflake Data Cloud creates an external stage by default.</p>
File Compression	<p>Determines whether or not files are compressed before they are transferred to the target directory.</p> <p>The following options are available:</p> <ul style="list-style-type: none"> <li>- <b>None</b>. Files are not compressed.</li> <li>- <b>GZIP</b>. Files are compressed using GZIP compression.</li> </ul> <p>Applicable for all sources that support the file ingestion task except for Microsoft Azure Blob Storage V3 and Amazon S3 V2.</p>

## File format and copy options

When you configure a file ingestion task to transfer a large number of files to Snowflake Data Cloud, specify the copy option and the file format to load the data.

Select a Snowflake Data Cloud connection in a file ingestion task and then specify the copy option and the file format in the target options to determine how to load the files to a Snowflake Data Cloud target table.

The copy option specifies the action that the task performs when an error is encountered while loading data from a file.

You can specify the following copy option to abort the COPY statement if any error is encountered:

```
ON_ERROR = ABORT_STATEMENT
```

**Note:** The file ingestion task for Snowflake Data Cloud is certified for only the ABORT\_STATEMENT for ON\_ERROR copy option.

When you load files, you can specify the file format and define the rules for the data files. The task uses the specified file format and rules while bulk loading data into Snowflake Data Cloud tables.

The following list describes some of the format type options:

- `RECORD_DELIMITER = '<character>' | NONE`. Single character string that separates records in an input file.
- `FIELD_DELIMITER = '<character>' | NONE`. Specifies the single character string that separates records in an input file.
- `FILE_EXTENSION = '<string>' | NONE`. Specifies the extension for files unloaded to a stage.
- `SKIP_HEADER = <integer>`. Number of lines at the start of the file to skip.
- `DATE_FORMAT = '<string>' | AUTO`. Defines the format of date values in the data files or table.
- `TIME_FORMAT = '<string>' | AUTO`. Defines the format of time values in the data files or table.
- `TIMESTAMP_FORMAT = <string>' | AUTO`. Defines the format of timestamp values in the data files or table.

### Example of File format and copy options for loading files to Snowflake

You want to create a CSV file format and define the following rules to load files to Snowflake:

- Delimit the fields using the pipe character (|).
- Files include a single header line that will be skipped.

Specify the following file format: `file_format = (type = csv field_delimiter = '|' skip_header = 1)`

You can specify both the copy options and file format by using the following character: `&&`

For example, `file_format = (type = csv field_delimiter = ',' skip_header = 2)&&on_error=ABORT_STATEMENT`

Similarly, use the following file format in the **File Format and Copy Options** field to load data into separate columns:

- For JSON: `on_error='ABORT_STATEMENT'&&file_format = (type = json)&&MATCH_BY_COLUMN_NAME=CASE_INSENSITIVE`
- For AVRO: `on_error='ABORT_STATEMENT'&&file_format = (type = avro)&&MATCH_BY_COLUMN_NAME=CASE_INSENSITIVE`
- For ORC: `on_error='ABORT_STATEMENT'&&file_format = (type = orc)&&MATCH_BY_COLUMN_NAME=CASE_INSENSITIVE`
- For PARQUET: `on_error='ABORT_STATEMENT'&&file_format = (type = parquet)&&MATCH_BY_COLUMN_NAME=CASE_INSENSITIVE`

The string `MATCH_BY_COLUMN_NAME` specifies whether to load the semi-structured data into the columns in the target table that match the corresponding columns represented in the data. `CASE_SENSITIVE`, `CASE_INSENSITIVE`, and `NONE` are the supported options. Default is `NONE`.

Consider the following criteria for a column to match between the data and table:

- The column represented in the data must have the same name as the column in the table. The column names are either case-sensitive (`CASE_SENSITIVE`) or case-insensitive (`CASE_INSENSITIVE`).
- The column can be in any order.
- The column in the table must have a data type that is compatible with the values in the column represented in the data. For example, string, number, and Boolean values can be loaded into a variant column.

For more information about the various file formats that you can specify and the copy option, see the Snowflake Data Cloud documentation at the following website:

<https://docs.snowflake.net/manuals/sql-reference/sql/copy-into-table.html#copy-options-copyoptions>

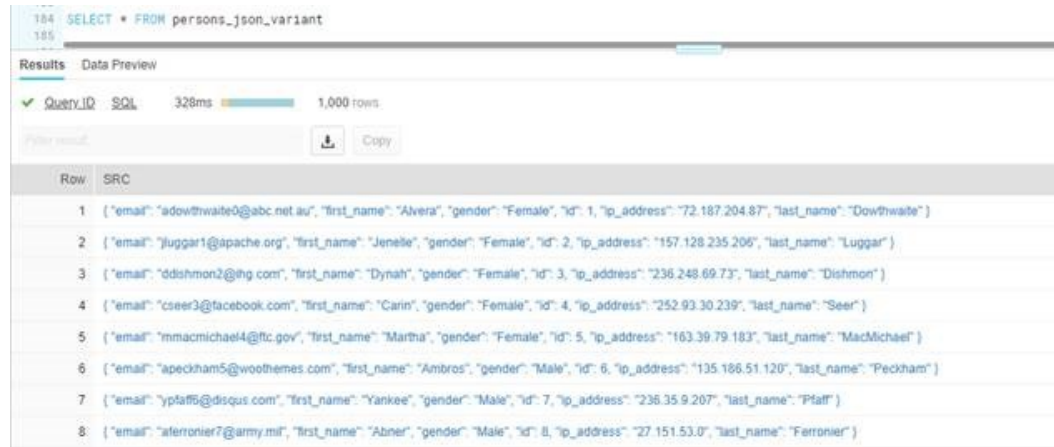
## Rules and guidelines for loading JSON files

Consider the following rule and guideline when you load files of the JSON format to Snowflake Data Cloud.

When you load files of the JSON format to Snowflake Data Cloud, the target table must have only one column of variant type.

To load files of JSON format to columnar format, consider the following tasks:

For example, see the following data view in a table with variant column:



Row	SRC
1	{ "email": "adownthwaite0@abc.net.au", "first_name": "Alvera", "gender": "Female", "id": 1, "ip_address": "72.187.204.87", "last_name": "Dowthwaite" }
2	{ "email": "juggar1@apache.org", "first_name": "Jenelle", "gender": "Female", "id": 2, "ip_address": "157.128.235.206", "last_name": "Luggar" }
3	{ "email": "ddishmon2@ihg.com", "first_name": "Dynah", "gender": "Female", "id": 3, "ip_address": "236.248.69.73", "last_name": "Dishmon" }
4	{ "email": "cseer3@facebook.com", "first_name": "Carin", "gender": "Female", "id": 4, "ip_address": "252.93.30.239", "last_name": "Seer" }
5	{ "email": "mmacmichael4@ftc.gov", "first_name": "Martha", "gender": "Female", "id": 5, "ip_address": "163.39.79.183", "last_name": "MacMichael" }
6	{ "email": "apeckham5@woothemes.com", "first_name": "Ambros", "gender": "Male", "id": 6, "ip_address": "135.186.51.120", "last_name": "Peckham" }
7	{ "email": "yplatt5@disqus.com", "first_name": "Yankee", "gender": "Male", "id": 7, "ip_address": "236.35.9.207", "last_name": "Platt" }
8	{ "email": "aferronier7@army.mil", "first_name": "Abner", "gender": "Male", "id": 8, "ip_address": "27.151.53.0", "last_name": "Feronier" }

To update the table to columnar format, run the following SQL query from the **Post-processing Commands** field in the mapping task:

```
INSERT INTO PERSONS_JSON SELECT parse_json($1):email,  
parse_json($1):first_name,  
parse_json($1):gender,  
parse_json($1):id,  
parse_json($1):ip_address,  
parse_json($1):last_name from PERSONS_JSON_VARIANT
```

After you run the mapping task, the Secure Agent copies the data in columnar format to Snowflake:

Row	EMAIL	FIRST_NAME	GENDER	ID	IP_ADDRESS	LAST_NAME
1	adownthwaite0@abc.net.au	Alvera	Female	1	72.187.204.87	Dowthwaite
2	juggar1@apache.org	Jenelle	Female	2	157.128.235.206	Luggar
3	ddishmon2@ihg.com	Dynah	Female	3	236.248.69.73	Dishmon
4	cseer3@facebook.com	Carin	Female	4	252.93.30.239	Seer
5	mmacmichael4@ftc.gov	Martha	Female	5	163.39.79.183	MacMichael
6	apeckham5@woothemes.com	Ambros	Male	6	135.186.51.120	Peckham
7	yplatt5@disqus.com	Yankee	Male	7	236.35.9.207	Platt
8	aferronier7@army.mil	Abner	Male	8	27.151.53.0	Feronier

## External stage

When you configure a file ingestion task to load files from a Microsoft Azure Blob Storage or Amazon S3 source to the Snowflake Data Cloud tables, specify the external staging directory to use in Snowflake.

You must specify the external stage name for the specific connection type that you want to use in the **Target Options** section in the file ingestion task.

The external stage field value is mandatory when you run a file ingestion task to load files from Microsoft Azure Blob Storage to Snowflake Data Cloud where the connection type in the source is Microsoft Azure Blob Storage V2. When the source connection type is Amazon S3 V2, and you do not specify an external stage for Amazon S3 V2 in the Snowflake Data Cloud target options, Snowflake creates an external stage directory by default.



Ensure that the source directory path in the **Source Options** of the file ingestion task is the same as the directory path provided in the URL of the external stage created for the Microsoft Azure Blob Storage V2 or Amazon S3 V2 connection in Snowflake Data Cloud.

For example, an external stage for Microsoft Azure Blob Storage created using an Azure account name and a blob container with a folder path has the following stage URL: 'azure://<URL>/<blob container>/<folder path>'. The stage uses the file format you specify in the **Target Options** of the file ingestion task.

The following image shows the stage name and the stage URL for a Microsoft Azure Blob Storage V2 connection in Snowflake Data Cloud:

**Create Stage**

Staged files will be stored in the specified Azure location

Name\*

Schema Name

URL\*

Azure SAS Token

Encryption Master Key

Comment

[Show SQL](#)

In the example, the stage URL is `azure://adapterdevblob.blob.core.windows.net/snowflakemi/MI/` and the external stage name is `MFT_BLOB1`.

When you create a file ingestion job, in the **Folder Path** field in the **Source Options** of the Microsoft Azure Blob Storage V2 source, specify the following `<Blob Container>/<folder path>` path from the stage URL: `/snowflakemi/MI`

The following image shows the specified source folder path in the **Source Options** section:

**Source Details**

Source Type: ☒ Source Connection ☐ File Listener

Connection Type:

Connection:  [View](#)

Description:

Account Name: adapterdevblob

**Source Options**

File Pickup: ☒ By Pattern ☐ By File List

Source Directory:  [Add Parameters](#)

☐ Include files from sub-folders

☐ File Pattern

☒ File Date

☐ File Size

☐ Skip duplicate files

☐ Check File Stability

Batch Size:

In the **Target Options** for Snowflake Data Cloud, specify the following name of the created external stage:  
MFT\_BLOB1

The following figure shows the configured external stage field in the **Target Options** section:

**Target Details**

Connection Type: \* Snowflake Data Cloud

Connection: \* ASnowflakeoldconn View

Description: informatica

Additional JDBC URL Parameter:

**Target Options**

Warehouse: \* TEST\_WH [Add Parameters](#)

Database: \* SALES [Add Parameters](#)

Schema: \* MFT [Add Parameters](#)

Target Table Name: \* TEST\_MI [Add Parameters](#)

Role: \* UNIT

Pre SQL:

Post SQL:

☐ Truncate Target Table ?

File Format And Copy Options: ? on\_error=ABORT\_STATEMENT

External Stage: ? MFT\_BLOB1

#### Rules and Guidelines for Snowflake Data Cloud file ingestion tasks

When you configure a file ingestion task to write to Snowflake Data Cloud from the supported file ingestion sources, you must specify a batch size for the maximum number of files to be transferred in a batch. Specify a value for the batch size in the required source properties of the file ingestion task. When you specify a batch size, the performance of the task is optimized.

The default batch size is 5. When you write from Amazon S3 or Azure Blob Storage sources to a Snowflake target, you can specify a maximum batch size of 1000 in the Amazon S3 or Azure Blob Storage source properties. For other file ingestion supported sources, you must specify a batch size between 1 and 20.

## Configuring file processing actions

You can define file processing actions, such as compress and encrypt, that Mass Ingestion Files performs on files before it transfers them.

1. To add file processing actions, click the plus sign in the Actions tab.

The **Action Details** window appears.

2. Perform the following file processing actions:

Action	Description
Compress	To compress the files, select <b>Compress</b> . Then select one of the following action types: <ul style="list-style-type: none"><li>- Zip</li><li>- Gzip</li><li>- Tar</li></ul>
Decompress	To decompress compressed files, select <b>Decompress</b> . Then select one of the following action types: <ul style="list-style-type: none"><li>- Unzip</li><li>- Gunzip</li><li>- Untar</li></ul> <p><b>Note:</b> Use the action type that corresponds to the compression action type that was used to compress the file. For example, for a .zip file, use the Unzip method.</p>
Encrypt	To encrypt files by using the PGP encryption method, select <b>Encrypt</b> . Then select <b>PGP</b> and enter the key ID of the user who decrypts the file. <ul style="list-style-type: none"><li>- To add your sign key, select <b>Sign</b>. The key ID and the key passphrase enables.</li><li>- Enter your private key ID and key passphrase.</li></ul> <p><b>Note:</b> For more information about securing files that file ingestion transfers, see <a href="#">"Mass Ingestion Files security" on page 395</a>.</p>
Decrypt	To decrypt PGP-encrypted files, select <b>Decrypt</b> . Then select <b>PGP</b> and enter the key passphrase of the user of the target directory.
Flatten file structure	To move files from multiple folders to a single folder in the target directory, select <b>File Structure</b> and then select <b>Flatten</b> .
Virus scan	To scan files for viruses by using the ICAP protocol, select <b>Virus Scan</b> . Then select <b>ICAP</b> and enter the <b>ICAP Server URL</b> or the server where the files are scanned. ICAP sends a response code that indicates whether the malwares are detected in the files. <p><b>Note:</b> Use the ICAP server of the organization.</p>

3. Click **Save**.

To add another action, click the plus sign. To delete an action, click **Delete**. To change the order in which the file ingestion task processes files, drag and drop the sequence of actions.

4. Click **Next**. The Runtime Options tab appears.

## Configuring runtime options

You can run a file ingestion task manually, or you can schedule the task to run at a specific time or when a file is ready to transfer. You can receive notifications if the task fails. You can run multiple jobs and file batches concurrently. You can also select the log level that a job creates.

1. On the **Runtime Options** page, under **Schedule Details**, select one of the following options:
  - Do not run this task on a schedule. The task does not run according to a defined schedule. You can run the task manually.
  - Run this task on a schedule. Select a schedule by which the task runs.
  - Run this task by file listener. The task runs when the file listener notifies the task of a file event. A file event occurs when files arrive to the monitored folder or the files in the monitored folder are updated or deleted.  
You must create a file listener that listens to the folder where files arrive. For more information about creating a file listener, see the section "File listeners" in the *Components* help.
2. Under **Failure Management**, select the **Send a notification on failure** option to receive notifications if the task fails and if the task detects infected files. Enter a comma-separated list of email addresses to which to send the notifications.  
  
If the file ingestion task detects an infected file, it copies the file from the source to the quarantine directory. The default directory path is `<agent location>/data/quarantine`.
3. Under **Advanced Options**, select **Allow concurrency** to run multiple file ingestion task jobs concurrently.  
**Warning:** Running concurrent jobs might cause unexpected results if the targets include duplicate files.
4. Select the number of file batches to run in parallel. Default is 1.
5. Select the log level to determine the level of detail in the logs that the job creates. Select one of the following options:
  - Normal. Logs project-level information, such as the name of the user that ran the project, when the project started, any variables that were passed, and when the project stopped. It also logs any errors encountered.
  - Silent. Additionally logs the start and stop times of tasks.
  - Verbose. Additionally logs task-level details, such as the names of the files that were processed.
  - Debug. Additionally logs detailed debugging information, such as message responses from servers.The default value is Normal.
6. Click **Save**.

## Running a file ingestion task

You can run a file ingestion task in the following ways:

- To run a file ingestion task manually, on the **Explore** page, navigate to the task. In the row that contains the task, click **Actions** and select **Run**.  
Alternatively, you can run the task manually from the **Task Details** page. To access the **Task Details** page, click **Actions** and select **View**. In the **Task Details** page, select **Run**.
- To run a file ingestion task on a schedule, edit the task in the file ingestion task wizard to associate the task with a schedule.

# Aborting a file ingestion job

You can abort a file ingestion job that is in the **Up and Running** state.

To abort a job, open the **My Jobs** page, and select **Abort** from the **Actions** menu for the job.

## Key ring command reference

A file ingestion task encrypts and decrypts files using the Pretty Good Program (PGP) method. An Informatica Intelligent Cloud Services administrator uses the command line interface (CLI) to create key IDs and key passphrases. The administrator can then share them with the Informatica Intelligent Cloud Services user to encrypt and decrypt files.

You can run the key ring commands if you have the privileges to update files in the agent location. A PGP configuration file is created when you install the agent. The PGP configuration file consists of the properties that lists the location of the public key ring and the secret key ring. You must update the properties to change the location of the existing key ring. For more information about updating the properties, see the *Administrator* help.

The default location of the PGP configuration file is `<agent location>/apps/MassIngestionRuntime/<latest version no>/conf/pgp-configuration.properties`.

Use the `createKeyRing` command to create a key ring in the key ring location that is defined in the PGP configuration file.

To create key IDs and add them to the key ring, use the `createKeyPair` command. A key ID consists of a public key and a private key. To import public keys from different partners or use an existing key pair and import it to the current agent key ring location, use the `importKeys` command.

### createKeyRing

Creates a key ring. A key ring consists of a public key ring and a secret key ring.

If the key ring exists, the command displays an error indicating that a key ring already exists.

The `createKeyRing` command uses the following syntax:

```
<--command|-c> createKeyRing
```

The following sample command creates a key ring and saves the key ring in the location that is defined in the PGP configuration file:

```
./pgp_cli.sh -c createKeyRing
```

The command displays the following output:

```
KeyRing created successfully
```

### createKeyPair

Creates a key pair. The key pair or the key ID consists of a public key and a private keys.

The `createKeyPair` command uses the following syntax:

```
<--command|-c> createKeyPair
<--name|-n> key_name
<--passphrase|-p> passphrase
```

```
[<--size|-s> size]
<--expiration|-e> expiration_date>
<--email|-m> email
```

The following table describes createKeyPair options and arguments:

Option	Argument	Description
--name -n	key_name	Required. The name of the key pair.
--passphrase -p	passphrase	Required. The passphrase of the PGP key.
--size -s	size	Optional. The size of the PGP key in bits. Enter one of the following values: - 512 - 1024 - 2048 - 4096 Default is 512.
--expiration -e	expiration_date	Required. The date when the PGP key pair expires. Use the following date format: dd-mm-yyyy
--email -m	email	Required. The email ID of the user.

**Note:** The type argument uses the RSA PGP key.

The following sample command creates a key pair and adds the key pair to the key ring.

```
./pgp_cli.sh -c createKeyPair -n Mykeypair -p Mykeypassphrase -s 1024 -e 10-12-2019 -m
abc@informatica.com
```

The command displays the following output:

```
12:09:2017 INFO Key pair was successfully created and added to your key ring. The key
ID is '0x23149FC8C38658EA'.
12:09:2017 INFO Key Pair created successfully.
```

## listKeys

Lists all keys in key ring.

The listKeys command uses the following syntax:

```
<--command|-c> listKeys
```

The following sample command lists keys that are in the key ring:

```
./pgp_cli.sh -c listKeys
```

The command displays the following output:

```
12:10:38 INFO Default system locale: English (United States)
12:10:38 INFO Listing Keys.
12:10:38 INFO Total keys : 2
Key ID : 0x23149FC8C38658EA User : Mykeypair <abc@informatica.com./pgp_cli.sh>
Description : Key Pair Key Type : RSA Key Size : 1024 Expiration Date : Tue Dec 10
23:59:59 IST 2019
```

Key ID : 0x7B1E52AFB29030A6 User : new <a@b.com> Description : Key Pair Key Type : RSA  
Key Size : 1024 Expiration Date : Sat Sep 28 23:59:59 IST 2019

## importKeys

Imports keys from an external file to the key ring.

To import public keys from an external file or to use an existing key pair and import it to the current agent key ring location, use the importKeys command.

The importKeys command uses the following syntax:

```
<--command|-c> importKeys <--location|-l> location
```

The following table describes importKeys options and arguments:

Option	Argument	Description
-location -l	location	Required. The file name and location of the file that contains key pairs or public keys to import.

The following sample command imports keys from the key pair to the key ring:

```
./pgp_cli.sh -c importKeys -l /root/RSFiles/SubFolder1/SubFolder2/file1.asc
```

The command displays the following output:

```
12:37:09 INFO Default system locale: English (United States)
12:37:10 INFO Importing Keys.
12:37:10 INFO Public key '0x23149FC8C38658EA' with user ID 'doctest
<abc@informatica.com./pgp_cli.sh>' was imported successfully.
12:37:10 INFO 1 public keys and 0 secret keys were successfully imported into your key
ring.
12:37:10 INFO Import Finished.
```

## exportKeyPairs

Exports key pairs from the key ring to a file.

The exportKeyPairs command uses the following syntax:

```
<--command|-c> exportKeyPairs
<--ids|-i> list_of_key_ids
<--location|-l> location
```

The following table describes exportKeyPairs options and arguments:

Option	Argument	Description
--ids -i	list_of_key_ids	Required. Comma-separated list of key IDs in the key ring.
--location -l	location	Required. The file name and location of the file to export key pairs from the key ring.

The following sample command exports key pairs from the key ring to a local repository:

```
./pgp_cli.sh -c exportKeyPairs -i 0x23149FC8C38658EA -l /root/RSFiles/SubFolder1/file.asc
```



The command displays the following output:

```
12:28:18 INFO Default system locale: English (United States)
12:28:18 INFO Exporting Key Pairs.
12:28:18 INFO Export Finished.
```

## exportPublicKeys

Exports public keys from the key ring to a file.

The exportPublicKeys command uses the following syntax:

```
<--command|-c> exportPublicKeys
<--ids|-i> list_of_key_ids
<--location|-l> location
```

The following table describes exportPublicKeys options and arguments:

Option	Argument	Description
--ids -i	list_of_key_ids	Required. Comma-separated list of PGP key IDs in the key ring.
--location -l	location	The file name and location file to export public key from the key ring.

The following sample command exports public keys to a local repository:

```
./pgp_cli.sh -c exportPublicKeys -i 0x23149FC8C38658EA -l /root/RSFiles/SubFolder1/
SubFolder2/file1.asc
```

The command displays the following output:

```
12:32:10 INFO Default system locale: English (United States)
12:32:10 INFO Exporting Public Keys.
12:32:10 INFO Export Finished.
```

## deleteKeys

Deletes keys from the key ring.

The deleteKeys command uses the following syntax:

```
<--command|-c> deleteKeys <--ids|-i> list_of_key_ids
```

The following table describes deleteKeys options and arguments:

Option	Argument	Description
--ids -i	list_of_key_ids	Required. Comma-separated list of key IDs in the key ring.

The following sample command deletes keys:

```
./pgp_cli.sh -c deleteKeys -i 0x23149FC8C38658EA
```

The command displays the following output:

```
12:36:46 INFO Default system locale: English (United States)
12:36:46 INFO Deleting Key.
12:36:47 INFO Key '0x23149FC8C38658EA' was deleted
12:36:47 INFO Delete Finished.
```

## changePassphrase

Changes the passphrase of the key.

The changePassphrase command uses the following syntax:

```
<--command|-c> changePassphrase  
<--ids|-i> key_id  
<--old-passphrase|-o> old_passphrase  
<--passphrase|-p> new_passphrase
```

The following table describes changePassphrase options and arguments:

Option	Argument	Description
--ids -i	key_id	Required. Comma-separated list of PGP key IDs in the key ring.
--old-passphrase -o	old_passphrase	Required. The old passphrase of the PGP key ring.
--passphrase -p	new_passphrase	Required. The new passphrase of the PGP key ring.

The following sample command replaces the old key passphrase to the new key passphrase:

```
./pgp_cli.sh -c changePassphrase -i 0xDA70CEEDF703DCBE -o Mykeypassphrase -p  
Mynewkeypassphrase
```

The command displays the following output:

```
12:46:36 INFO Default system locale: English (United States)  
12:46:36 WARN Unable to load pgp configuration file : ./conf/pgp-  
configuration.properties (No such file or directory)  
12:46:36 INFO Changing Key Pair.  
12:46:36 INFO Passphrase for the key '0xDA70CEEDF703DCBE' was changed successfully.  
Please make sure to save this passphrase in a secure place.  
12:46:36 INFO Key Passphrase changed successfully.
```

## CHAPTER 7

# Mass Ingestion Streaming

Mass Ingestion Streaming is a separately licensed ingestion type of the Mass Ingestion service. Mass Ingestion Streaming can ingest data at scale from any streaming data sources, such as logs, clickstream, social media, and IoT sources. Use Mass Ingestion Streaming to ingest high-volume, real-time data from streaming sources to on-premises and cloud storage. You can also track and monitor the progress of the ingestion.

To gather operational intelligence from streaming data or to perform real-time data warehousing, you need to collect and analyze the data before it becomes obsolete or corrupted. Use Mass Ingestion Streaming to combine or separate data from streaming sources in real time. You can apply simple transformations on the data to ensure the data ingested is ready for analytics.

The Mass Ingestion service has an easy-to-use interface that runs in Informatica Intelligent Cloud Services. Use the Mass Ingestion Streaming service to define, deploy, undeploy, and monitor ingestion jobs. A job is an executable instance of an ingestion task. You can collect streaming and IoT data from different sources, apply simple transformations to the data, and then ingest the data to different types of targets. You can ingest data from sources, such as Amazon Kinesis, event logs, Google PubSub, JMS, Kafka, MQTT, OPC UA, and REST V2. You can stream data to targets, such as Amazon Kinesis, Amazon S3, Azure Event Hubs, Databricks Delta, Google Cloud Storage, Google PubSub, Kafka, and Microsoft Azure Data Lake Storage.

## Use cases

Mass Ingestion Streaming can help you fulfill multiple usage requirements.

Consider using Streaming ingestion in the following scenarios:

- **Real-time analytics.** Ingest streaming and IoT data into messaging systems, such as Apache Kafka or Amazon Kinesis, for real-time analytics. Real-time analytics can help companies identify business opportunities and revenue streams that can result in increased profits and improved customer service.
- **Data integration.** Ingest streaming and IoT data into cloud data lakes, such as Amazon S3, for integrating data in real-time to provide up-to-the-minute information.

# Mass Ingestion Streaming sources

You can ingest high volume, real-time data from supported streaming sources to on-premises and cloud targets that Mass Ingestion Streaming supports. You can ingest data in the form of events or messages.

Mass Ingestion Streaming supports the following data sources:

- Amazon Kinesis Streams
- AMQP
- Azure Event Hubs Kafka
- Flat File
- Google PubSub
- JMS
- Kafka
  - Apache Kafka
  - Confluent Kafka
  - Amazon Managed Streaming (Amazon MSK)
- MQTT
- OPC UA
- REST V2

To determine the connectors to use for these source types, see *Connectors and Connections > Mass Ingestion Streaming connectors*.

## Amazon Kinesis Streams sources

Use a Kinesis Streams source to read data from an Amazon Kinesis Stream. To create a Kinesis Streams source connection, use the Kinesis connection type.

Kinesis Streams is a real-time data stream processing service that Amazon Kinesis offers within the AWS ecosystem. Kinesis Streams is a customizable option that you can use to build custom applications to process and analyze streaming data. As Kinesis Streams cannot automatically scale to meet data in-flow demand, you must manually provision enough capacity to meet system needs.

Before you use a Kinesis stream source, perform the following tasks:

1. In the Amazon Kinesis console, create and configure an Amazon Kinesis stream.
2. In the Amazon Web Services (AWS) Identity and Access Management (IAM) service, create a user.
3. Download the access key and secret access key that are generated during the user creation process.
4. Associate the user with a group that has permissions to write to the Kinesis stream.

Mass Ingestion Streaming does not support profile based and cross account authentication. Amazon Web Services credentials used for Amazon Kinesis must have permissions to access to Amazon DynamoDB and Amazon CloudWatch services.

## AMQP sources

Use an Advanced Message Queuing Protocol (AMQP) source to read messages from an AMQP message queue. To create an AMQP source connection, use the AMQP connection type.

AMQP is a message-oriented standard with queuing, routing, reliability and security features. AMQP is a wire-level, platform-agnostic protocol that you can use to facilitate business transactions by passing real-time message streams.

Using the AMQP connector, you can read messages from AMQP brokers, monitor a message queue, and handle subscribe patterns for brokered messaging. The streaming ingestion task uses RabbitMQ as the AMQP broker. RabbitMQ is a distributed message broker system that is fast, scalable, and durable. RabbitMQ uses AMQP 0-9-1 messaging protocol for the secure transfer of messages.

In a streaming ingestion task, you can use an AMQP source to subscribe to a stream of incoming messages. The AMQP broker stores the messages in the message queue until the streaming ingestion job receives the message off the queue. When the streaming ingestion job receives a message, the job acknowledges the receipt of the message. The acknowledged message is then removed from the message queue.

You can use the AMQP source when you have long-running tasks that you want to run as reliable background jobs. You can also choose to use an AMQP source for communication between applications where one part of the system needs to notify another part, such as order handling in a webshop.

## Azure Event Hubs Kafka sources

You can configure a Kafka source to connect to Azure Event Hubs. To create an Azure Event Hubs Kafka source connection, use the Kafka connection type.

When you create a standard or dedicated tier Event Hubs namespace, the Kafka endpoint for the namespace is enabled by default. You can then use the Azure Event Hubs enabled Kafka connection as a source connection while configuring a streaming ingestion task. Enter the Event Hubs name as the topic name.

The Azure Event Hubs source information that you enter while configuring a streaming ingestion task is same as that of a normal Kafka source configuration. For more information about Azure Event Hubs Kafka source properties, see [“ Azure Event Hubs Kafka source properties” on page 467](#).

Configure the following properties while creating a Kafka connection in Administrator:

- **Kafka Broker List:** `NAMESPACENAME.servicebus.windows.net:9093`
- **Additional Connection Properties:**  
`security.protocol=SASL_SSL,sasl.mechanism=PLAIN,sasl.kerberos.service.name=Kafka`
- **SSL Mode:** One-Way
- **SSL TrustStore File Path:** Path to a trusted root cert on your file system. For example,  
`<AGENT_HOME>/jdk/jre/lib/security/cacerts`
- **SSL TrustStore Password:** Truststore password.
- **Additional Security Properties:** `sasl.jaas.config=org.apache.kafka.common.security.plain.PlainLoginModule required username="$ConnectionString" password="Endpoint=sb://mynamespace.servicebus.windows.net/;SharedAccessKeyName=XXXXXX;SharedAccessKey=XXXXXX";`

For more information about creating an Azure Event Hubs Kafka source connection, see the *Connections* help.

**Note:** Event Hubs for Kafka is available only on standard and dedicated tiers. The basic tier doesn't support Kafka on Event Hubs.

## Flat File sources

Use a flat file as a source to read incoming real-time data. Configure a flat file connection to read data from flat files that are stored in the same directory.

A streaming ingestion task reads each row in a flat file source and ingests the data to a configured target. When a flat file is continuously updated in real time, the streaming ingestion task reads only the newly added content instead of reading the complete file again.

Streaming ingestion can read data from the delimited flat files. The delimiter character must be a carriage return (`\r`), a line feed (`\n`), or a combination of both.

## Google PubSub sources

Use a Google PubSub source to read messages from the configured Google Cloud PubSub subscription. To create a Google PubSub source connection, use the Google PubSub connection type.

Google PubSub is an asynchronous messaging service that decouples services that produce events from services that process events. You can use Google PubSub as a messaging-oriented middleware or for event ingestion and delivery for streaming analytics pipelines. Google PubSub offers durable message storage and real-time message delivery with high availability and consistent performance at scale. You can run Google PubSub servers in all the available Google Cloud regions around the world.

Before you use Google PubSub connector, you must ensure that you meet the following prerequisites:

- Your organization has the Google PubSub Connector license.
- You have a Google service account JSON key to access Google PubSub.
- You have the `client_email`, `client_id`, and `private_key` values for the Google service account. You need these details when you create a Google PubSub connection in Administrator.

In a streaming ingestion task, you can use a Google PubSub source to subscribe to messages from a Google PubSub topic.

## JMS sources

Use a JMS source to read data from a JMS provider. To create a JMS source connection, use the JMS connection type.

JMS providers are message-oriented middleware systems that send JMS messages. The JMS source reads JMS messages either from a JMS provider message queue or from a JMS provider based on the message topic.

The JMS source can read the following JMS message types:

- `Message`. Contains only header and properties fields
- `TextMessage`. Contains a string object. `TextMessages` can contain XML or JSON message data.
- `BytesMessage`. A stream of uninterpreted bytes. Use a `BytesMessage` for encoding a message body to match an existing message format. `BytesMessages` generally do not include property fields.
- `MapMessage`. Contains a set of name or value pairs. The names are in string format. The values are of Java primitive datatypes.

### JMS message delivery destination types

You can choose one of the following JMS message delivery destination types:

- **Queue.** The JMS message producer delivers messages to a single consumer. The consumer must be registered to consume messages from the queue. If no consumers are registered to the queue, the queue retains the messages until a consumer registers to it.
- **Topic.** The JMS message producer delivers messages to all active consumers who subscribe to the topic. Several producers can send messages to the topic destination, and each message can be delivered to several subscribers. If no consumers are registered to the topic, the topic doesn't retain the message. You can make the subscription sharable, durable or both. A sharable subscription enables one or more consumers to access a single subscription. A durable subscription retains the message for inactive subscribers until subscribers consume the message or until the message expires.

## Kafka sources

Use a Kafka source to read messages from a Kafka topic. To create a Kafka source connection, use the Kafka connection type.

Kafka is a publish-subscribe messaging system. It is an open-source distributed streaming platform that persists the streaming data in a Kafka topic. Any topic can then be read by any number of systems that need data in real-time. Kafka can serve as an interim staging area for streaming data that can be consumed by different downstream consumer applications can consume.

Kafka runs as a cluster comprised of one or more servers each of which is called a broker. Kafka brokers stream data in the form of messages. These messages are published to a topic. When you create a Kafka source, you create a Kafka consumer to read messages from a Kafka topic.

In a streaming ingestion task, you can use a Kafka source to subscribe to a stream of incoming data. When you configure a Kafka source to read from a Kafka topic, you can specify the topic name or use a Java supported regular expression to subscribe to all topics that match a specified pattern.

You can use the same Kafka connection to create an Amazon Managed Streaming for Apache Kafka (Amazon MSK) or a Confluent Kafka source connection. You can then use the Amazon MSK source or the Confluent Kafka source in a streaming ingestion task to read messages from an Apache Kafka or a Confluent Kafka topic.

## MQTT sources

Use an MQTT source to read data from an MQ Telemetry Transport (MQTT) broker. To create an MQTT source, use the MQTT connection type.

MQTT is a publish-subscribe messaging system. It is a simple, lightweight, and persistent messaging protocol. It is designed for constrained devices and low-bandwidth, high-latency, or unreliable networks. Both publishers and subscribers are MQTT clients. MQTT decouples the publisher from the subscriber, so a broker manages the client connections.

An MQTT broker receives all messages, filters the messages, determines which client subscribed to each message, and then sends messages to the subscribed clients. If multiple MQTT sources connect to one MQTT broker, each connection must have a unique identifier. When you run a streaming ingestion job to ingest data from an MQTT source, Streaming Ingestion first writes the data to an internal queue before writing the data to a target.

**Note:** An MQTT source must have a unique client identifier. If two MQTT sources have the same client identifier, the MQTT broker rejects both the clients and the streaming ingestion job gets into *running with warning* state.

Mass Ingestion Streaming supports MQTT Quality of Service (QoS) level 1. Level 1 indicates that the client sends the message to the broker at least once, but the message might be delivered more than once. After the broker acknowledges the message receipt, the client deletes the message from the outbound queue. The QoS Level is restricted to client to broker or broker to client communication.

## OPC UA sources

Use an OPC UA source to read messages from an OPC UA application tag. To create an OPC UA source connection, use the OPCUA connection type.

Open Platform Communications (OPC) is one of the important communication protocols for Industry 4.0 and the IIoT (Industrial Internet Of Things). OPC Unified Architecture (OPC UA) is a machine-to-machine communication protocol used for industrial automation. OPC UA provides a flexible and adaptable mechanism to move data between enterprise systems, monitoring devices, and sensors that interact with real-world data. You can use OPC UA to establish communication for simple downtime status or for massive amounts of highly complex plant-wide information.

The OPC UA source is a client that collects data from OPC servers. Data points in OPC are tags that represent data from devices and provide real-time access to data. In a streaming ingestion task, you can create an OPC UA source to read the incoming data based on the list of tags that you provide. You must mention the tags in a JSON array format.

## REST V2 sources

Use a REST V2 source to read data from a web service application. To create a REST V2 source connection, use the REST V2 connection type.

REST V2 source connector is a generic connector for cloud applications with REST API. It supports swagger specification version 2.0. The swagger specification file contains operation ID, path parameters, query parameters, header fields, and payload details.

When you can create a REST V2 source connection in Administrator, for a streaming ingestion task, you can configure one of the following REST authentication types:

- Basic
- OAuth1.0
- OAuth2.0 Client Credentials
- OAuth2.0 Authorization Code
- JWT bearer token authentication

## Mass Ingestion Streaming targets

You can ingest streaming data from a supported source to any on-premises and cloud targets that Mass Ingestion Streaming supports.

Mass Ingestion Streaming supports the following targets:

- Amazon Kinesis Data Firehose
- Amazon Kinesis Streams
- Amazon S3



- Databricks Delta
- Flat file
- Google Cloud Storage
- Google PubSub
- JDBC V2
- Kafka
  - Apache Kafka
  - Confluent Kafka
  - Amazon Managed Streaming (Amazon MSK)
- Microsoft Azure Data Lake Store Gen2
- Microsoft Azure Event Hubs

To determine the connectors to use for these target types, see *Connectors and Connections > Mass Ingestion Streaming connectors*.

## Amazon Kinesis Data Firehose target

Use a Kinesis target to receive data from a source and write the data to an Amazon Kinesis Data Firehose target. To create a Kinesis target, use the Amazon Kinesis connection type.

Kinesis Firehose is a real-time data stream processing service that Amazon Kinesis offers within the AWS ecosystem. Use Kinesis Firehose to batch, encrypt, and compress data. Kinesis Firehose can automatically scale to meet system needs.

To configure access for Kinesis Firehose as a target, perform the following tasks:

- Create an AWS account with the required IAM permissions for the IAM user to use the AWS Kinesis Data Firehose service.
- Define a Firehose delivery stream. Configure source as Direct PUT or other sources.
- Grant required permissions to the IAM user credentials based on the target the user is writing to. For a list of permissions, see the AWS documentation at <https://docs.aws.amazon.com/firehose/latest/dev/controlling-access.html#access-to-firehose>.

## Amazon Kinesis Streams target

Use a Kinesis target to receive data from source services and write the data to an Amazon Kinesis Stream. To create a Kinesis target, use the Amazon Kinesis connection type.

Kinesis Streams is a real-time data stream processing service that Amazon Kinesis offers within the AWS ecosystem. Kinesis Streams is a customizable option that you can use to build custom applications to process and analyze streaming data. As Kinesis Streams cannot automatically scale to meet data in-flow demand, you must manually provision enough capacity to meet system needs.

Before you use a Kinesis stream target, perform the following tasks:

1. In the Amazon Kinesis console, create and configure an Amazon Kinesis stream.
2. In the Amazon Web Services (AWS) Identity and Access Management (IAM) service, create a user.
3. Download the access key and secret access key that are generated during the user creation process.
4. Associate the user with a group that has permissions to write to the Kinesis stream.

## Amazon S3 target

Use an Amazon S3 V2 connector to write the streaming data to an Amazon S3 target.

Amazon Simple Storage Service (Amazon S3) is storage service in which you can copy data from a streaming source and simultaneously move data to any target. You can use Amazon S3 to transfer the data from a list of configured source connections to an Amazon S3 target. You can accomplish these tasks by using the AWS Management Console web interface.

You can use an Amazon S3 object as a target in a streaming ingestion task. You can configure the Amazon S3 target and advanced properties for a target object.

## Databricks Delta target

Use a streaming ingestion task to write data to a Databricks Delta target. To create a Databricks Delta target, use the Databricks Delta connection type. The Databricks Delta target requires a Databricks cluster version 6.3 or later.

Databricks Delta is an open source storage layer that provides ACID transactions and works on top of existing data lakes. Databricks uses proprietary Delta software to manage stored data and allow fast access to the data.

You can access Delta Lake tables built on top of the following storage types:

- Azure Data Lake Storage (ADLS) Gen2
- Amazon Web Services (AWS) S3

The Databricks Delta target writes data to one or more Delta Lake tables on Databricks. You can use the Databricks Delta target in a streaming ingestion task for the following use cases:

- Ingest bulk data from all streaming sources into Databricks Delta tables
- Merge change data capture (CDC) from all streaming sources and write to Databricks Delta tables

The Databricks Delta connection uses a JDBC URL to connect to the Databricks cluster. When you configure the target, you specify the JDBC URL and credentials to use to connect to the cluster. You also define the connection information that the target uses to connect to the staging location in Amazon S3 or Azure Data Lake Storage Gen2.

You specify the tables in Delta Lake to which you want to write the data. The target writes data from record fields to table columns based on matching names.

## Flat file target

Use a streaming ingestion task to write data from various sources to a flat file target. The task writes real-time streaming data from various sources to the file system of the Secure Agent that runs the dataflow.

A streaming ingestion task writes the data to the staging directory with the file name that you provide. When the task adds new content to the file, it precedes it with a New Line character (\n) in the target.

The flat file target performs the file rollover action and you can configure the rollover properties. The file rollover process closes the current file and creates a new file on the basis of the file size, event count, or time. To configure rollover, specify the **Rollover Size**, **Rollover Events Count**, or the **Rollover Time** properties of the target service. The rollover process moves the file from the staging directory to the target and renames the file. The file name format of the renamed file is the original file name with an addition of the time stamp and counter information (yyyy\_mm\_dd-hh\_mm\_ss\_counter). For example, during rollover, the file streaming.txt is renamed to streaming-2021\_08\_16-17\_17\_30\_4.txt.

You can implement a combination of the rollover properties. For example, if you set the rollover events count to 1000, the rollover size to 1 GB, and the rollover time to 1 hour, the task rolls the file over when the file reaches a size of 1 GB even if the 1000 events are not accumulated and the 1-hour period has not elapsed.

## Google Cloud Storage V2 target

Use a streaming ingestion task to write data to a Google Cloud Storage target. To create a Google Cloud Storage target, use the Google Cloud Storage V2 connection type.

You can use Google Cloud Storage to stream multimedia, store custom data analytics pipelines, or distribute large data objects to users through direct download. You can write data to Google Cloud Storage for data backup. In the event of a database failure, you can read the data from Google Cloud Storage and restore it to the database.

Google Cloud Storage offers different storage classes based on factors such as data availability, latency, and price. Google Cloud Storage has the following components:

- **Projects.** In Google Cloud Storage, all resources are stored within a project. Project is a top-level container that stores billing details and user details. You can create multiple projects. A project has a unique project name, project ID, and project number.
- **Buckets.** Each bucket acts like a container that stores data. You can use buckets to organize and access data. You can create more than one bucket but you cannot nest buckets. You can create multiple folders within a bucket and you can also nest folders. You can define access control lists to manage objects and buckets. An access control list consists of permission and scope entries. Permission defines the access to perform a read or write operation. Scope defines a user or a group who can perform the operation.
- **Objects.** Objects comprise the data that you upload to Google Cloud Storage. You can create objects in a bucket. Objects consist of object data and object metadata components. The object data is a file that you store in Google Cloud Storage. The object metadata is a collection of name-value pairs that describe object qualities.

Before you use Google Cloud Storage V2 Connector, you must complete the following prerequisite tasks:

1. Ensure that you have a Google service account to access Google Cloud Storage.
2. Ensure that you have the `client_email`, `project_id`, and `private_key` values for the service account. You will need to enter these details when you create a Google Cloud Storage connection in the Administrator.
3. Ensure that you have enabled the Google Cloud Storage JSON API for your service account. Google Cloud Storage V2 Connector uses the Google API to integrate with Google Cloud Storage.
4. Verify that you have write access to the Google Cloud Storage bucket that contains the target file.
5. Ensure that you have enabled a license to use a Cloudera CDH or Hortonworks HDP package in your organization.

When you deploy a streaming ingestion task, the Secure Agent uses the Google Cloud Storage API to perform the specified operation and writes data to Google Cloud Storage files. You can write data into a Google Cloud Storage target. You cannot perform update, upsert, or delete operations on a Google Cloud Storage target.

## Google PubSub target

Use a streaming ingestion task to write data to a Google PubSub topic. To create a Google PubSub target, use the Google PubSub connection type.

Google PubSub is an asynchronous messaging service that decouples services that produce events from services that process events. You can use Google PubSub as a messaging-oriented middleware or for event ingestion and delivery for streaming analytics pipelines. Google PubSub offers durable message storage and real-time message delivery with high availability and consistent performance at scale. You can run Google PubSub servers in all the available Google Cloud regions around the world.

Before you use Google PubSub connector, you must ensure that you meet the following prerequisites:

- Your organization has the Google PubSub Connector license.
- You have a Google service account JSON key to access Google PubSub.
- You have the `client_email`, `client_id`, and `private_key` values for the Google service account. You need these details when you create a Google PubSub connection in Administrator.

In a streaming ingestion task, you can use a Google PubSub target to publish messages to a Google PubSub topic.

## JDBC V2 target

Use a streaming ingestion task to write data to a database target. To create a JDBC V2 target, use the JDBC V2 connection type.

You can use a JDBC V2 target to write data to a database table. The target consumes data in JSON format. The target ignores fields that don't map to the table columns of the database.

Consider the following prerequisites before you configure JDBC V2 as a target:

- A table to write data must exist before deploying the streaming ingestion task.
- Copy the database driver files to the following directory:  
`<Secure Agent installation directory>/apps/Streaming_Ingestion_Agent/ext`

**Note:** The JDBC V2 target accepts data only in a simple JSON or an array of a simple JSON format.

## Kafka target

Use a streaming ingestion task to write data to a Kafka target. To create a Kafka target, use the Kafka connection type.

Kafka is a publish-subscribe messaging system. It is an open-source distributed streaming platform. This platform allows systems that generate data to persist their data in real-time in a Kafka topic. Any topic can then be read by any number of systems who need that data in real-time. Kafka can serve as an interim staging area for streaming data that can be consumed by different downstream consumer applications.

Kafka runs as a cluster that comprises of one or more Kafka brokers. Kafka brokers stream data in the form of messages, publishes the messages to a topic, subscribes the messages from a topic, and then writes it to the Kafka target.

When you create a Kafka target, you create a Kafka producer to write Kafka messages. You can use each Kafka target in a streaming ingestion job that writes streaming Kafka messages. When you configure a Kafka target, specify the topic to publish the messages and the IP address and port on which the Kafka broker runs. If a Kafka topic does not exist in the target, instead of manually creating topics you can also configure your brokers to auto-create topics when a non-existent topic is published to.

You can use the same Kafka connection to create an Amazon Managed Streaming for Apache Kafka (Amazon MSK) or a Confluent Kafka connection. You can then use the Amazon MSK or the Confluent Kafka target in a streaming ingestion task to write messages to an Apache Kafka or a Confluent Kafka target.

## Microsoft Azure Data Lake Storage Gen2 target

Use a streaming ingestion task to write data to a Microsoft Azure Data Lake Storage Gen2 target. To create a Microsoft Azure Data Lake Storage Gen2 target, use the Microsoft Azure Data Lake Storage Gen2 connection type.

Microsoft Azure Data Lake Storage Gen2 is a next-generation data lake solution for big data analytics. You can store data in the form of directories and sub-directories, making it efficient for data access and manipulation. You can store data of any size, structure, and format. You can process large volumes of data to achieve faster business outcomes. Data scientists and data analysts can use data in the data lake to find out specific patterns before you move the analyzed data to a data warehouse. You can use big data analytics available on top of Microsoft Azure Blob storage.

The streaming ingestion task writes data to Microsoft Azure Data Lake Storage Gen2 based on the specified conditions.

For more information about Microsoft Azure Data Lake Storage Gen2, see the Microsoft Azure Data Lake Storage Gen2 documentation.

## Microsoft Azure Event Hubs target

Use a streaming ingestion task to write data to an Azure Event Hubs target. To create an Azure Event Hubs target, use the Azure Event Hubs connection type.

Azure Event Hubs is a highly scalable data streaming platform and event ingestion service that receives and processes events. Azure Event Hubs can ingest and process large volumes of events with low latency and high reliability. It is a managed service that can handle message streams from a wide range of connected devices and systems.

Any entity that sends data to an event hub is an event publisher. Event publishers can publish events using HTTPS or Kafka 1.0 and later. Event publishers use a Shared Access Signature (SAS) token to identify themselves to an event hub and can have a unique identity or use a common SAS token.

For more information about Event Hubs, see the Microsoft Azure Event Hubs documentation.

# Transformations in Mass Ingestion Streaming

Transformations are part of a streaming ingestion task. Transformations represent the operations that you want to perform when ingesting streaming data.

Each transformation performs a specific function. For example, a Filter transformation filters data from the ingested data based on a specified condition.

When you create a streaming ingestion task, adding a transformation is optional. Each transformation type has a unique set of options that you can configure.

You can use the following transformations in streaming ingestion tasks:

- Combiner
- Filter
- Python
- Splitter
- Format Converter

You can add multiple transformations to a streaming ingestion task. In such a case, the order of transformations is important because the source data undergoes each transformation in the given order. The output of one transformation becomes the input to the next one in the task flow.

In a streaming ingestion task, you can add only one Combiner transformation and one Format Converter transformation. The Format Converter transformation must be the last transformation in the task flow. If the task includes both a Combiner transformation and a Format Converter transformation, the Format Converter transformation must be the last transformation in the task flow, preceded by the Combiner transformation.

## Supported data formats

Each supported data format of the incoming streaming data supports only a specified transformation type.

Streaming ingestion transformations can process streaming data in the following formats:

- Binary. Any type of structured and unstructured data.
- JSON. Minimal, readable format for structuring data.
- XML. Structured text data.

If no transformation is configured for a task, the incoming data is consumed in its original format.

## Combiner transformation

A Combiner transformation combines multiple events from a streaming source into a single event based on the specified conditions.

A Combiner transformation processes binary data and JSON data. For JSON message formats, the Combiner transformation combines the incoming data into an array of data and returns JSON array objects as output. For binary message formats, it combines the incoming data based on the specified conditions.

In a streaming ingestion task, you can add only one Combiner transformation. If the task includes both a Combiner transformation and a Format Converter transformation, the Format Converter transformation must be the last transformation in the task flow, preceded by the Combiner transformation. If the task doesn't include a Format Converter transformation, the Combiner transformation must be the last transformation in the task flow.

You can use one of the following conditions for a Combiner transformation:

- Minimum number of events
- Maximum aggregate size
- Time limit

For example, consider the following events:

- Record created
- Record published

If you use comma (,) as a delimiter, the Combiner transformation returns the following combined event:

Record created,Record Published

**Note:** When you process binary data with a Combiner transformation, you cannot use a regular expression as a delimiter.

## Filter transformation

The Filter transformation filters data out of the incoming streaming events based on a specified filter condition.

You can filter data based on one or more conditions. For example, to work with data within a date range, you can create conditions to remove data based on the specified dates.

## Python transformation

A Python transformation runs the Python script to transform incoming data from a streaming source.

A Python transformation processes binary, JSON, and XML data. The Python transformation uses the two variables, *inputData* and *outputData* for storing the incoming data and outgoing data.

Incoming data of XML or JSON message formats are stored as string in the *inputData* variable and the outgoing data of XML or JSON message formats are stored as string in the *outputData*. Incoming data of binary message format are stored as `numpy.ndarray` in the *inputData* variable. Outgoing data of binary message format are stored as `bytearray` in the *outputData* variable.

Binary data in the *inputData* variable are encoded as ASCII characters. You must decode the data accordingly. Also ensure that the Python transformation script handles any non-ASCII characters present in the *inputData* variable.

Before using a Python transformation, create a directory, Python home to install Python. After installing Python in the Python home directory, ensure to install the third-party libraries, NumPy and Jep (Java Embedded Python) in the same directory as Python home.

In one Secure Agent, you cannot use two different versions of Python to run the same Python transformation.

### Sample Python scripts for JSON

```
import json
temp=json.loads(inputData)
temp["name"]="Mr "+temp["name"]
outputData=json.dumps(temp)
#####
inputData: { "name":"John", "age":30, "city":"New York"}
outputData: { "name":"Mr John", "age":30, "city":"New York"}
```

### Sample Python scripts for binary

```
temp = ''.join(str(chr(c)) for c in inputData)
temp += " - this is edited again text"
outputData = bytearray(temp, 'utf-8')
#####
inputData: Sample text
outputData: Sample text - this is edited again text
```

### Sample Python scripts for XML

```
import xml.etree.ElementTree as ET
myroot = ET.fromstring(inputData)
for x in myroot:
    if x.tag=="body":
        x.tag="Msg"
xmlstr = ET.tostring(myroot)
outputData=xmlstr.decode('utf-8')
#####
inputData: <note><to>You</to><from>Me</from><heading>Message</heading><body>Happy
Coding</body></note>
outputData: <note><to>You</to><from>Me</from><heading>Message</heading><Msg>Happy
Coding</Msg></note>
```

## Splitter transformation

A Splitter transformation splits multiline messages or message arrays into separate messages based on the conditions that you specify before ingesting them into targets.

The Splitter transformation splits binary, JSON, and XML messages based on the condition that you specify and passes the separated messages into new files before ingesting them into targets. Use the Splitter transformation to split complex messages into logical components. For example, if a message contains an error code and error message separated by a comma, you can use the comma to separate the code and message into different files.

### Binary messages

In binary message format, the Splitter transformation divides the messages based on line boundaries or byte sequence. The maximum number of lines determines the line boundaries. Each output split file contains no more than the configured number of lines or bytes. The default value for the line boundaries is 1. The default for the byte sequence is ','.

### JSON messages

In JSON message format, the Splitter transformation divides a JSON file into separate files based on the array element specified by a JSONPath expression. Each generated file is comprised of an element of the specified array. The generated file is transferred to the downstream target or transformation in the task. If the specified JSONPath is not found or does not evaluate to an array element, the original file is routed to *failure* and no files are generated. The default JSONPath Expression is '\$'.

### XML messages

In XML message format, the Splitter transformation splits an XML message into many files based on the level of input depth. Each of these files contain a child or descendant of the original file.

## Format Converter transformation

The Format Converter transformation converts the data format of XML and JSON incoming messages to Parquet format, based on the specified conditions, before streaming them into the data lake.

You can add only one Format Converter transformation to a streaming ingestion task. The Format Converter transformation must be the last transformation in the task flow.

You can specify the date, time, and timestamp format of incoming data. If the format is not specified, it is considered in milliseconds since the epoch (Midnight, January 1, 1970, GMT).

## Configuring a streaming ingestion task

When you create a streaming ingestion task, you define a source and a target. You can optionally define a transformation to transform the data. Use the task wizard to configure streaming ingestion tasks.

To configure a streaming ingestion task, perform the following steps:

1. Define the basic information of a task.
2. Configure a source.
3. Configure a target.
4. Optionally, add one or multiple transformations.
5. Optionally, set the runtime options.



As you work through the task wizard, you can click **Save** to save your work at any time after you configure the target. When you have completed the wizard, click **Save** to save the task.

Before you begin, verify that the conditions that are described in ["Before you begin" on page 465](#).

## Before you begin

Before you create streaming ingestion tasks, verify that the following conditions are met:

- Check that your organization has licenses for Mass Ingestion Streaming and the streaming ingestion packages.
- The Mass Ingestion Streaming is running on the Secure Agent.
- Source and target connections exist.

## Defining basic task information

Use the **Definition** tab in the streaming ingestion task wizard to define the task details.

1. Click **New > Streaming Ingestion Task**.  
The **Definition** page of the task wizard appears.
2. Configure the following properties:

Property	Description
Name	A name for the streaming ingestion task. The names of streaming ingestion tasks must be unique within the organization. Task names can contain alphanumeric characters, spaces, and underscores. Names must begin with an alphabetic character or underscore. Task names are not case-sensitive.
Location	The project folder to store the task.
Runtime Environment	Runtime environment that contains the Secure Agent. The Secure Agent runs the task.
Description	Optional. Description about the task. Maximum length is 4,000 characters.

3. Click **Next**

## Configuring a source

To configure a source, select a source connection from which you want to ingest streaming data and then configure source properties. Before you configure a source, ensure that the connection to the source is created in the Administrator service.

1. On the **Source** page, select a connection.  
The streaming ingestion task supports the following sources:
  - Amazon Kinesis Streams
  - AMQP

- Apache Kafka
- Flat file
- Google PubSub
- JMS
- MQTT
- OPC UA
- REST V2

The connection type populates automatically based on the connection that you select.

2. Based on the source that you select, enter the required details.

Options that appear on the **Source** tab of the task wizard vary based on the type of source that you select.

3. Under **Advanced Properties**, enter the required information.

4. Click **Next**.

The **Target** tab appears.

## Amazon Kinesis Streams source properties

The following table describes the Amazon Kinesis Streams source properties on the **Source** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the Amazon Kinesis Stream source connection.
Connection Type	The Amazon Kinesis connection type. The connection type populates automatically based on the connection that you select.
Stream	Name of the Kinesis Stream from which you want to read data.

The following table describes the advanced properties for Amazon Kinesis Streams sources in the **Source** tab when you define a streaming ingestion task:

Property	Description
Append GUID to DynamoDB table name	Specifies whether or not to add a GUID as a suffix to the Amazon DynamoDB table name. If disabled, you must enter the Amazon DynamoDB table name. Default is enabled.
Amazon DynamoDB	Amazon DynamoDB table name to store the checkpoint details of the Kinesis source data. The Amazon DynamoDB table name is generated automatically. However, if you enter a name of your choice, the streaming ingestion task prefixes the given name to the auto-generated name.

For more information about Kinesis Streams, see the Amazon Web Services documentation.

## AMQP source properties

The following table describes the Advanced Message Queuing Protocol (AMQP) source properties on the **Source** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the AMQP source connection.
Connection Type	The AMQP connection type. The connection type populates automatically based on the connection that you select.
Queue	Name of the existing AMQP queue from which the streaming ingestion task reads the messages. This queue is pre-defined by the AMQP administrator.
Auto Acknowledge messages	You can choose True or False. If you choose True, the AMQP broker automatically acknowledges the received messages.
Batch Size	The maximum number of messages that must be pulled in a single session. Default is 10 messages.

## Azure Event Hubs Kafka source properties

You can create a Kafka connection with Azure Event Hubs namespace.

When you create a standard or dedicated tier Event Hubs namespace, the Kafka endpoint for the namespace is enabled by default. You can then use the Azure Event Hubs enabled Kafka connection as a source connection while creating a streaming ingestion task. Enter the Event Hubs name as the topic name.

The following table describes the Kafka source properties on the **Source** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the Kafka source connection.
Connection Type	The Kafka connection type. The connection type populates automatically based on the connection that you select.
Topic	Name of the Event Hubs from which you want to read the events. You can either enter the topic name manually or fetch the already created metadata of the Kafka enabled Event Hubs connection. 1. Click <b>Select</b> . The <b>Select Source Object</b> dialog box appears showing all the available topics. 2. Select the required topic and click <b>OK</b> .

The following table describes the advanced properties for Kafka sources in the **Source** tab when you define a streaming ingestion task:

Property	Description
Consumer Configuration Properties	Comma-separated list of configuration properties for the consumer to connect to Kafka. Specify the values as key-value pairs. For example, <code>key1=value1, key2=value2</code> . The <code>group.id</code> property of Kafka consumer is autogenerated. You can override this property.

**Note:** Event Hubs for Kafka is available only on standard and dedicated tiers. The basic tier doesn't support Kafka on Event Hubs.

## Flat File source properties

The following table describes the flat file source properties on the **Source** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the flat file source connection.
Connection Type	The Flat file connection type. The connection type appears automatically based on the connection that you select.
Initial Start Position	Starting position from which the data is to be read in the file to tail. You can choose one of the following positions to start reading: <ul style="list-style-type: none"><li>- Beginning of File. Read from the beginning of the file to tail. Do not ingest any data that has already been rolled over.</li><li>- Current Time. Read from the most recently updated part of the file to tail. Do not ingest any data that has already been rolled over or any data in the file to tail that has already been written.</li></ul>
Tailing Mode	Tail a file or multiple files based on the logging pattern. You can choose one of the following modes: <ul style="list-style-type: none"><li>- Single File. Tail only one file.</li><li>- Multiple Files. Tail all the files indicated in the base directory. In this mode, you can enter a regular expression to indicate the files to tail.</li></ul>
File	Absolute path with the name of the file you want to read. Name of the file to tail or regular expression to find the files to tail. Enter the base directory for multiple files mode.

The following table describes the advanced properties that you can configure for flat file sources on the **Source** tab when you define a streaming ingestion task:

Connection Property	Description
Rolling Filename Pattern	Name pattern for the file that rolls over. If the file to tail rolls over, the file name pattern is used to identify files that have rolled over. The underlying streaming ingestion Secure Agent recognizes this file pattern. When the Secure Agent restarts, and the file has rolled over, it picks up from where it left off. You can use asterisk (*) and question mark (?) as wildcard characters to indicate that the files are rolled over in the same directory. For example, <code>\${filename}.log.*</code> . Here, asterisk (*) represents the successive version numbers that would be appended to the file name.

## Google PubSub source properties

The following table describes the Google PubSub source properties on the **Source** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the Google PubSub source connection.
Connection Type	The Google PubSub connection type. The connection type populates automatically based on the connection that you select.
Subscription	Name of the subscription on the Google PubSub service from which messages should be pulled. The Google PubSub connection supports only the pull delivery type for a subscription.
Batch Size	Maximum number of messages that the Cloud service bundles together in a batch. Default is 1.

## JMS source properties

The following table describes the JMS source properties on the **Source** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the JMS source connection.
Connection Type	JMS connection type. The connection type populates automatically based on the connection that you select.
Destination Type	Type of destination that the source service sends the JMS message to. You can choose one of the following destination types: <ul style="list-style-type: none"><li>- Queue. The JMS provider delivers messages to a single consumer who is registered for the queue.</li><li>- Topic. The JMS provider delivers messages to all active consumers who subscribe to the topic. When you use this destination type, multiple consumers can read the message.</li></ul> Default is <b>Queue</b> .
Shared Subscription	Enables multiple consumers to access a single subscription. Applies to the topic destination type. Default is false.
Durable Subscription	Enables inactive subscribers to retain messages and deliver retained messages when the subscribers reconnect. Applies to the topic destination type. Default is false.
Subscription Name	Name of the subscription. Applies to the topic destination type, when the topic subscription is sharable, durable, or both. If no value is specified, the ingestion service generates a unique subscription name.
JMS Destination	Name of the queue or topic that the JMS provider delivers the message to.

The following table describes the advanced properties for JMS sources in the **Source** tab when you define a streaming ingestion task:

Property	Description
Client ID	Optional. Unique identifier that identifies the JMS connection. The streaming ingestion task generates a unique client ID if a value isn't specified for an unshared durable subscription.

## Kafka source properties

When you define Kafka as the source of a streaming ingestion task, you must configure the mandatory Kafka source properties on the **Source** tab. Optionally, provide a comma-separated list of consumer configuration properties.

The following table describes the mandatory Kafka source properties:

Property	Description
Connection	Name of the Kafka source connection.
Connection Type	The Kafka connection type. The connection type populates automatically based on the connection that you select.
Topic	Kafka source topic name or a Java supported regular expression for the Kafka source topic name pattern to read the events from. You can either enter the topic name manually or fetch the metadata of the Kafka connection. To select the metadata of the Kafka connection perform the following actions: 1. Click <b>Select</b> . The <b>Select Source Object</b> dialog box appears, showing all the topics or topic patterns available in the Kafka broker. 2. Select the topic and click <b>OK</b> . <b>Note:</b> When you add a new Kafka source topic to a streaming ingestion job that is in <i>Up and Running</i> state, redeploy the job immediately to avoid data loss from the new topics.

### Consumer Configuration Properties

On the **Advanced Properties** section of the **Source** tab, in **Consumer Configuration Properties**, you can provide a comma-separated list of optional consumer configuration properties. Specify the values as key-value pairs.

The following table describes the consumer configuration properties that you can configure for Kafka sources:

Property	Description
group.id	Specifies the name of the consumer group the Kafka consumer belongs to. If <code>group.id</code> doesn't exist when you construct the Kafka consumer, the task creates the consumer group automatically. This property is auto-generated. You can override this property. Default is <code>key1=value1, key2=value2</code> .
auto.offset.reset	<p>Specifies the behavior of the consumer when there is no committed position or when an offset is out of range.</p> <p>You can use the following types of auto offset reset:</p> <ul style="list-style-type: none"> <li>- Earliest. Resets the offset position to the beginning of the topic.</li> <li>- Latest. Resets the offset position to the latest position of the topic.</li> <li>- None.</li> </ul> <p>When you read data from a Kafka topic or use a topic pattern and the offset of the last checkpoint is deleted during message recovery, provide the following property to recover the messages from the next available offset:</p> <pre>auto.offset.reset=earliest</pre> <p>Otherwise, the streaming ingestion task reads data from the latest offset available.</p>
message-demarcator	<p>Kafka source receives messages in batches. You can contain all Kafka messages in a single batch for a given topic and partition. This property allows you to provide a string to use as a demarcation for multiple Kafka messages. If you don't provide a value, each Kafka message is triggered as a single event.</p> <p>You can use the following delimiters as demarcators:</p> <ul style="list-style-type: none"> <li>- New line. Separates the new content with a new line feed. Enter the following value to use a new line as a message demarcator:  <pre>message-demarcator=\${literal('&amp;#10;'):unescapeXml() }</pre> </li> <li>- Comma. Separates the new content with a comma. Enter the following value to use a comma as a message demarcator:  <pre>message-demarcator=\${literal('&amp;#44;'):unescapeXml() }</pre> </li> <li>- Semicolon. Separates the new content with a semicolon. Enter the following value to use a semicolon as a message demarcator:  <pre>message-demarcator=\${literal('&amp;#59;'):unescapeXml() }</pre> </li> <li>- Tab. Separates the new content with a tab. Enter the following value to use a tab as a message demarcator:  <pre>message-demarcator=\${literal('&amp;#09;'):unescapeXml() }</pre> </li> </ul>
max.poll.records	<p>Specifies the maximum number of records returned in a single call to poll.</p> <p>For example, <code>max.poll.records=100000</code></p>

## MQTT source properties

The following table describes the MQTT source properties on the **Source** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the MQTT source connection.
Connection Type	The MQTT connection type. The connection type populates automatically based on the connection that you select.
Topic	Name of the MQTT topic.

The following table describes the advanced properties for the MQTT source on the **Source** tab when you define a streaming ingestion task:

Connection Property	Description
Client ID	Optional. Unique identifier that identifies the connection between the MQTT source and the MQTT broker. The client ID is the file-based persistence store that the MQTT source uses to store messages when they are being processed.  If you do not specify a client ID, the streaming ingestion task uses the client ID provided in the MQTT connection. However, if you have not specified the client ID even in the MQTT connection, the streaming ingestion task generates a unique client ID.
Max Queue Size	Optional. The maximum number of messages that the processor can store in memory at the same time.  Default value is 1024 bytes.

## OPC UA source properties

The following table describes the OPC UA source properties on the **Source** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the OPC UA source connection.
Connection Type	The OPC UA connection type. The connection type populates automatically based on the connection that you select.
Tag List Specified As	Format in which the list of tags is specified. Select one of the following formats: <ul style="list-style-type: none"><li>- <b>List of Tags.</b> List of tags to be read by the OPC client, specified as a JSON array.</li><li>- <b>Path for Tags File.</b> File containing list of tags to be read by the OPC client, specified as a JSON array.</li></ul>



Property	Description
Tags or File Path	List of tags or path to the file containing the list of tags to be read, specified as a JSON array. The list of tags or file path cannot exceed 2048 characters.
Minimum Publish Interval	The minimum publish interval of subscription notification messages, in milliseconds. Set this property to a lower value to detect the rapid change of data. Default is 1,000 milliseconds.

## REST V2 source properties

The following table describes the REST V2 source properties on the **Source** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the REST V2 source connection.
Connection Type	The REST V2 connection type. The connection type populates automatically based on the connection that you select.
REST Endpoints	List of REST endpoints specified in the input swagger file. These endpoints appear based on the chosen REST connection.
Scheme	List of schemes specified in the swagger definition. The selected scheme is used to create a the URL.
Poll Interval	Interval between two consecutive REST calls. Default is 10 seconds.
Action on Unsuccessful Response codes	Action required for unsuccessful REST calls. You can choose of the following actions: <ul style="list-style-type: none"> <li>- Raise Alert</li> <li>- Route to Downstream. Route the response to the downstream processors.</li> <li>- Route to Reject Directory: Route the response to the reject directory configured in Runtime Options page.</li> </ul>

Based on the defined operation ID in the selected **REST Endpoints** property, the dynamic properties such as, **Path**, **Query**, and **Payload** appear at the lower section of the REST V2 source page.

- **Headers.** Adds header to a REST call.
- **Path.** Consists of multiple path parameters as specified in swagger definition. You cannot edit the **Path Key**. You can only enter corresponding values for the path keys.
- **Query.** Consists of query parameters. Query parameters are similar to path parameters.
- **Payload.**
  - **Sample Payload.** A read-only text box that shows schema of request body for a PUT, POST, or PATCH request. For example, { "name" : "string", "salary" : "string", "age" : "string" }.
  - **Body.** The request body to be sent incase of PUT, POST, or PATCH request. You can copy a sample request body from a sample payload and then can replace the values as appropriate.

You can define any of these properties as mandatory in the swagger specification file. Then the same property is considered mandatory while configuring a streaming ingestion REST V2 source. If you do not

define a REST endpoint in the swagger specification file, the corresponding section does not appear in the streaming ingestion REST V2 page.

## Configuring a target

To configure a target, select a target connection to which you want to transfer the streaming data and then configure the target properties. Before you configure a target, ensure that the connection to the target is created in the Administrator service.

1. On the **Target** page, select a connection.

The streaming ingestion task supports the following targets:

- Amazon Kinesis Data Firehose
- Amazon Kinesis Streams
- Amazon S3 V2
- Apache Kafka
- Databricks Delta
- Flat file
- Google PubSub
- Google Cloud Storage V2
- JDBC V2
- Microsoft Azure Data Lake Storage Gen2
- Microsoft Azure Event Hubs

2. Based on the target that you select, enter the required details.

Options that appear on the **Target** tab of the task wizard vary based on the type of target that you select.

3. Under **Advanced Properties**, enter the required information.

4. Perform one of the following tasks:

- To add a transformation, click **Next**.  
The **Transformation** tab appears.

- To save the task, click **Save**.

You can then deploy the streaming ingestion task. For more information about deploying the streaming ingestion task, see [“Deploying a streaming ingestion task” on page 487](#).

## Amazon Kinesis Data Firehose target properties

The following table describes the Amazon Kinesis Data Firehose target properties on the **Target** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the Amazon Kinesis Data Firehose target connection.
Connection Type	The Amazon Kinesis connection type. The connection type populates automatically based on the connection that you select.
Stream Name/ Expression	Kinesis stream name or a regular expression for the Kinesis stream name pattern. Use the <code>\$expression\$</code> format for the regular expression. <code>\$expression\$</code> evaluates the data and sends the matching data to capturing group 1.

For more information about Kinesis Data Firehose, see the Amazon Web Services documentation.

## Amazon Kinesis Streams target properties

The following table describes the Amazon Kinesis Streams target properties on the **Target** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the Amazon Kinesis Stream target connection.
Connection Type	The Amazon Kinesis connection type. The connection type populates automatically based on the connection that you select.
Stream Name/ Expression	Kinesis stream name or a regular expression for the Kinesis stream name pattern. Use the <code>\$expression\$</code> format for the regular expression. <code>\$expression\$</code> evaluates the data and sends the matching data to capturing group 1.

For more information about Kinesis Streams, see the Amazon Web Services documentation.

## Amazon S3 target properties

The following table describes the Amazon S3 target properties on the **Target** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the Amazon S3 target connection.
Connection Type	The Amazon S3 V2 connection type. The connection type populates automatically based on the connection that you select.
Object Name/ Expression	Amazon S3 file name or a regular expression for the Amazon S3 file name pattern. Use the <code>\$expression\$</code> format for the regular expression. <code>\$expression\$</code> evaluates the data and sends the matching data to capturing group 1.

The following table describes the Amazon S3 advanced target properties that you can configure on the **Target** tab when you define a streaming ingestion task:

Property	Description
Minimum Upload Part Size	Optional. Minimum upload part size when uploading a large file as a set of multiple independent parts, in megabytes. Use this property to tune the file load to Amazon S3. Default value is 5120 MB.
Multipart Upload Threshold	Optional. Multipart download minimum threshold to determine when to upload objects in multiple parts in parallel. Default value is 5120 MB.

## Azure Event Hubs target properties

The following table describes the Azure Event Hubs target properties on the **Target** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the Azure Event Hubs target connection.
Connection Type	The Azure Event Hubs connection type. The connection type populates automatically based on the connection that you select.
Event Hub	The name of the Azure Event Hubs.

The following table describes the Azure Event Hubs advanced target properties that you can configure on the **Target** tab when you define a streaming ingestion task:

Property	Description
Shared Access Policy Name	Optional. The name of the Event Hub Namespace Shared Access Policy. The policy must apply to all data objects that are associated with this connection. To read from Event Hubs, you must have Listen permission. To write to an Event Hub, the policy must have Send permission.
Shared Access Policy Primary Key	Optional. The primary key of the Event Hub Namespace Shared Access Policy.

## Databricks Delta target properties

The following table describes the Databricks Delta target properties on the **Target** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the Databricks Delta target connection.
Connection Type	The Databricks Delta connection type. The connection type populates automatically based on the connection that you select.
Use Existing Cluster	Choose whether you want to use the existing cluster or provision a new cluster. Choose <b>True</b> to use the existing cluster. If you choose <b>True</b> , provide the existing cluster ID.
Retry Attempts	The maximum number of times the Secure Agent retries the REST API calls to Databricks when an error occurs due to network connection or the REST endpoint returns 5xx HTTP error code. Default is 0.
Retry Delay Interval	The time Interval, in milliseconds, at which the Secure Agent must retry the REST API call when an error occurs due to network connection or the REST endpoint returns 5xx HTTP error code. Default is 1,000 milliseconds.
Job Status Poll Interval	Poll interval in seconds at which the Secure Agent checks the status of the job completion, in milliseconds.
Staging Location	Relative directory path to store the staging files. <ul style="list-style-type: none"><li>- If the Databricks cluster is deployed on AWS, use the relative path of the Amazon S3 staging bucket.</li><li>- If the Databricks cluster is deployed on Azure, use the relative path of the Azure Data Lake Store Gen2 staging file system name.</li></ul>
Target Table Name	Name of the Databricks Delta table to append.

The following table describes the Databricks Delta target advanced properties that you can configure on the **Target** tab when you define a streaming ingestion task:

Property	Description
Data Location	Relative path to store the data. If you do not provide a value, a managed table with the table name specified in <b>Target Table Name</b> property is created.
Target Database Name	Overrides the database name provided in the Databricks Delta connection in Administrator.

For a Databricks Delta target, the source messages must be only in JSON format.

**Note:** In a streaming ingestion job with Databricks Delta target, when you change the source schema to include additional data columns, Informatica recommends that you redeploy the job to include the change data capture.

When you use a Filter transformation in a streaming ingestion task with a Databricks Delta target, ensure that the ingested data conforms to a valid JSON data format. The Filter transformation with JSONPath filter type

validates the incoming data. If the incoming data does not conform to a valid JSON data format, the streaming ingestion task rejects the data. The rejected data then moves into the configured reject directory. If you do not have a reject directory already configured, the rejected data is lost.

Informatica recommends that you use a Combiner transformation in the streaming ingestion task that contains a Databricks Delta target. Add the Combiner transformation before writing to the target. The streaming ingestion task then combines all the staged data before writing into the Databricks Delta target.

## Flat file target properties

The following table describes the flat file target properties on the **Target** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the flat file target connection.
Connection Type	The flat file connection type. The connection type appears based on the connection that you select.
Staging Directory Location	Path to the staging directory on the Secure Agent. Specify the staging directory where to stage the files when you write data to a flat file target. Ensure that the directory has sufficient space and you have write permissions to the directory.
Rollover Size *	The file size, in kilobytes (KB), at which the task moves the file from the staging directory to the target. For example, set the rollover size to 1 MB and name the file target.log. If the source service sends 5 MB to the target, the streaming ingestion task first creates the target.log.<timestamp> file. When the size of target.log.<timestamp> reaches 1 MB, the task rolls the file over.
Rollover Events Count *	Number of events or messages to accumulate for file rollover. For example, if you set the rollover events count to 1000, the task rolls the file over when the file accumulates 1000 events.
Rollover Time *	Length of time, in milliseconds, for a target file to roll over. After the time period has elapsed, the target file rolls over. For example, if you set rollover time as 1 hour, the task rolls the file over when the file reaches a period of 1 hour.
File Name	The name of the file that the task creates on the target.
* Specify a value for at least one rollover option to perform target file rollover.	

## Google Cloud Storage target properties

The following table describes the Google Cloud Storage target properties on the **Target** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the Google Cloud Storage target connection.
Connection Type	The Google Cloud Storage connection type. The connection type populates automatically based on the connection that you select.
Number of Retries	The number of times the streaming ingestion task retries to write to the Google Cloud Storage target. Default is 6.
Bucket	The container to store, organize, and access objects that you upload to Google Cloud Storage.
Key	Name of the Google Cloud Storage target object.

The following table describes the Google Cloud Storage advanced target properties on the **Target** tab when you define a streaming ingestion task:

Property	Description
Proxy Host	Host name of the outgoing proxy server that the Secure Agent uses.
Proxy Port	Port number of the outgoing proxy server.
Content Type	The file content type. You can specify any MIME types, such as application.json, multipart, text, or html. These values are not case sensitive. Default is text.
Object ACL	Access control associated with the uploaded object. Choose one of the following types of authentication: <ul style="list-style-type: none"><li>- <b>Authenticated Read.</b> Gives the bucket or object owner FULL_CONTROL permission and gives all authenticated Google account holders READ permission.</li><li>- <b>Bucket Owner Full Control.</b> Grants full control permission to the bucket or object owner and grants read permission to all the authenticated Google account holders.</li><li>- <b>Bucket Owner Read Only.</b> Grants full control permission to the object owner and grants read permission to the bucket owner. Use this type only with objects.</li><li>- <b>Private.</b> Gives the bucket or object owner FULL_CONTROL permission for a bucket or object.</li><li>- <b>Project Private.</b> Gives permission to the project team based on their roles. Anyone who is part of the team has READ permission and project owners and project editors have FULL_CONTROL permission. This is the default ACL for newly created buckets.</li><li>- <b>Public Read Only.</b> Gives the bucket owner FULL_CONTROL permission and gives all anonymous users READ and WRITE permission. This ACL applies only to buckets. When you apply this to a bucket, anyone on the Internet can list, create, overwrite, and delete objects without authenticating.</li></ul>
Server Side Encryption Key	Server-side encryption key for the Google Cloud Storage bucket. Required if the Google Cloud Storage bucket is encrypted with SSE-KMS.
Content Disposition Type	Type of RFC-6266 Content Disposition to be attached to the object. Choose either <b>Inline</b> or <b>Attachment</b> .

## Google PubSub target properties

The following table describes the Google PubSub target properties on the **Target** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the Google PubSub target connection.
Connection Type	The Google PubSub connection type. The connection type populates automatically based on the connection that you select.
Topic	Name of the target Google PubSub topic.
Batch Size	Maximum number of messages that the Cloud service bundles together in a batch. Default is 1.

## JDBC V2 target properties

The following table describes the JDBC V2 target properties on the **Target** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the JDBC V2 target connection.
Connection Type	The JDBC V2 connection type. The connection type populates automatically based on the connection that you select.
Table name	Name of the table to insert data to in JSON format.

## Kafka target properties

The following table describes the Kafka target properties that on the **Target** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the Kafka target connection.
Connection Type	The Kafka connection type. The connection type populates automatically based on the connection that you select.
Topic Name/ Expression	<p>Kafka topic name or a Java supported regular expression for the Kafka topic name pattern. Use the <code>\$expression\$</code> format for the regular expression. <code>\$expression\$</code> evaluates the data and sends the matching data to capturing group 1.</p> <p>You can either enter the topic name manually or fetch the already created metadata of the Kafka connection.</p> <ol style="list-style-type: none"><li>1. Click <b>Select</b>. The <b>Select Target Object</b> dialog box appears showing all the topics available in the Kafka broker. However, Kafka topic name patterns do not appear in the list.</li><li>2. Select the required topic and click <b>OK</b>.</li></ol>



The following table describes the Kafka advanced target properties on the **Target** tab when you define a streaming ingestion task:

Property	Description
Producer Configuration Properties	The configuration properties for the producer.
Metadata Fetch Timeout in milliseconds	The time after which the metadata is not fetched.
Batch Flush Size in bytes	The batch size of the events after which a streaming ingestion task writes data to the target.

## Microsoft Azure Data Lake Storage Gen2 target properties

The following table describes the Microsoft Azure Data Lake Storage Gen2 (ADLS Gen2) target properties on the **Target** tab when you define a streaming ingestion task:

Property	Description
Connection	Name of the Microsoft Azure Data Lake Storage Gen2 target connection.
Connection Type	The ADLS Gen2 connection type. The connection type populates automatically based on the connection that you select.
Write Strategy	<p>The operation type to write data to ADLS Gen2 file. If the file exists in ADLS Gen2 storage, you can select to overwrite, append, fail, or rollover the existing file. Default is <b>Append</b>.</p> <ul style="list-style-type: none"> <li>- <b>Append</b>. Add data to an existing file inside a directory.</li> <li>- <b>Overwrite</b>. Delete existing data in an existing file and insert newly read data.</li> <li>- <b>Fail</b>. Write data to an existing file fails.</li> <li>- <b>Rollover</b>. Close the current file to which data is being written to and create a new file based on the configured rollover value.</li> </ul> <p><b>Note:</b> When you edit or redeploy a streaming ingestion job that contains a target with the rollover strategy, all the files in the staging directory are moved to the target directory even if the defined rollover conditions are not met.</p>
Interim Directory	<p>Path to the staging directory in ADLS Gen2. Specify the staging directory where you want to stage the files when you write data to ADLS Gen2. Ensure that the directory has sufficient space and you have write permissions to the directory. Applicable when you select the <b>Write Strategy</b> as Rollover.</p> <p>While configuring an ADLS Gen 2 target in a streaming ingestion job, if you do not specify any value for the rollover properties, the files remain in the interim directory. When you stop or undeploy the streaming ingestion job, these files in the interim directory are moved to the target location, by default.</p>
Rollover Size	<p>Target file size, in kilobytes (KB), at which to trigger rollover. Applicable when you select the <b>Write Strategy</b> as Rollover.</p>
Rollover Events Count	<p>Number of events or messages that you want to accumulate for the rollover. Applicable when you select the <b>Write Strategy</b> as Rollover.</p>

Property	Description
Rollover Time	Length of time, in milliseconds, for a target file to roll over. After the time period has elapsed, the target file rolls over. Applicable when you select the <b>Write Strategy</b> as Rollover.
File Name/ Expression	File name or a regular expression for the file name pattern. Use the <code>\$expression\$</code> format for the regular expression. <code>\$expression\$</code> evaluates the data and sends the matching data to capturing group 1.

The following table describes the Microsoft Azure Data Lake Storage Gen2 (ADLS Gen2) advanced target properties on the **Target** tab when you define a streaming ingestion task:

Property	Description
Filesystem Name Override	Overrides the default file system name provided in connection. This file system name is used write to a file at run time.
Directory Override	Overrides the default directory path. The ADLS Gen2 directory that you use to write data. Default is root directory. The directory path specified while creating the target overrides the path specified while creating a connection.
Compression Format	Optional. Compression format to use before the streaming ingestion task writes data to the target file. Use one of the following formats: <ul style="list-style-type: none"> <li>- None</li> <li>- Gzip</li> <li>- Bzip2</li> <li>- Zlib</li> <li>- Deflate</li> </ul> Default is <b>None</b> . To read a compressed file from the data lake storage, the compressed file must have specific extensions. If the extensions used to read the compressed file are not valid, the Secure Agent does not process the file.

## Configuring a transformation

You can specify the data format of the streaming data. Based on the data format, you can configure a transformation.

- On the **Transformation** page, select the format of the streaming data.  
The streaming ingestion transformations support the following data formats:
  - Binary
  - JSON
  - XML
- Based on the selected data format, select one of the supported transformations, and configure it.

3. To add more than one transformation, click **Add Transformations**.
  - a. On the **Transformation** tab, click **Add Transformation**.  
The **New Transformation** dialog box appears.
  - b. Based on the transformation type you select, enter the required properties.
  - c. Click **Save**.  
The saved transformation appears under **Transformations** in the **Transformation Details** wizard.
4. Perform one of the following tasks:
  - To configure the runtime options for the task, click **Next**.  
The **Runtime Options** tab appears.
  - To save the task, click **Save**.  
You can then deploy the streaming ingestion task. For more information about deploying the streaming ingestion task, see [“Deploying a streaming ingestion task” on page 487](#).

## Adding a transformation

1. On the **Transformation** tab, click **+** to add a transformation.



The screenshot shows the 'Transformation Details' wizard. At the top, there's a section for 'Incoming Message Format' with a dropdown menu set to 'Binary'. Below this is a table titled 'Transformations' with columns 'Name' and 'Type'. The table is currently empty, and a red box highlights a '+' button in the top right corner of the table area, indicating where to click to add a new transformation. Below the table, it says 'No data to display'.

- The **New Transformation** dialog box appears.
2. Based on the transformation type you select, enter the required properties.
  3. Click **Save**.  
The saved transformation appears under **Transformations** in the **Transformation Details** wizard.

## Combiner transformation properties

The following table describes the properties you can configure for a Combiner transformation:

Property	Description
Transformation Type	Select Combiner.
Transformation Name	Name of the Combiner transformation.
Minimum Number of Events	Minimum number of events to collect before the transformation combines the events into a single event. Default is 1.

Property	Description
Maximum Aggregate Size	Maximum size of the combined events in megabytes. If not specified, this transformation waits to meet any of the other two conditions before combining the events.
Time Limit	Maximum time to wait before combining the events. If not specified, this transformation waits for the other conditions before combining the events or waits forever.
Delimiter	Symbols used to specify divisions between data strings in the transformed data. Applicable only for the binary data format.
Append the delimiter character to the last record in each batch	When there are many batches with events or records, you can choose whether to use the delimiter character at the end of the last record in each batch. This enables the delimiter character to act as the division between each batch.

## Filter transformation properties

The following table describes the properties you can configure for a Filter transformation:

Property	Description
Transformation Type	Select Filter.
Transformation Name	Name of the Filter transformation.
Filter Type	Type of filter to evaluate the incoming data. Use one of the following filter types: <ul style="list-style-type: none"> <li>- JSON Path. An expression that consists of a sequence of JSON properties.</li> <li>- Regular Expression. A range or pattern of values.</li> <li>- XPath. An expression that selects nodes or node-sets in an XML document.</li> </ul>
Expression	Expression for the filter type that you select.

## Python transformation properties

The following table describes the properties you can configure for a Python transformation:

Property	Description
Transformation Type	Select Python.
Transformation Name	Name of the Python transformation.
Script Input Type	Python script input type. You can either enter the Python script in <b>Script Body</b> or provide the path to the Python script available in the <b>Script Path</b> .
Python Path	Directory to the Python path libraries.

## Splitter transformation properties

The following table describes the properties you can configure for a Splitter transformation for binary message format:

Property	Description
Transformation Type	Select Splitter.
Transformation Name	Name of the Splitter transformation.
Split Type	Split condition to evaluate the incoming data. Use one of the following split types: <ul style="list-style-type: none"><li>- Line Split.</li><li>- Content Split.</li></ul>
Line Split Count	The maximum number of lines that each output split file contains, excluding header lines.
Byte Sequence	Specified sequence of bytes on which to split the content.

The following table describes the properties you can configure for a Splitter transformation for JSON message format:

Property	Description
Split Expression	Split condition to evaluate the incoming data. Use one of the following split types: <ul style="list-style-type: none"><li>- Array Split.</li><li>- JSONPath Expression.</li></ul>
JSONPath Expression	A JSONPath expression that specifies the array element to split into JSON or scalar fragments. The default JSONpath Expression is \$.

The following table describes the properties you can configure for a Splitter transformation for XML message format:

Property	Description
Split Depth	The XML nesting depth to start splitting the XML fragments. The default split depth is 1.

## Format Converter transformation properties

When you define a streaming ingestion task and add a Format Converter transformation, provide values for transformation properties on the **New Transformation** page of the task wizard.

The following table describes the properties you can configure for a Format Converter transformation:

Property	Description
Transformation Type	Select Format Converter.
Transformation Name	Name of the Format Converter transformation.
Convert to Format	The streaming ingestion task converts incoming data to the selected format. Currently, the Format Converter transformation converts the incoming data only to Parquet format.
Date Format *	Enter the format of dates in input fields. For example, MM/dd/yyyy.
Time Format *	Enter the format of time in input fields. For example, HH/mm/ss.
Timestamp Format *	Enter the format of timestamps in input fields. For example, the epoch timestamp for 10/11/2021 12:04:41 GMT (MM/dd/yyyy HH:mm:ss) is 1633953881 and the timestamp in milliseconds is 1633953881000.
Expect Records as Array	Determines whether to expect a single record or an array of records. Select this property to expect arrays in each record. Applies only to XML incoming messages. By default, this property is deselected.
* If the format is not specified, it is considered in milliseconds since the epoch (Midnight, January 1, 1970, GMT).	

## Configuring runtime options


You can configure additional runtime options for the streaming ingestion task. Runtime options include settings to manage reject events and to notify users about errors.

1. On the **Runtime Options** page, under **Notification Management**, select to either disable the notifications or set the time limit after which you want to receive a notification if an error occurs.  
You can set the time in minutes or hours. Default is 10 minutes.
2. Enter a list of email addresses to which you want to send notifications if a task runs with error.  
Use commas to separate a list of email addresses. Note that email notification options configured for the organization are not used here.
3. Under **Agent Parameters**, perform the following tasks:
  - a. Specify a directory to store the rejected events.  
Rejected events are not stored by default.  
Note that filtered events in a regular expression are not moved to the reject directory.
  - b. To purge the log files, specify the maximum file size for the log file after which the log file is purged.  
You can set the file size in megabytes or gigabytes. Default is 10 MB.
  - c. Choose the severity of an event that you want to log.  
The default log level is `Info`.

The supported log levels are:

- Debug. Logs all messages along with additional debug level messages.
- Error. Logs only the error messages.
- Info. Logs all errors, warnings, and important informational messages.
- Warn. Logs all errors and warning messages.

4. Under **Advanced Parameters**, perform the following tasks to improve the performance of the streaming ingestion task:

- a. Click the  icon next to **Advanced Parameters**.

The **Key** and **Value** fields appear.

- b. Specify a valid parameter property and its value.

5. To save the task, click **Save**.

You can then deploy the streaming ingestion task. For more information about deploying the streaming ingestion task, see [“Deploying a streaming ingestion task” on page 487](#).

## Deploying a streaming ingestion task

After you create a streaming ingestion task, you must deploy it on the Secure Agent for the task to run as a job. Before deploying the task, ensure that the Secure Agent configured for the task is running.

- To deploy a task, perform one of the following actions:

- After you save a task, click **Deploy**.
- On the **Explore** page, open the project that contains the task, and then select **Deploy** from the **Actions** menu for the task.

A message about successful deployment of the task appears. Your job is now queued to run.

When you edit or undeploy a job, deploying the job again does not affect any other jobs running on the same Secure Agent or Secure Agent group.

When you edit or redeploy a streaming ingestion job that contains a target with the rollover strategy, all the files in the staging directory are moved to the target directory even if the defined rollover conditions are not met.

## Undeploying a streaming ingestion job

You can undeploy a streaming ingestion job from the Secure Agent.

- On the **Explore** page, open the project that contains the job, and then select **Undeploy** from the **Actions** menu for the job.

Alternatively, you can undeploy a streaming ingestion job from the action menu in a job row on the **Mass Ingestion** page in Monitor and Operational Insights.

When you undeploy a streaming ingestion job, the Secure Agent does not save the previous state histories of the job. So, when you deploy the streaming ingestion task again, the task runs as completely new job.

## Stopping and resuming streaming ingestion jobs

You can stop and resume streaming ingestion jobs. Stop or resume a job on the **My Jobs** page in Mass Ingestion or **All Jobs** page on the Mass Ingestion page in Monitor and Operational Insights.

### Stop a job.

You can stop a streaming ingestion job that is Up and Running, Running with Error, or Running with Warnings.

To stop a job, open the **My Jobs** page, and then select **Stop** from the **Actions** menu for the job.

### Resume a job.

You can resume a stopped streaming ingestion job.

To resume a job, open the **My Jobs** page, and then select **Resume** from the **Actions** menu for the job.

Alternatively, you can stop or resume a streaming ingestion job from the **Actions** menu in a job row on the **Mass Ingestion** page in Monitor and Operational Insights.

When you stop a streaming ingestion job, the Secure Agent saves the previous state histories of the job. When you resume the stopped streaming ingestion job, the job starts running from the last saved state.

## Mass Ingestion Streaming REST API

Use streaming ingestion resources to deploy, undeploy, start, and stop streaming ingestion tasks and to monitor streaming ingestion jobs.

When you use the streaming ingestion resource, use the following request header format:

```
<METHOD><base URL>  
Content-Type: application/json  
Accept: application/json  
IDS-SESSION-ID: <SessionId>
```

## Dataflows

Use the Dataflows resource to deploy, undeploy, start, and stop streaming ingestion tasks.

Use the following base URL:

```
<server URI>/sisvc/api/v1/Dataflows('<dataflow ID>')/OData.SI.<API name>
```

**Note:** If you use a tool such as Postman that automatically includes the HTTP version, do not enter the HTTP version in the URL. If the HTTP version appears twice in the URL, the request fails.

### Deploying a streaming ingestion task

Use a POST request to deploy a streaming ingestion task.

#### POST request

To deploy a streaming ingestion task, use the following URL:

```
<server URI>/sisvc/api/v1/Dataflows('<dataflow ID>')/OData.SI.Deploy
```

A request body is not required because the URL passes the dataflow ID.



## POST request example

To deploy a streaming ingestion task, you might send a request similar to the following example:

```
POST <serverUrl>/sisvc/api/v1/Dataflows('50077311-d4a4-437c-9218-c3596d1f182f') /
OData.SI.Deploy
Content-Type: application/json
Accept: application/json
IDS-SESSION-ID: 210oeVx22Rujiej7yTokmT
```

## POST response example

If the request is successful, you might receive a response similar to the following example:

```
{
  "@odata.context": "$metadata#OData.SI.DeploymentResult",
  "successful": true,
  "code": null,
  "errorMessage": null
}
```

## Undeploying a streaming ingestion task

Use a POST request to undeploy a streaming ingestion task.

### POST request

To undeploy a streaming ingestion task, use the following URL:

```
<server URI>/sisvc/api/v1/Dataflows('<dataflow ID>')/OData.SI.Undeploy
```

A request body is not required because the URL passes the dataflow ID.

### POST request example

To undeploy a streaming ingestion task, you might send a request similar to the following example:

```
POST <serverUrl>/sisvc/api/v1/Dataflows('50077311-d4a4-437c-9218-c3596d1f182f') /
OData.SI.Undeploy
Content-Type: application/json
Accept: application/json
IDS-SESSION-ID: 210oeVx22Rujiej7yTokmT
```

### POST response example

If the request is successful, you might receive a response similar to the following example:

```
{
  "@odata.context": "$metadata#OData.SI.DeploymentResult",
  "successful": true,
  "code": null,
  "errorMessage": null
}
```

## Starting a streaming ingestion task

Use a POST request to start a streaming ingestion task.

### POST request

To start a streaming ingestion task, use the following URL:

```
<server URI>/sisvc/api/v1/Dataflows('<dataflow ID>')/OData.SI.Start
```

A request body is not required because the URL passes the dataflow ID.

## POST request example

To start a streaming ingestion task, you might send a request similar to the following example:

```
POST <serverUrl>/sisvc/api/v1/Dataflows('50077311-d4a4-437c-9218-c3596d1f182f') /
OData.SI.Start
Content-Type: application/json
Accept: application/json
IDS-SESSION-ID: 210oeVx22Rujiej7yTokmT
```

## POST response example

If the request is successful, you might receive a response similar to the following example:

```
{
  "@odata.context": "$metadata#OData.SI.DeploymentResult",
  "successful": true,
  "code": null,
  "errorMessage": null
}
```

## Stopping a streaming ingestion task

Use a POST request to stop a streaming ingestion task.

### POST request

To stop a streaming ingestion task, use the following URL:

```
<server URI>/sisvc/api/v1/Dataflows('<dataflowID>')/OData.SI.Stop
```

A request body is not required because the URL passes the dataflow ID.

### POST request example

To stop a streaming ingestion task, you might send a request similar to the following example:

```
POST <serverUrl>/sisvc/api/v1/Dataflows('d7572789-dc4c-4c56-bbeb-3772736d61aa') /
OData.SI.Stop
Content-Type: application/json
Accept: application/json
IDS-SESSION-ID: 210oeVx22Rujiej7yTokmT
```

### POST response example

If the request is successful, you might receive a response similar to the following example:

```
{
  "@odata.context": "$metadata#OData.SI.DeploymentResult",
  "successful": true,
  "code": null,
  "errorMessage": null
}
```

## jobs

Use the jobs resource to get the details of a streaming ingestion job.

### GET request

To request the details of a streaming ingestion job, use the following URL:

```
<server URI>/sisvc/monitor/v1/jobs/<dataflow ID>/<run ID of the job>
```

## GET request example

To request the details of a streaming ingestion job, you might send a request similar to the following example:

```
POST https://usw1-ing.dm2-us.informaticacloud.com/sisvc/monitor/v1/jobs/
1948938e-3923-4602-aba8-f122e3d66faf/42559
Content-Type: application/json
Accept: application/json
IDS-SESSION-ID: 2l0oeVx22Rujiej7yTokmT
```

## GET response

Returns the jobs object if successful or an error object if an error occurs.

If successful, the response includes the following information about a streaming ingestion job:

Parameter	Type	Description
assetId	String	ID of the streaming ingestion job.
assetName	String	Name of the streaming ingestion job.
duration	Integer	The time it took to deploy the job.
endTime	Integer	End time of deploying the job, in UTC time.
startTime	Integer	Start time of deploying the job, in UTC time.
extraData	String	Additional information including the task ID, deployed version, and the Secure Agent group ID.
runId	Integer	Run ID of the streaming ingestion job. The ID changes for every deployment.
orgId	String	ID of the organization the logged in user belongs to.
runtimeEnv	String	ID of the Secure Agent that deployed the streaming ingestion job.
startedBy	String	Name of the user who created the streaming ingestion task.
status	String	The status of the streaming ingestion job. A job can be in one of the following status: <ul style="list-style-type: none"><li>- Deploying. The job is being deployed.</li><li>- Up and Running. The job is running.</li><li>- Running with Warning. The job is running with warnings.</li><li>- Running with Error. The job is running with error.</li><li>- Undeployed. The job is undeployed.</li><li>- Stopped. The job was intentionally stopped.</li></ul>

## GET response example

If the request to get the details of a streaming ingestion job is successful, you might receive a response similar to the following example:

```
{
  "assetId": "1948938e-3923-4602-aba8-f122e3d66faf",
  "assetName": "testmonitor",
  "assetType": "SI_DATAFLOW",
  "correlationId": null,
  "duration": 1543,
  "endTime": "2022-02-14T04:04:13.000+0000",
  "extraData": "{\"id\":\"0RwiUUb9bVwjL67dWOKjoI\",\"version\":1,\"agentGroupId\":null}",
  "location": "Default",
```

```

    "runId": 42559,
    "orgId": "21Fy0UUNlnbjhaoT3TSqw",
    "runtimeEnv": "011ZFB2500000000000N",
    "startedBy": "siga_new",
    "status": "Undeployed",
    "startTime": "2022-02-14T03:38:30.000+0000",
    "deployedVersion": 1
  }

```

## MIJobs

Use the MIJobs resource to get a list of the available streaming ingestion jobs.

### GET request

To request a list of the available streaming ingestion jobs, use the following URL:

```
<server URI>/mijobmonitor/api/v1/MIJobs
```

You can include the following query parameters in the URI:

Parameter	Type	Required	Description
\$count	Boolean	No	Displays the number of ingestion jobs in the database.
\$filter	String	No	Filters the job based on the input. You can filter using one of the following fields: <ul style="list-style-type: none"> <li>- assetName</li> <li>- assetType</li> <li>- startedBy</li> <li>- status</li> </ul> You can filter jobs using single or multiple fields.
\$orderby	String	No	Sorts the order of the jobs. You can sort the jobs using the following fields: <ul style="list-style-type: none"> <li>- assetName</li> <li>- assetType</li> <li>- status</li> <li>- runtimeEnv</li> <li>- startTime</li> </ul> You can sort jobs using single or multiple fields.
\$skip	Integer	No	Skips the number of streaming ingestion jobs that you specify. For example, you might want to skip the first five streaming ingestion jobs. Consider the \$filter and \$orderby parameter values, if specified.
\$top	Integer	No	Displays the number of top streaming ingestion jobs that you specify. For example, you might want to view the top ten streaming ingestion jobs. Consider the \$filter and \$orderby parameter values, if specified.

### GET request example

To get a list of the available streaming ingestion jobs, you might send a request similar to the following example:

```

POST https://usw1-ing.dm2-us.informaticacloud.com/mijobmonitor/api/v1/MIJobs?$count=true&
$filter=(startedBy eq 'siga_new')&$orderby=deployTime desc&$skip=0&$top=25
Content-Type: application/json
Accept: application/json
IDS-SESSION-ID: 2l0oeVx22Rujiej7yTokmT

```

## GET response

Returns the MIjobs object if successful or an error object if an error occurs.

If successful, the response includes the following information about the streaming ingestion job:

Parameter	Type	Description
assetName	String	Name of the streaming ingestion job.
runId	Integer	Run ID of the streaming ingestion job. The ID changes for every deployment.
orgId	String	ID of the organization the logged in user belongs to.
runtimeEnv	String	ID of the Secure Agent that deployed the streaming ingestion job.
startTime	Integer	Date and start time of deploying the job, in UTC time.
endTime	Integer	Date and end time of deploying the job, in UTC time.
deployTime	Integer	Date and time of deploying the job, in UTC time.
undeployTime	Integer	Date and time of undeploying the job, in UTC time.
startedBy	Integer	Name of the user who created the streaming ingestion task.
status	String	The status of the streaming ingestion job. A job can be in one of the following status: <ul style="list-style-type: none"><li>- Deploying. The job is being deployed.</li><li>- Up and Running. The job is running.</li><li>- Running with Warning. The job is running with warnings.</li><li>- Running with Error. The job is running with error.</li><li>- Undeployed. The job is undeployed.</li><li>- Stopped. The job was intentionally stopped.</li></ul>
extraData	String	Additional information including the task ID, the location of the streaming ingestion job, and the Secure Agent ID.

## GET response example

If the request to get a list of available streaming ingestion jobs is successful, you might receive a response similar to the following example:

```
{
  "@odata.context": "$metadata#Collection(OData.MI.JobMonitor.MIJob)",
  "@odata.count": 421,
  "value": [
    {
      "assetId": "7ce6bbc7-f0e2-4278-bd6d-d1187f4a1420",
      "assetName": "SIdeployJms",
      "assetType": "SI_DATAFLOW",
      "runId": 33015,
      "duration": 300000,
      "orgId": "1Pm6cSfPcAqfgeV57Fn3u4",
      "runtimeEnv": "011U5M080000000000003",
      "startTime": "2021-04-29T13:09:48.000+0000",
      "endTime": "2021-04-29T13:14:48.000+0000",
      "deployTime": "2021-04-29T13:09:48.000+0000",
      "undeployTime": "2021-04-29T13:14:48.000+0000",
      "startedBy": "siqa_new",
      "status": "Undeployed",
      "outOfSync": true,
      "extraData": "{\"taskId\":\"7Z4ZzjXc9QViT4t2okiHuz\",\"runtimeEnv"
```

```

\":"011U5M25000000000002\","location\":"RestAutomation\"},"",
    "deployedVersion": 1,
    "replace": null,
    "lastUpdateTime": 0
  },
  {
    "assetId": "a03b9aa1-4a4a-47ee-808d-ddc0ee7b3a4a",
    "assetName": "kafka to kafka test",
    "assetType": "SI_DATAFLOW",
    "runId": 33527,
    "duration": 204988000,
    "orgId": "iPm6cSfPcAqfgeV57Fn3u4",
    "runtimeEnv": "011U5M080000000000002",
    "startTime": "2021-05-04T05:41:39.000+0000",
    "endTime": "2021-05-06T14:38:07.000+0000",
    "deployTime": "2021-05-04T05:41:39.000+0000",
    "undeployTime": "2021-05-06T14:38:07.000+0000",
    "startedBy": "siqa_new",
    "status": "Undeployed",
    "outOfSync": true,
    "extraData": "{\"taskId\":\"8V21nib7Sggiw3QoDRi5uK\",\"runtimeEnv\":"011U5M250000000000002\","location\":"Default\"},"",
    "deployedVersion": 1,
    "replace": null,
    "lastUpdateTime": 0
  }
]
}

```

## status

Use the status resource to get the status of a streaming ingestion job.

### GET request

To request the status of a streaming ingestion job, use the following URL:

```
<server URI>/sisvc/monitor/v1/status/dataflows/<dataflow ID>
```

### GET request example

To get the status of a streaming ingestion job, you might send a request similar to the following example:

```

POST https://usw1-ing.dm2-us.informaticacloud.com/sisvc/monitor/v1/status/dataflows/
1948938e-3923-4602-aba8-f122e3d66faf
Content-Type: application/json
Accept: application/json
IDS-SESSION-ID:2l0oeVx22Rujiej7yTokmT

```

### GET response

Returns the job status object if successful or an error object if an error occurs.

If successful, the response includes the following information about the status of a streaming ingestion job:

Parameter	Type	Description
dataflowName	String	Name of the streaming ingestion job.
dataflowId	Integer	ID of the streaming ingestion job.

Parameter	Type	Description
status	String	The status of the streaming ingestion job. A job can be in one of the following status: <ul style="list-style-type: none"> <li>- Deploying. The job is being deployed.</li> <li>- Up and Running. The job is running.</li> <li>- Running with Warning. The job is running with warnings.</li> <li>- Running with Error. The job is running with error.</li> <li>- Undeployed. The job is undeployed.</li> <li>- Stopped. The job was intentionally stopped.</li> </ul>
timestamp	Integer	Time, in milliseconds, when the Secure Agent records the status of the streaming ingestion job.
reports	Array	Status details of each node.
graph	String	The throughput information for the source and target of the job.
runId	Integer	Run ID of the streaming ingestion job. The ID changes for every deployment.

### GET response example

If the request to get the status of a streaming ingestion job is successful, you might receive a response similar to the following example:

```
{
  "dataflowName": "testmonitor",
  "dataflowId": "1948938e-3923-4602-aba8-f122e3d66faf",
  "status": "Running",
  "timestamp": 1644839755000,
  "reports": [
    {
      "name": "testmonitor_testmonitor_source",
      "id": "a5684428-f41f-4d24-b73f-33c232314a91",
      "status": "Running",
      "timestamp": 1644839756000,
      "message": null
    },
    {
      "name": "testmonitor_testmonitor_target",
      "id": "4f59b5fb-b5b2-4b83-994b-0d3e56f67e22",
      "status": "Running",
      "timestamp": 1644839756000,
      "message": null
    }
  ],
  "graph": "{\\\"agentId\\\":\\\"011ZFB080000000000N\\\",\\\"nodes\\\":[{\\\"id\\\":\\\"a5684428-f41f-4d24-b73f-33c232314a91\\\",\\\"name\\\":\\\"testmonitor_source\\\",\\\"serviceType\\\":\\\"source\\\",\\\"config\\\":[{\\\"key\\\":\\\"_nativeName\\\",\\\"value\\\":\\\"src\\\"},{\\\"key\\\":\\\"consumerProperties\\\",\\\"value\\\":null}],\\\"connectionId\\\":\\\"011ZFB0B00000000000KJ\\\",\\\"type\\\":\\\"\\\",\\\"metaMetadata\\\":\\\"\\\"},{\\\"id\\\":\\\"4f59b5fb-b5b2-4b83-994b-0d3e56f67e22\\\",\\\"name\\\":\\\"testmonitor_target\\\",\\\"serviceType\\\":\\\"target\\\",\\\"config\\\":[{\\\"key\\\":\\\"_nativeName\\\",\\\"value\\\":\\\"trgt\\\"},{\\\"key\\\":\\\"batchSize\\\",\\\"value\\\":\\\"1048576\\\"},{\\\"key\\\":\\\"mdFetchTimeout\\\",\\\"value\\\":\\\"5000\\\"},{\\\"key\\\":\\\"producerProperties\\\",\\\"value\\\":null}],\\\"connectionId\\\":\\\"011ZFB0B00000000000KJ\\\",\\\"type\\\":\\\"\\\",\\\"metaMetadata\\\":\\\"\\\"}],\\\"edges\\\":[{\\\"id\\\":\\\"6ae185ea-7e6e-4bf6-bd9e-0be5ef3a8e78\\\",\\\"name\\\":\\\"testmonitor_source testmonitor_target\\\",\\\"from\\\":\\\"testmonitor_source\\\",\\\"to\\\":\\\"testmonitor_target\\\",\\\"type\\\":\\\"success\\\",\\\"config\\\":[],\\\"metaMetadata\\\":\\\"\\\"}],\\\"runtimeOptions\\\":null}\\\",\\\"version\\\": 1,\\\"runId\\\": 42563\\\"}
```

## statistics

Use the statistics resource to get the statistics of a streaming ingestion job.

The streaming ingestion job should be in one of the following status before you can view its statistics:

- Deploying
- Up and Running
- Running with Warning
- Running with Error
- Stopped

### GET request

To request the statistics of a streaming ingestion job, use the following URL:

```
<server URI>/sisvc/monitor/v1/statistics/dataflows/<dataflow ID>
```

You can include the following query parameters in the URI:

Parameter	Type	Required	Description
intervals	Integer	Yes	Time, in seconds, to display statistics for a streaming ingestion job. For example, if you specify 30 seconds, the response displays job statistics for the last 30 seconds.
overall	Boolean	No	Displays the statistics from the time the job is deployed.

### GET request example

To request the statistics of a streaming ingestion job, you might send a request similar to the following example:

```
POST https://usw1-ing.dm2-us.informaticacloud.com/sisvc/monitor/v1/statistics/
dataflows/7f1daca9-3983-4677-930f-a9529802c56b?intervals=30&overall=true
Content-Type: application/json
Accept: application/json
IDS-SESSION-ID: 210oeVx22Rujiej7yTokmT
```

### GET response

Returns the statistics object if successful or an error object if an error occurs.

If successful, the response includes the following information about the statistics of a streaming ingestion job:

Parameter	Type	Description
dataflowId	String	ID of the streaming ingestion job.
dataflowRunId	Integer	Run ID of the streaming ingestion job.
startTime	Integer	Start time of the streaming ingestion job, in milliseconds.
stopTime	Integer	Stop time of the streaming ingestion job, in milliseconds.



Parameter	Type	Description
inMessages	Integer	The number of messages that arrive at a node. A node is a source, transformation, or target, that is used in the streaming ingestion task. The value is zero for a source node.
outMessages	Integer	The number of messages that transfer from a node. The value is zero for a target node.
inBytes	Integer	The total size of incoming messages in bytes. The value is zero for a source node.
outBytes	Integer	The total size of outgoing messages in bytes. The value is zero for a target node.
nodes	Array	Information about streaming data in the source and the target used in the task.
intervals	Integer	The statistics of the job for the time interval you specify in the request. Applies when you set an interval.

### GET response example

If the request to get the statistics of a streaming ingestion job is successful, you might receive a response similar to the following example:

```
{
  "dataflowId": "7f1daca9-3983-4677-930f-a9529802c56b",
  "dataflowName": "newnew",
  "dataflowVersion": 1,
  "dataflowRunId": 54231,
  "snapshotCount": 171,
  "overall": {
    "dataflowId": "7f1daca9-3983-4677-930f-a9529802c56b",
    "dataflowName": "newnew",
    "dataflowVersion": 1,
    "dataflowRunId": 54231,
    "traits": {},
    "interval": 6007,
    "startTime": 1646649995000,
    "stopTime": 1646656000000,
    "nodes": [
      {
        "name": "newnew_newnew_source",
        "id": "17a51cdf-1f27-481e-81b8-d2e8ff60ec28",
        "inMessages": 0,
        "outMessages": 0,
        "inBytes": 0,
        "outBytes": 0,
        "nodeType": "Unknown"
      },
      {
        "name": "newnew_newnew_target",
        "id": "c30d6db4-6a3b-40d3-adfb-88779a972098",
        "inMessages": 0,
        "outMessages": 0,
        "inBytes": 0,
        "outBytes": 0,
        "nodeType": "Unknown"
      }
    ]
  },
  "intervals": {
    "30": {
```

```

        "dataflowId": "7f1daca9-3983-4677-930f-a9529802c56b",
        "dataflowName": "newnew",
        "dataflowVersion": null,
        "dataflowRunId": 54231,
        "traits": {},
        "interval": 30,
        "startTime": 1646655972683,
        "stopTime": 1646656002683,
        "nodes": []
      }
    }
  }
}

```

## history

Use the history resource to get the history of a streaming ingestion job.

### GET request

To request the history of a streaming ingestion job, use the following URL:

```
<server URI>/sisvc/monitor/v1/history/dataflows/<dataflow ID>
```

### GET request example

To get the history of a streaming ingestion job, you might send a request similar to the following example:

```

POST https://usw1-ing.dm2-us.informaticacloud.com/sisvc/monitor/v1/history/dataflows/
1948938e-3923-4602-aba8-f122e3d66faf
Content-Type: application/json
Accept: application/json
IDS-SESSION-ID: 2l0oeVx22Rujiej7yTokmT

```

### GET response

Returns the job history object if successful or an error object if an error occurs.

If successful, the response includes the following information about the history of a streaming ingestion job:

Parameter	Type	Description
dataflowName	String	Streaming ingestion job name.
dataflowId	Integer	Streaming ingestion job ID.
deployedAt	Integer	The start time of deploying the job, in UTC time.
undeployedAt	Integer	The time when the job finished undeploying, in UTC.
runID	Integer	Run ID of the streaming ingestion job. The ID changes for every deployment.

### GET response example

If the request to get the history of a streaming ingestion job is successful, you might receive a response similar to the following example:

```

[
  {
    "dataflowId": "1948938e-3923-4602-aba8-f122e3d66faf",
    "dataflowName": "testmonitor",
    "deployedAt": 1644809910000,
    "undeployedAt": 1644811453000,
    "dataflowVersion": 1,
    "runId": 42559,
  }
]

```

```

        "overall": null,
        "intervals": {},
        "graph": null
    },
    {
        "dataflowId": "1948938e-3923-4602-aba8-f122e3d66faf",
        "dataflowName": "testmonitor",
        "deployedAt": 1644811513000,
        "undeployedAt": 1644838813000,
        "dataflowVersion": 1,
        "runId": 42561,
        "overall": null,
        "intervals": {},
        "graph": null
    }
]

```

## CHAPTER 8

# Monitoring Mass Ingestion Jobs

You can monitor the progress, performance, and status of ingestion jobs from the Mass Ingestion, Monitor, and Operational Insights services.

Depending on the service you use and type of ingestion job, you can view following monitoring information:

- On the **My Jobs** page in the Mass Ingestion service, monitor the ingestion jobs for the ingestion tasks that you deployed. You can view a list of your jobs that includes general job properties such as the task type, runtime environment, start time, duration, and current job state.
- On the **Mass Ingestion** page in either the Monitor or Operational Insights services, monitor *all* types of ingestion jobs that any member of your organization deployed. You can view the following types of information:
  - Summary counts of ingestion jobs by task type and job state.
  - Recent jobs that require your attention because they have errors or warnings.
  - A list of all ingestion jobs by type, including the general job properties.
- From either the list of your jobs or list of all jobs, you can drill down to details for a specific job by clicking the job name. You can view additional job overview information, performance statistics, and information about past job runs.

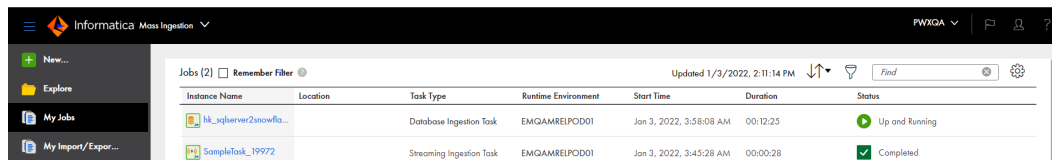
**Note:** Usually, a job name corresponds to the ingestion task name. For application ingestion and database ingestion jobs, the job name has the format *taskname-job\_instance\_number*. The number is incremented each time the job is deployed.

## Monitoring your ingestion jobs

On the **My Jobs** page in Mass Ingestion, you can monitor the ingestion jobs for the tasks that you deployed.

The **My Jobs** page shows information about each job instance, including its status.

For example, the following image shows the **My Jobs** page with a database ingestion job and a streaming ingestion job:



Instance Name	Location	Task Type	Runtime Environment	Start Time	Duration	Status
tk_sqlserver2unowila...		Database Ingestion Task	EMGAMRELPD001	Jan 3, 2022, 3:58:08 AM	00:12:25	Up and Running
SampleTask_19972		Streaming Ingestion Task	EMGAMRELPD001	Jan 3, 2022, 3:45:28 AM	00:00:28	Completed

For descriptions of the columns, see [“Job properties” on page 504](#). These columns are the same as those shown for all ingestion jobs on the **All Jobs** tab of the **Mass Ingestion** page in both the Monitor and Operational Insights services.

To find a job in a long list, use any of the following methods:

- To *sort* the listed jobs, click a column heading or click the Sort icon and select a field to sort by. The default sort order for application ingestion jobs, database ingestion jobs, and streaming ingestion jobs is the time of task deployment, from latest to earliest. The default sort order for file ingestion jobs is the job start time, from latest to earliest.
- To *find* a job based on the job name, enter the job instance name, or any part of the name, in the *Find* text box. With a partial name, the Find operation looks for that particular string anywhere in the instance name. In Mass Ingestion and Operational Insights services, you can include the percent sign (%) wildcard within an instance name search string to represent one or more characters, such as "ing2%798". Do not include the following symbols: question mark (?), number sign (#), or ampersand (&). If you include any of these symbols, the Find operation returns no results.
- To *filter* the list of jobs, click the Filter icon. Then click **Add Field** and enter filter criteria for one or more of the listed fields. For the **Instance Name** field, you can enter the full instance name or part of the name. In Mass Ingestion and Operational Insights services, you can include the percent sign (%) wildcard in the instance name value to represent one or more characters within the name, for example, "vp%test3". Your filter is saved for your user name only, for the current session until you change it. In Mass Ingestion and Operational Insights services, you can save the filter for the subsequent sessions by selecting the **Remember Filter** check box. To clear existing filter criteria, click the Filter icon again.

From the action (...) menu at the right end of each job row, you can perform some actions on the job, depending on the job status and task type.

## Monitoring all ingestion jobs

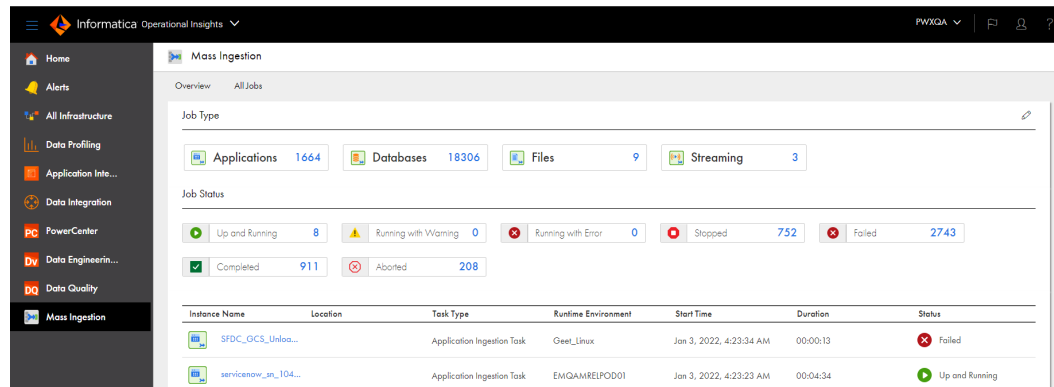
On the **Mass Ingestion** page in both Monitor and Operational Insights services, you can monitor all ingestion jobs that were deployed from the Mass Ingestion service, including application ingestion jobs, database ingestion jobs, file ingestion jobs, and streaming ingestion jobs.

The **Mass Ingestion** page has the following tabs:

- The **Overview** tab displays buttons that you can use to filter the list of ingestion jobs by job type and state.
- The **All Jobs** tab lists all types of ingestion jobs that any member in your organization created and deployed. It includes the same column properties as on the **My Jobs** page in the Mass Ingestion service.

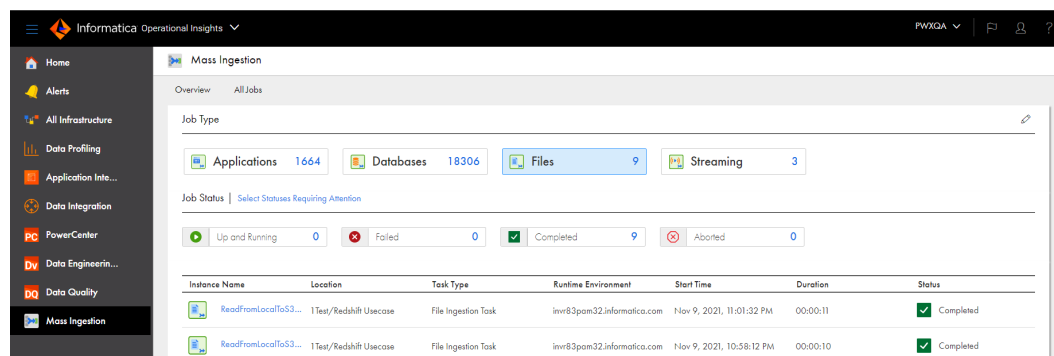
## Overview tab

The **Overview** tab initially lists all types of ingestion jobs. You can use the buttons at the top to filter the jobs by job type and status. Each button shows the number of jobs with that job type or status. For example:



**Note:** For application ingestion jobs and database ingestion jobs, the **Location** column is not populated.

The following example shows the **Overview** tab for file ingestion jobs with all state buttons displayed:



To control the status buttons that appear on the **Overview** tab, click the Edit (pencil) icon. Then in the **Reorder Job Status** dialog box, select the **Visibility** check box next to each job status for which you want to display buttons and jobs. For Mass Ingestion Applications and Mass Ingestion Databases, a subset of the statuses is displayed by default: Up and Running, Running with Warning, Stopped, Failed, Completed, and Aborted.

To rearrange the order of the job status buttons, click the **Edit** icon. Then in the **Reorder Job Status** dialog box, select and drag a job status row up or down.

To filter the list of jobs on the **Overview** tab, use in any of the following methods:

- To see only application ingestion jobs, database ingestion jobs, file ingestion jobs, or streaming ingestion jobs, click the button for a job type. The selected button is highlighted in blue. To see all types of ingestion jobs again, click the selected button again.
- To see the jobs that match a particular job status, click a status button. You cannot select multiple status buttons at the same time on the **Overview** tab. To clear the filter, click the selected status button again.
- To see the jobs with a status that might require your attention, click **Select States Requiring Attention** above the job status buttons. This option is displayed only after you select a job type. This option lists the application ingestion jobs and database ingestion jobs with the **Failed** or **Running with Warning** state, file ingestion jobs with the **Failed** state, or streaming ingestion jobs with the **Running with Error** or **Running with Warning** state. To clear the filter, click **Select States Requiring Attention** again.

**Note:** All filters that you set on the **Overview** tab or in the **Reorder Job Status** dialog box are active only for the current session or until you change them during the session.

## All Jobs tab

The **All Jobs** tab lists all ingestion jobs of any type and with any status.

For example, the following image shows the **All Jobs** tab for application ingestion jobs:

Instance Name	Location	Task Type	Runtime Environment	Start Time	Duration	Status
NetSuite_Comb...	Krishna	Application Ingestion Task	EMQAMREIPOD01	Jan 3, 2022, 4:18:27 AM	00:04:04	Up and Running
servicenow_at_1...	spoorthi/servicenow	Application Ingestion Task	EMQAMREIPOD01	Jan 3, 2022, 4:08:31 AM	00:00:26	Completed
MSD_GBQ_com...	padmini/MSD365	Application Ingestion Task	imr79pom28	Jan 3, 2022, 3:58:08 AM	00:12:25	Failed
CDC_SFDC_Oen...	UpgradeTesting_Jan2022...	Application Ingestion Task	Geet_Linux	Jan 3, 2022, 3:53:10 AM	00:01:49	Failed
servicenow_at_1...	spoorthi/servicenow	Application Ingestion Task	EMQAMREIPOD01	Jan 3, 2022, 3:45:28 AM	00:00:28	Completed

If the list of jobs is long, use any of the following methods to make finding the job easier:

- To **sort** the listed jobs, click a column heading or click the Sort arrows icon and select a field to sort by. The default sort order for application ingestion jobs, database ingestion jobs, and streaming ingestion jobs is the time of task deployment, from latest to earliest. The default sort order for file ingestion jobs is the job start time, from latest to earliest.
- To **find** a job based on the job name, enter the job instance name, or any part of the name, in the *Find* text box. With a partial name, the Find operation looks for that particular string anywhere in the instance name. In Mass Ingestion and Operational Insights services, you can include the percent sign (%) wildcard within an instance name search string to represent one or more characters, such as "ing2%798". Do not include the following symbols: question mark (?), number sign (#), or ampersand (&). If you include any of these symbols, the Find operation returns no results.
- To **filter** the list of jobs, click the Filter icon. Then click **Add Field** and enter filter criteria for one or more of the listed fields. For the **Instance Name** field, you can enter the full instance name or part of the name. In Mass Ingestion and Operational Insights services, you can include the percent sign (%) wildcard in the instance name value to represent one or more characters within the name, for example, "vp%test3". Your filter is saved for your user name only, for the current session until you change it. In Mass Ingestion and Operational Insights services, you can save the filter for the subsequent sessions by selecting the **Remember Filter** check box. To clear existing filter criteria, click the Filter icon again.

From the action (...) menu at the right end of each job row, you can perform some actions on the job, depending on the job status and task type.

# Job properties

The lists of ingestion jobs on the **My Jobs** page in the Mass Ingestion service and on the **All Jobs** tab on the **Mass Ingestion** page in Monitor and Operational Insights display properties for each job. The properties provide a high-level view of the job status.

The following table describes the job properties:

Property	Description
Instance Name	<p>The generated name of the job instance in the following format: <code>&lt;task_name&gt;_&lt;instance_number&gt;</code></p> <p>You can click the instance name to view detailed information about the job.</p> <p><b>Note:</b> If you edit the name of the associated ingestion task, the job name remains the same.</p>
Location	<p>The project or project\subfolder, where the task definition associated with the job exists. For example: <code>Myproject\Oracle</code></p> <p><b>Note:</b> For Mass Ingestion Files and Mass Ingestion Streaming, this property displays a value only for jobs newly deployed in Fall 2020 October and later releases. For Mass Ingestion Databases, this property displays a value only for jobs newly deployed in Fall 2020 November and later releases. This property is blank for any ingestion jobs that were deployed prior to the Fall 2020 releases. For Mass Ingestion Applications, this property is blank.</p> <p>If you move a task definition to another folder, the <b>Location</b> value is not updated.</p>
Task Type	<p>The type of ingestion task. This value must be <b>Application Ingestion Task</b>, <b>Database Ingestion Task</b>, <b>File Ingestion Task</b>, or <b>Streaming Ingestion Task</b>.</p>
Runtime Environment	<p>The name of the runtime environment in which the job runs.</p>
Start Time	<p>For application ingestion and database ingestion jobs, the date and time when the job was deployed.</p> <p>For file ingestion jobs, the date and time when the job started.</p> <p>For streaming ingestion jobs, the date and time when the job was deployed.</p>
Duration	<p>For application ingestion and database ingestion jobs, the amount of time that the job has run since it was deployed. For jobs that are in a Completed, Stopped, Failed, or Aborted state, the amount of time between the date and time the job was deployed and when it acquired its current state.</p> <p>For file ingestion jobs, the amount of time that the job has run.</p> <p>For streaming ingestion jobs, the amount of time that the job has been running.</p>
Status	<p>The current status of the job, such as Deploying, Up and Running, or Undeployed.</p> <p>The set of valid statuses vary by type of ingestion task. For more information, see the "Job Overview tab" section in <a href="#">"Database ingestion job details" on page 511</a> or <a href="#">"Streaming ingestion job details" on page 518</a> or the "Results" section in <a href="#">"File ingestion job details" on page 516</a>.</p>



# Viewing details for an ingestion job

On the **My Jobs** page in the Mass Ingestion service or on the **All Jobs** tab of the **Mass Ingestion** page in both the Monitor and Operational Insights services, you can drill down on a specific ingestion job to display job details.

To view job details, click the job name in the jobs list. A page for the job appears. The details vary by type of ingestion job.

## Application ingestion job details

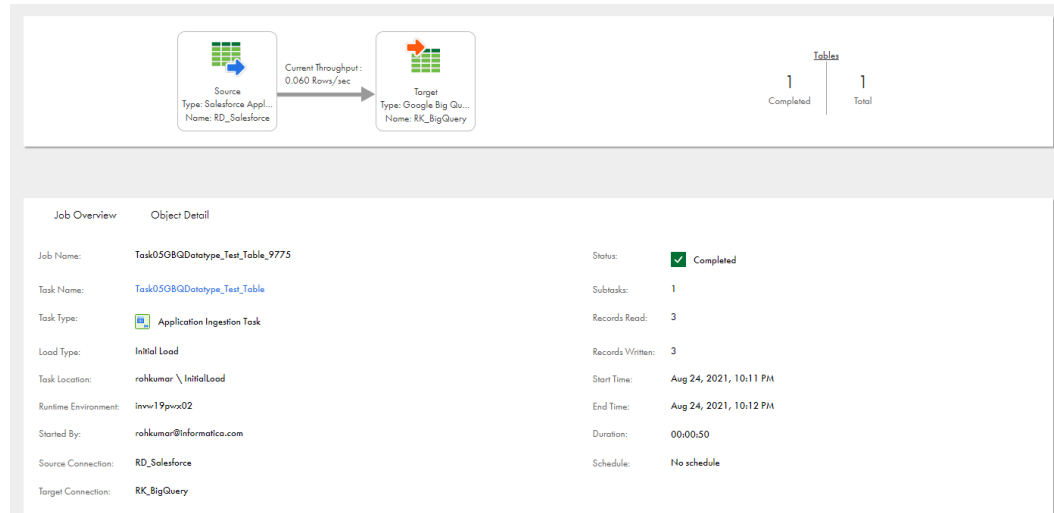
For application ingestion jobs, you can view job-specific details on the **Job Overview**, **Object Detail**, and **Alerts** tabs. Access these tabs by drilling down on a job from the **My Jobs** page in Mass Ingestion or from the **All Jobs** tab on the Mass Ingestion page in either the Operational Insights or Monitor service.

### Job Overview tab

On the **Job Overview** tab, view detailed information for the entire job, including the name of the associated task, the load type, source and target connection names, current status, number of records read and written, start and end times, and run duration. For incremental load jobs and combined initial and incremental load jobs, you can also download the job log.

The following image shows the **Job Overview** tab for a completed application ingestion job:

**Note:** The diagram at the top of the page displays the calculated data throughput, in rows per second, if the job successfully propagated data to the target, regardless of the job's current status. If the calculated value is 0, indicating no data has flowed to the target, the throughput is not displayed.



The following table describes the job overview properties:

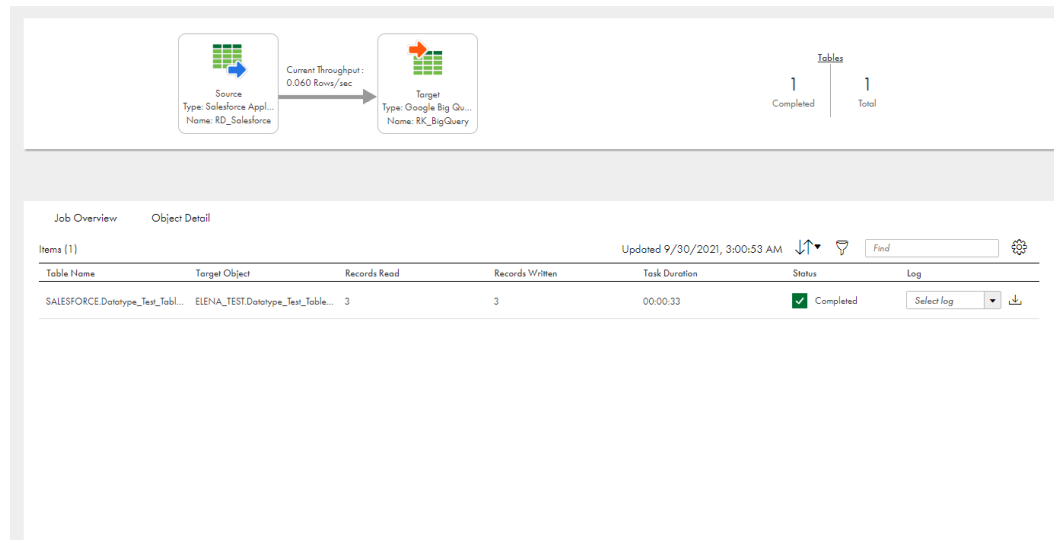
Property	Description
Job Name	The name of the job. Application ingestion job names have format <i>task name-job instance number</i> .
Task Name	The name of the associated ingestion task. You can click the task-name link to view or edit task details in Mass Ingestion, if necessary. If you edit the task, you must redeploy it for the updated task definition to be used for a job.
Task Type	The type of task, which is <b>Application Ingestion Task</b> .
Load Type	The type of load operation that the job performs. Options are: <ul style="list-style-type: none"> <li>- <b>Initial Load</b>. Loads a snapshot of source data read at a specific point-in-time to a target.</li> <li>- <b>Incremental Load</b>. Loads incremental data changes to a target on a continuous basis, until the job is stopped or ends.</li> <li>- <b>Initial and Incremental Load</b>. Performs an initial load and then automatically switches to an incremental load.</li> </ul>
Task Location	The project or project folder that contains the ingestion task definition.
Runtime Environment	The name of the runtime environment that the job uses to run.
Started By	The name of the user who started the job.
Source Connection	The name of the source connection.
Target Connection	The name of the target connection.
Status	The status of the job, which can be one of the following values: <ul style="list-style-type: none"> <li>- <b>Up and Running</b>. The job is running.</li> <li>- <b>Running with Warning</b>. The job is running with a warning. This state can also occur when one or more table-specific subtasks fail but some subtasks are still running.</li> <li>- <b>On Hold</b>. The job is in a paused state while the Mass Ingestion Databases (DBMI) agent is being updated.</li> <li>- <b>Stopping</b>. The job is stopping in response to a Stop request.</li> <li>- <b>Stopped</b>. The job was intentionally stopped.</li> <li>- <b>Failed</b>. The job ended abnormally, the task deployment to the job failed, or one or more object-specific subtasks failed. Also, for an initial load job, the job was stopped.</li> <li>- <b>Deploying</b>. The job is being deployed.</li> <li>- <b>Deployed</b>. The job has been deployed.</li> <li>- <b>Aborting</b>. The job is stopping immediately in response to an Abort request.</li> <li>- <b>Aborted</b>. The job has been aborted.</li> <li>- <b>Undeploying</b>. The job is being undeployed.</li> <li>- <b>Undeployed</b>. The job has been undeployed.</li> <li>- <b>Completed</b>. The job completed successfully.</li> </ul>
Subtasks	The number of subtasks that the application ingestion job used to propagate data from source objects to the target. When a job runs, it uses a separate subtask to process each source object.

Property	Description
Records Read	<p>The number of records that were read from the source.</p> <p><b>Note:</b> The first time you run a job associated with an application ingestion combined initial and incremental load task, the <b>Records Read</b> count might be greater than the total number of object-level DML change records read. This behavior occurs because the initial load portion of combined processing always starts after change data capture begins. As a result, some change records are included in the <b>Records Read</b> count before initial load processing starts and adds more records to the count.</p>
Records Written	<p>For a Microsoft Azure Synapse Analytics target, the number of records written to the intermediate Microsoft Azure Data Lake Storage files.</p> <p>For a Snowflake target, the number of records written to the internal staging area that is created when the job runs.</p>
Start Time	The date and time when the job was deployed.
End Time	The date and time when the job ended because it completed processing, was stopped, or failed. This field is not displayed for running jobs
Duration	The amount of time, in the hh:mm:ss format, that the job ran before it ended.
Log	<p>For incremental load jobs and combined initial and incremental load jobs, you can download the job execution log for the entire job run. Select one of the following log types:</p> <ul style="list-style-type: none"> <li>- <b>Complete Log.</b> The entire log, including all types of messages. It is available for any job that ran, regardless of its state.</li> <li>- <b>Error.</b> The error log, which includes messages only for errors that occurred. It is available for Failed jobs only. Use this log to determine the reason for the job failure, for example, the deployment failed. If the log file ends with an ellipsis (...), the log has been truncated because of its long length. In this case, download the Complete Log to see all error messages.</li> </ul> <p>To download a log to your local system, click the Download icon.</p> <p><b>Note:</b> For initial load jobs, you can get the job log for a specific source object from the <b>Object Detail</b> tab.</p>

## Object Detail tab

On the **Object Detail** tab, view statistics and status information by source object from the last run of an application ingestion job.

The following image shows the **Object Detail** tab for an application ingestion job:



The following table describes the properties that are displayed for each object:

Column	Description
Table Name	<p>The name of the source object or view for which data was propagated to the target.</p> <p>For an incremental load job or a combined initial and incremental load job, click the arrow icon to the left of the object name to display detailed counts of LOBs, Inserts, Deletes, Updates, and DDL statements processed. For a combined initial and incremental load job, the Unload Count field is also displayed to show the number of records that the initial load portion of processing read from the source. The following usage notes apply to the detailed CDC counts:</p> <ul style="list-style-type: none"> <li>- The counts are only for the current job run. If you stop and restart the job, the counts start over from zero. Do not use these counts to identify the number of rows written to the target.</li> <li>- The counts are based on rows read from the source and do not reflect the records written to the target. Target write operations might be optimized by combining operations and reducing the number of physical writes. In this case, the counts might not match the number of write operations.</li> <li>- The value N/A means that the count value is not applicable for the count type or the value has not yet been calculated.</li> <li>- The Unload Count might not reflect the number of source records at the time the job is started or resynchronized because of a delay in the start of unload processing. Between the time of the unload request and start of unload processing, rows might be added to or deleted from the source table.</li> </ul>
Target Object	The name of the target table that is mapped to the source object.
Records Read	For an initial load job, the number of records that were read from the source. For other load types, this information is available only at the job-level on the <b>Job Overview</b> tab.
Records Written	<p>This information is available only at the job-level on the <b>Job Overview</b> tab.</p> <p>For a Microsoft Azure Synapse Analytics target, the number of records written to the intermediate Microsoft Azure Data Lake Storage files.</p> <p>For a Snowflake target, the number of records written to the internal staging area that is created when the job runs.</p>

Column	Description
Task Duration	<p>For an initial load job, the amount of time the subtask that processed the source table ran before it completed or was stopped. For other load types, this information is available only at the job-level on the <b>Job Overview</b> tab.</p> <p>When a job runs, it uses a separate subtask to process each source table.</p>
Stage	<p>For a combined initial and incremental load job, this column shows the stage in the transition from initial load processing to CDC processing for the table-specific job subtask. This column does not appear for other load types.</p> <p>The stage can be one of the following values:</p> <ul style="list-style-type: none"> <li>- <b>Not Started.</b> Initial load processing has not yet started for the table, or if an error occurred and the object is in the <b>Error on Retry</b> state, the next attempt to process the object has not yet started.</li> <li>- <b>Started.</b> Initial load processing has started.</li> <li>- <b>Unloading.</b> The subtask is unloading data from the object as part of initial load processing.</li> <li>- <b>Unloaded.</b> The subtask has finished unloading data from the object as part of initial load processing.</li> <li>- <b>Completed.</b> The subtask completed initial load processing of the object.</li> <li>- <b>Normal.</b> The subtask completed initial load processing of the object and has started CDC processing of the object.</li> <li>- <b>Cancelled.</b> Initial load processing was cancelled or stopped.</li> <li>- <b>Error.</b> The subtask detected an error in the source table.</li> </ul>
Status	<p>The status of the job subtask for the source object.</p> <p><b>Note:</b> If the job stops running, the subtask status reflects the status last collected before the job ended. For example, the job might be aborted but the subtask is in a Running status.</p> <p>The state can be one of the following values:</p> <ul style="list-style-type: none"> <li>- <b>Queued.</b> The subtask has not yet started running.</li> <li>- <b>Starting.</b> The subtask is starting.</li> <li>- <b>Started.</b> For a combined initial and incremental load job, the subtask has started.</li> <li>- <b>Running.</b> The subtask is running.</li> <li>- <b>On Hold.</b> The subtask, as well as the job, is in a paused state while the Mass Ingestion Databases (DBMI) agent is being updated.</li> <li>- <b>Completed.</b> The subtask completed processing successfully.</li> <li>- <b>Stopping.</b> The subtask is stopping in response to a Stop request.</li> <li>- <b>Stopped.</b> The subtask has stopped.</li> <li>- <b>Aborting.</b> The subtask is ending immediately in response to an Abort request.</li> <li>- <b>Aborted.</b> The subtask has been aborted.</li> <li>- <b>Failed.</b> The subtask ended unexpectedly.</li> <li>- <b>Error.</b> The subtask is in error and no longer writing data to the target table. For a combined initial and incremental load job, the subtask might be running and processing incremental change data but no data is being sent to the target.</li> <li>- <b>Error on Retry.</b> An error occurred on the last retry of subtask processing, and now the subtask is waiting to retry processing again.</li> </ul> <p><b>Note:</b> If a DDL change occurs on a source table and then you resume the job, the table subtask state might not change as expected until the first DML operation occurs on the source table.</p>

Column	Description
Log	<p>For initial load jobs, you can download the job execution log for a source object. Select one of the following log types:</p> <ul style="list-style-type: none"> <li>- <b>Complete Log</b>. The complete log for the object subtask from job execution.</li> <li>- <b>Error</b>. The log that contains error messages. This log type is available only for a Failed subtask.</li> </ul> <p>For incremental load jobs, you can get the complete log for the entire job run from the <b>Job Overview</b> tab.</p> <p>For combined initial and incremental load jobs, you can download the <b>Stage Log</b>. This log covers the transition from initial to incremental loading for a source object.</p> <p>To download a log locally, click the Download icon.</p> <p><b>Note:</b> If you undeployed the job, you can download the log for a table only if the associated task has not been deleted.</p>
Actions menu > Resync	<p>For a subtask in a combined initial and incremental load job, if the subtask stage is <b>Normal</b> and the subtask state is any state other than <b>Queued</b> or <b>Starting</b>, the Actions menu is displayed on the right end of the subtask row. From the Actions menu, you can select <b>Resync</b> to resynchronize the source and target objects. For more information, see "Resynchronizing source and target objects" in Mass Ingestion help.</p>

**Note:** This tab shows information on the latest job run. This tab is blank for jobs that have not run or are resuming.

## Alerts tab

On the **Alerts** tab, view alert messages that appear for certain events.

**Note:** The **Alerts** tab displays alert messages when a source schema change is detected. Messages are displayed for all detected schema changes even if you set the schema drift options for the associated task to Ignore.

You can filter the list of alerts based on severity or a date range. To specify a date range, enter one of the following types of values in the **Filter** field:

- A **Custom** date range that consists of the beginning date and time and ending date and time that you select.
- **Any Time** for all stored alerts.
- **Today** for alerts issued today from midnight to 11:59 pm.
- **Last Week**, **Last Month**, or **Last Year** to show alerts from the beginning of last week, month, or year to present.

The following table describes the columns of information that are displayed for each alert message:

Column	Description
Level	Severity level of the alert message.
Code	Alphanumeric code that identifies the alert type.
Details	Description of the event that raised the alert message.
Time	Date and time when the event occurred.

**Note:** You can also configure alert notifications for application ingestion jobs from the **Alerts > Mass Ingestion Alerts** page in Operational Insights. Operational Insights then sends Mass Ingestion alert

notifications to the users you select when database ingestion jobs acquire certain statuses or detect a DDL change.

## Database ingestion job details

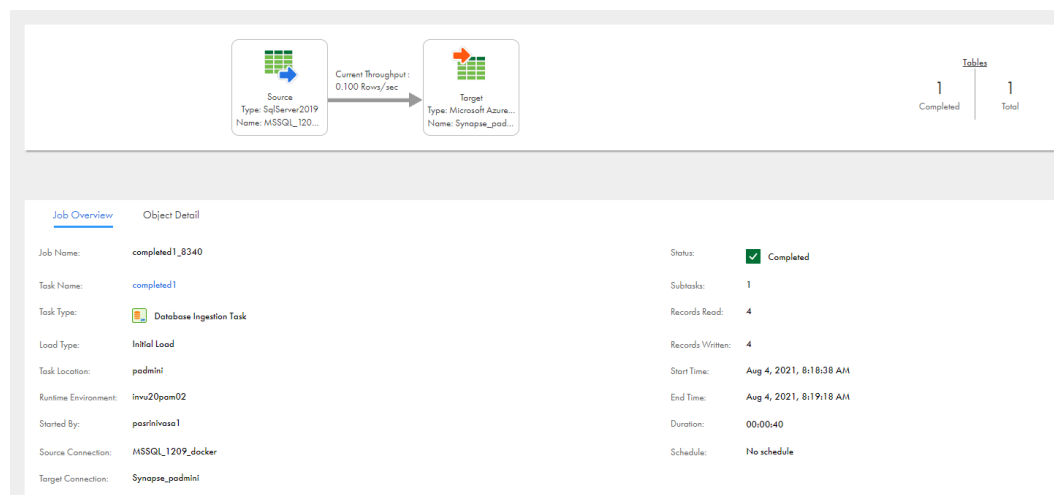
For database ingestion jobs, you can view job-specific details on the **Job Overview**, **Object Detail**, and **Alerts** tabs. Access these tabs by drilling down on a job from the **My Jobs** page in Mass Ingestion or from the **All Jobs** tab on the Mass Ingestion page in either the Operational Insights or Monitor service.

### Job Overview tab

On the **Job Overview** tab, view detailed information for the entire job, including the name of the associated task, the load type, source and target connection names, current status, number of records read and written, start and end times, and run duration. For incremental load jobs and combined initial and incremental load jobs, you can also download the job log.

The following image shows the **Job Overview** tab for a completed database ingestion job:

**Note:** The diagram at the top of the page displays the calculated data throughput, in rows per second, if the job has successfully propagated data to the target, regardless of the job's current status. If the calculated value is 0, indicating no data has flowed to the target, the throughput is not displayed.



The following table describes the job overview properties:

Property	Description
Job Name	The name of the job. Database ingestion job names have format <i>task name-job instance number</i> .
Task Name	The name of the associated ingestion task. You can click the task-name link to view or edit task details in Mass Ingestion, if necessary. If you edit the task, you must redeploy it for the updated task definition to be used for a job.
Task Type	The type of task, which is <b>Database Ingestion Task</b> .

Property	Description
Load Type	The type of load operation that the job performs. Options are: <ul style="list-style-type: none"> <li>- <b>Initial Load.</b> Loads a snapshot of source data read at a specific point-in-time to a target.</li> <li>- <b>Incremental Load.</b> Loads incremental data changes to a target on a continuous basis, until the job is stopped or ends.</li> <li>- <b>Initial and Incremental Load.</b> Performs an initial load and then automatically switches to an incremental load.</li> </ul>
Task Location	The project or project folder that contains the ingestion task definition.
Runtime Environment	The name of the runtime environment that the job uses to run.
Started By	The name of the user who started the job.
Source Connection	The name of the source connection.
Target Connection	The name of the target connection.
Status	The status of the job, which can be one of the following values: <ul style="list-style-type: none"> <li>- <b>Up and Running.</b> The job is running.</li> <li>- <b>Running with Warning.</b> The job is running with a warning. This state can also occur when one or more table-specific subtasks fail but some subtasks are still running.</li> <li>- <b>On Hold.</b> The job is in a paused state while the Mass Ingestion Databases (DBMI) agent is being updated.</li> <li>- <b>Stopping.</b> The job is stopping in response to a Stop request.</li> <li>- <b>Stopped.</b> The job was intentionally stopped.</li> <li>- <b>Failed.</b> The job ended abnormally, the task deployment to the job failed, or one or more table-specific subtasks failed. Also, for an initial load job, the job was stopped.</li> <li>- <b>Deploying.</b> The job is being deployed.</li> <li>- <b>Deployed.</b> The job has been deployed.</li> <li>- <b>Aborting.</b> The job is stopping immediately in response to an Abort request.</li> <li>- <b>Aborted.</b> The job has been aborted.</li> <li>- <b>Undeploying.</b> The job is being undeployed.</li> <li>- <b>Undeployed.</b> The job has been undeployed.</li> <li>- <b>Completed.</b> The job completed successfully.</li> </ul>
Subtasks	The number of subtasks that the database ingestion job used to propagate data from source tables to the target. When a job runs, it uses a separate subtask to process each source table.
Records Read	The number of records that were read from the source. <b>Note:</b> The first time you run a job associated with a database ingestion combined initial and incremental load task, the <b>Records Read</b> count might be greater than the total number of object-level DML change records read. This behavior occurs because the initial load portion of combined processing always starts after change data capture begins. As a result, some change records are included in the <b>Records Read</b> count before initial load processing starts and adds more records to the count.
Records Written	The number of records that were successfully propagated to an Amazon S3, Apache Kafka, flat file, or Microsoft Azure Data Lake Storage target. For a Microsoft Azure Synapse Analytics target, the number of records written to the intermediate Microsoft Azure Data Lake Storage files. For a Snowflake target, the number of records written to the internal staging area that is created when the job runs.

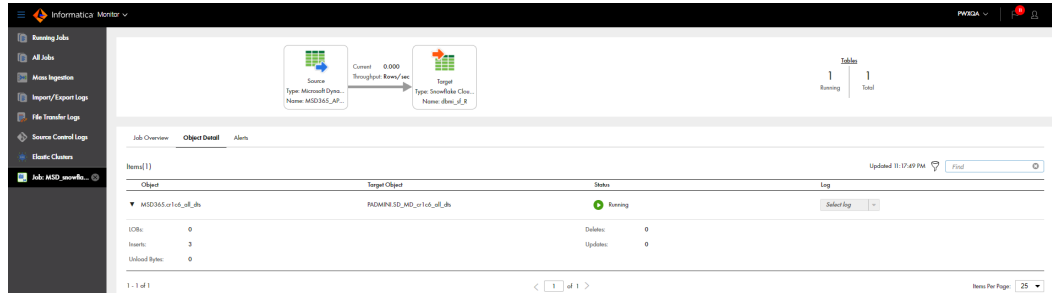


Property	Description
Start Time	The date and time when the job was deployed.
End Time	The date and time when the job ended because it completed processing, was stopped, or failed. This field is not displayed for running jobs
Duration	The amount of time, in the hh:mm:ss format, that the job ran before it ended.
Log	<p>For incremental load jobs and combined initial and incremental load jobs, you can download the job execution log for the entire job run. Select one of the following log types:</p> <ul style="list-style-type: none"> <li>- <b>Complete Log.</b> The entire log, including all types of messages. It is available for any job that ran, regardless of its state.</li> <li>- <b>Error.</b> The error log, which includes messages only for errors that occurred. It is available for Failed jobs only. Use this log to determine the reason for the job failure, for example, the deployment failed. If the log file ends with an ellipsis (...), the log has been truncated because of its long length. In this case, download the Complete Log to see all error messages.</li> </ul> <p>To download a log to your local system, click the Download icon.</p> <p><b>Note:</b> For initial load jobs, you can get the job log for a specific source object from the <b>Object Detail</b> tab.</p>

## Object Detail tab

On the **Object Detail** tab, view statistics and status information by source object from the last run of a database ingestion job.

The following image shows the **Object Detail** tab for a database ingestion job:



The following table describes the properties that are displayed for each object:

Column	Description
Object	<p>The name of the source table or view for which data was propagated to the target.</p> <p>For an incremental load job or a combined initial and incremental load job, click the arrow icon to the left of the object name to display detailed counts of LOBs, Inserts, Deletes, Updates, and DDL statements processed. For a combined initial and incremental load job, the Unload Count field is also displayed to show the number of records that the initial load portion of processing read from the source. The following usage notes apply to the detailed CDC counts:</p> <ul style="list-style-type: none"> <li>- The counts are only for the current job run. If you stop and restart the job, the counts start over from zero. Do not use these counts to identify the number of rows written to the target.</li> <li>- The counts are based on rows read from the source and do not reflect the records written to the target. Target write operations might be optimized by combining operations and reducing the number of physical writes. In this case, the counts might not match the number of write operations.</li> <li>- The value N/A means that the count value is not applicable for the count type or the value has not yet been calculated.</li> <li>- The Unload Count might not reflect the number of source records at the time the job is started or resynchronized because of a delay in the start of unload processing. Between the time of the unload request and start of unload processing, rows might be added to or deleted from the source table.</li> </ul>
Target Object	The name of the target object that is mapped to the source table.
Records Read	For an initial load job, the number of records that were read from the source. For other load types, this information is available only at the job-level on the <b>Job Overview</b> tab.
Records Written	<p>For an initial load job, the number of records that were successfully propagated to an Amazon S3, Apache Kafka, flat-file, or Microsoft Azure Data Lake Storage target. For other load types, this information is available only at the job-level on the <b>Job Overview</b> tab.</p> <p>For a Microsoft Azure Synapse Analytics target, the number of records written to the intermediate Microsoft Azure Data Lake Storage files.</p> <p>For a Snowflake target, the number of records written to the internal staging area that is created when the job runs.</p>
Task Duration	<p>For an initial load job, the amount of time the subtask that processed the source table ran before it completed or was stopped. For other load types, this information is available only at the job-level on the <b>Job Overview</b> tab.</p> <p>When a job runs, it uses a separate subtask to process each source table.</p>
Stage	<p>For a combined initial and incremental load job, this column shows the stage in the transition from initial load processing to CDC processing for the table-specific job subtask. This column does not appear for other load types.</p> <p>The stage can be one of the following values:</p> <ul style="list-style-type: none"> <li>- <b>Not Started.</b> Initial load processing has not yet started for the table, or if an error occurred and the table is in the <b>Error on Retry</b> state, the next attempt to process the table has not yet started.</li> <li>- <b>Started.</b> Initial load processing has started.</li> <li>- <b>Unloading.</b> The subtask is unloading data from the table as part of initial load processing.</li> <li>- <b>Unloaded.</b> The subtask has finished unloading data from the table as part of initial load processing.</li> <li>- <b>Completed.</b> The subtask completed initial load processing of the table.</li> <li>- <b>Normal.</b> The subtask completed initial load processing of the table and has started CDC processing of the table.</li> <li>- <b>Cancelled.</b> Initial load processing was cancelled or stopped.</li> <li>- <b>Error.</b> The subtask detected an error in the source table.</li> </ul>

Column	Description
Status	<p>The status of the job subtask for the source object.</p> <p><b>Note:</b> If the job stops running, the subtask status reflects the status last collected before the job ended. For example, the job might be aborted but the subtask is in a Running status.</p> <p>The state can be one of the following values:</p> <ul style="list-style-type: none"> <li>- <b>Queued.</b> The subtask has not yet started running.</li> <li>- <b>Starting.</b> The subtask is starting.</li> <li>- <b>Started.</b> For a combined initial and incremental load job, the subtask has started.</li> <li>- <b>Running.</b> The subtask is running.</li> <li>- <b>On Hold.</b> The subtask, as well as the job, is in a paused state while the Mass Ingestion Databases (DBMI) agent is being updated.</li> <li>- <b>Completed.</b> The subtask completed processing successfully.</li> <li>- <b>Stopping.</b> The subtask is stopping in response to a Stop request.</li> <li>- <b>Stopped.</b> The subtask has stopped.</li> <li>- <b>Aborting.</b> The subtask is ending immediately in response to an Abort request.</li> <li>- <b>Aborted.</b> The subtask has been aborted.</li> <li>- <b>Failed.</b> The subtask ended unexpectedly.</li> <li>- <b>Error.</b> The subtask is in error and no longer writing data to the target table. For a combined initial and incremental load job, the subtask might be running and processing incremental change data but no data is being sent to the target.</li> <li>- <b>Error on Retry.</b> An error occurred on the last retry of subtask processing, and now the subtask is waiting to retry processing again.</li> </ul> <p><b>Note:</b> If a DDL change occurs on a source table and then you resume the job, the table subtask state might not change as expected until the first DML operation occurs on the source table.</p>
Log	<p>For initial load jobs, you can download the job execution log for a source object. Select one of the following log types:</p> <ul style="list-style-type: none"> <li>- <b>Complete Log.</b> The complete log for the object subtask from job execution.</li> <li>- <b>Error.</b> The log that contains error messages. This log type is available only for a Failed subtask.</li> </ul> <p>For incremental load jobs, you can get the complete log for the entire job run from the <b>Job Overview</b> tab.</p> <p>For combined initial and incremental load jobs, you can download the <b>Stage Log</b>. This log covers the transition from initial to incremental loading for a source object.</p> <p>To download a log locally, click the Download icon.</p> <p><b>Note:</b> If you undeployed the job, you can download the log for a table only if the associated task has not been deleted.</p>
Actions menu > Resync	<p>For a subtask in a combined initial and incremental load job, if the subtask stage is <b>Normal</b> and the subtask status is any status other than <b>Queued</b> or <b>Starting</b>, the Actions menu is displayed on the right end of the subtask row. From the Actions menu, you can select <b>Resync</b> to resynchronize the source and target objects. For more information, see "Resynchronizing source and target objects" in Mass Ingestion help.</p>

**Note:** This tab shows information on the latest job run. This tab is blank for jobs that have not run or are resuming.

## Alerts tab

On the **Alerts** tab, view alert messages that appear for certain events.

**Note:** The **Alerts** tab displays alert messages when a source schema change is detected. Messages are displayed for all detected schema changes even if you set the schema drift options for the associated task to Ignore.

You can filter the list of alerts based on severity or a date range. To specify a date range, enter one of the following types of values in the **Filter** field:

- A **Custom** date range that consists of the beginning date and time and ending date and time that you select.

- **Any Time** for all stored alerts.
- **Today** for alerts issued today from midnight to 11:59 pm.
- **Last Week, Last Month, or Last Year** to show alerts from the beginning of last week, month, or year to present.

The following table describes the columns of information that are displayed for each alert message:

Column	Description
Level	Severity level of the alert message.
Code	Alphanumeric code that identifies the alert type.
Details	Description of the event that raised the alert message.
Time	Date and time when the event occurred.

**Note:** You can also configure alert notifications for database ingestion jobs from the **Alerts > Mass Ingestion Alerts** page in Operational Insights. Operational Insights then sends Mass Ingestion alert notifications to the users you select when database ingestion jobs acquire certain statuses or detect a DDL change.

## File ingestion job details

The job results for each file ingestion task instance display the status of the job, and success and error statistics.

To view detailed information about a file ingestion task, click the task name on the **My Jobs** page in Mass Ingestion or on the **All Jobs** tab of the Mass Ingestion page in both the Monitor and Operational Insights services.

You can download the job. The following image shows the details of a file ingestion job:

The screenshot displays the Informatica Operational Insights interface. On the left is a navigation sidebar with options: Home, Alerts, All Infrastructure, Application Integ..., Data Integration, Mass Ingestion, and a selected job 'adlgen2\_to\_fps...'. The main content area is titled 'adlgen2\_to\_fps\_1SatJun122021094156-124505'. It is divided into 'Job Properties' and 'Results' sections.

**Job Properties:**

- Task Name: adlgen2\_to\_fps\_1SatJun122021094156
- Task Type: File Ingestion Task
- Started By: filelistener
- Start Time: Jun 12, 2021, 10:31:16 AM
- End Time: Jun 12, 2021, 10:31:31 AM
- Duration: 00:00:15

**Results:**

- State: ✔ Success
- Session Log: [Download Session Log](#)
- Success Files: 14
- Error Files: 0
- Duplicate Files: 0
- Error Message: Job completed normally

Below these sections is a table titled 'File Events (14)' with columns: Name, File Size(Bytes), Status, Transfer Type, Start Time, Duration (ms), and Remarks. The table lists 14 file events, all with a 'Success' status.

Name	File Size(Bytes)	Status	Transfer Type	Start Time	Duration (ms)	Remarks
File7.txt	26	Success	FTPS Upload	Jun 12, 2021, 10:31:30 AM	358	
File6.txt	32	Success	FTPS Upload	Jun 12, 2021, 10:31:30 AM	265	
File7.txt	26	Success	ADLS Download	Jun 12, 2021, 10:31:29 AM	929	
File6.txt	32	Success	ADLS Download	Jun 12, 2021, 10:31:29 AM	928	
File5.txt	35	Success	FTPS Upload	Jun 12, 2021, 10:31:28 AM	297	
File3.txt	45	Success	FTPS Upload	Jun 12, 2021, 10:31:27 AM	303	

### Job Properties

The job properties for the file ingestion task instance display general properties about the instance.

The following table describes the job properties:

Property	Description
Task Name	The name of the associated ingestion task. You can click the task-name link to view or edit task details in Mass Ingestion.
Task Type	Task type. In this case, file ingestion task.
Started By	Name of the user or schedule that started the job.
Start Time	Date and time when the job was started.
End Time	Date and time when the job completed or stopped.
Duration	The amount of time the job ran before it completed or was stopped.

## Results

The job results for the file ingestion task instance display the status of the job and error statistics.

The job results include the following properties:

Property	Description
State	Job status. A job can have one of the following statuses: <ul style="list-style-type: none"><li>- <b>Running</b>. The job is still running.</li><li>- <b>Success</b>. The job completed successfully.</li><li>- <b>Failed</b>. The job did not complete because it encountered errors</li><li>- <b>Aborted</b>. The job was aborted.</li></ul>
Session Log	Allows you to download the session log file. By default, Informatica Intelligent Cloud Services stores session logs for 10 runs before it overwrites the logs with the latest runs. If you need the session logs for earlier runs, take a backup of the directory that holds the session log files. Session log files are written to the following directory: <Secure Agent installation directory>/apps/Data_Integration_Server/logs
Success Files	Number of files that are successfully transferred, downloaded, and uploaded to the target.
Error Files	Number of files that were not transferred to the target.
Duplicate Files	Number of files that were identified as duplicates.
Error Message	Error message, if any, that is associated with the job.

## File Events

This section shows the total number of files that the file ingestion task has transferred with information about each file.

The File Events section is updated each time the file ingestion task transfers a file, and the state of the file updates throughout the file transfer process. You can track the progress of a file transfer based on the state of the file.

The File Events section displays the following properties for each file:

Property	Description
Name	The name of the file.
File size	The size of the file in bytes.
Status	<p>The status of the file transfer. A file can have one of the following status:</p> <ul style="list-style-type: none"><li>- <b>Success</b>. The file transfer completed successfully.</li><li>- <b>Failed</b>. The file transfer did not complete because it encountered errors.</li><li>- <b>Processing</b>. The file transfer is still running.</li><li>- <b>Duplicate</b>. The task previously transferred a file with the same name, directory location, and size.</li><li>- <b>Interrupted</b>. The file transfer is interrupted because of network issues or changed server credentials during the file transfer. Run the file ingestion job to resume the transfer of the interrupted files. This status is applicable when the file ingestion task transfers file from the advanced FTP, advanced SFTP, or advanced FTPS sources.</li><li>- <b>In Doubt</b>. The previous task instance encountered errors while transferring the file. Applicable for tasks where the source is configured to skip duplicate files.</li><li>- <b>Quarantined</b>. The task marks any infected file it detects from a source as quarantined.</li></ul> <p>You can monitor the Status property to track the progress of the file transfer of each file.</p>
Transfer Type	<p>The type of file transfer. A file can have one of the following transfer types:</p> <ul style="list-style-type: none"><li>- <code>&lt;source_name&gt;Download</code>. The file is downloaded from source. <code>&lt;source_name&gt;</code> is the name of the source.</li><li>- <code>&lt;target_name&gt;Upload</code>. The file is uploaded to the target. <code>&lt;target_name&gt;</code> is the name of the target.</li><li>- Copy from Source. The file ingestion task is performing file processing actions.</li><li>- Copy to Target. The file is transferred from a local directory to a local directory.</li></ul>
Start time	Date and time when the file transfer started.
Duration	The length of time to transfer the file, in milliseconds.
Remarks	Applies to file events in Failed status. The message includes the reason for failure of the event based on the file transfer type.

## Streaming ingestion job details

To view detailed information about a streaming ingestion job, click the job name on the **My jobs** page in Mass Ingestion or on the **All Jobs** tab of the Mass Ingestion page in both the Monitor and Operational Insights services.

### Overview tab

The **Overview** tab displays general properties of the job. You can download the job log, too.

The following image shows the **Overview** tab for a streaming ingestion job:

Kafka To Kafka With Filter. Update Every: 30 sec. Last updated on Oct 1, 2020, 04:25 PM

Source (0 Events) → Filter → Segregator → Target (0 Events)

Overview | Alert | Performance | Past Run

Job Name	Kafka To Kafka With Filter	State	Running with Error
Version	3	Duration	1 Days 20:45:14
Task Type	Streaming Ingestion Task	Start Time	Sep 29, 2020, 07:40 PM
Task Location	Default	Runtime Environment	INWPF28BH8X-AAD
Started By	siqa_test	Download Log	Complete
Secure Agent	INWPF28BH8X-AAD		

The following table describes the job overview properties:

Property	Description
Job Name	The name of the job.
Version	The version number of the job.
Task Type	The task type of streaming ingestion task.
Task Location	The project or project folder that contains the streaming ingestion task.
Started By	The name of the user who deployed the job.
Secure Agent	The location where the Secure Agent is running. A warning symbol near the Secure Agent indicates that the Secure Agent is either offline or not reachable.
State	The state of the job. A job can have one of the following states: <ul style="list-style-type: none"> <li>- Deploying. The job is being deployed.</li> <li>- Up and Running. The job is running.</li> <li>- Running with Warning. The job is running with warnings.</li> <li>- Running with Error. The job is running with error. If a job continuously runs with warnings for seven minutes or for the time specified in the runtime option, the state of the job changes to Running with Error.</li> <li>- Undeployed. The job is undeployed.</li> <li>- Stopped. The job was intentionally stopped.</li> </ul>
Duration	Total time the job ran before it is undeployed. The total time is shown in hh:mm:ss format.
Start Time	The date and time when the job was deployed.

Property	Description
Runtime Environment	Name of the runtime environment that the job uses to run.
Download Log	<p>Level of log that you want to download for a running job.</p> <p>You can download one of the following logs:</p> <ul style="list-style-type: none"> <li>- Complete. The entire log, including all types of messages. It is available for any job that ran, regardless of its state.</li> <li>- Latest. Latest version of the log.</li> </ul> <p>To download a log to your local system, click the <b>Download</b> icon.</p>

## Alert tab

The **Alert** tab displays the alert messages when an event occurs.

The following image shows the **Alert** tab for a streaming ingestion job:

The screenshot shows the user interface for a streaming ingestion job named "KafkaToFlatfileTargetWithFormatConverterWith...". At the top, a diagram illustrates the data flow from a "Source" (represented by a green grid icon) to a "FormatConverter" (represented by a green double-headed arrow icon). Below the source icon, it indicates "499021 Events".

Below the diagram, there are four tabs: "Overview", "Alert", "Performance", and "Past Run". The "Alert" tab is currently selected, displaying a list of alert messages. Each message starts with a yellow warning icon and contains the following text: "KafkaToFlatfileTargetWithFormatConverterWithXML\_FormatConverter : ConvertRecord[id=6ee4a30e-9ed9-4e00-87bf-d6cc0e0...". To the right of each message is a blue link labeled "Show M...".



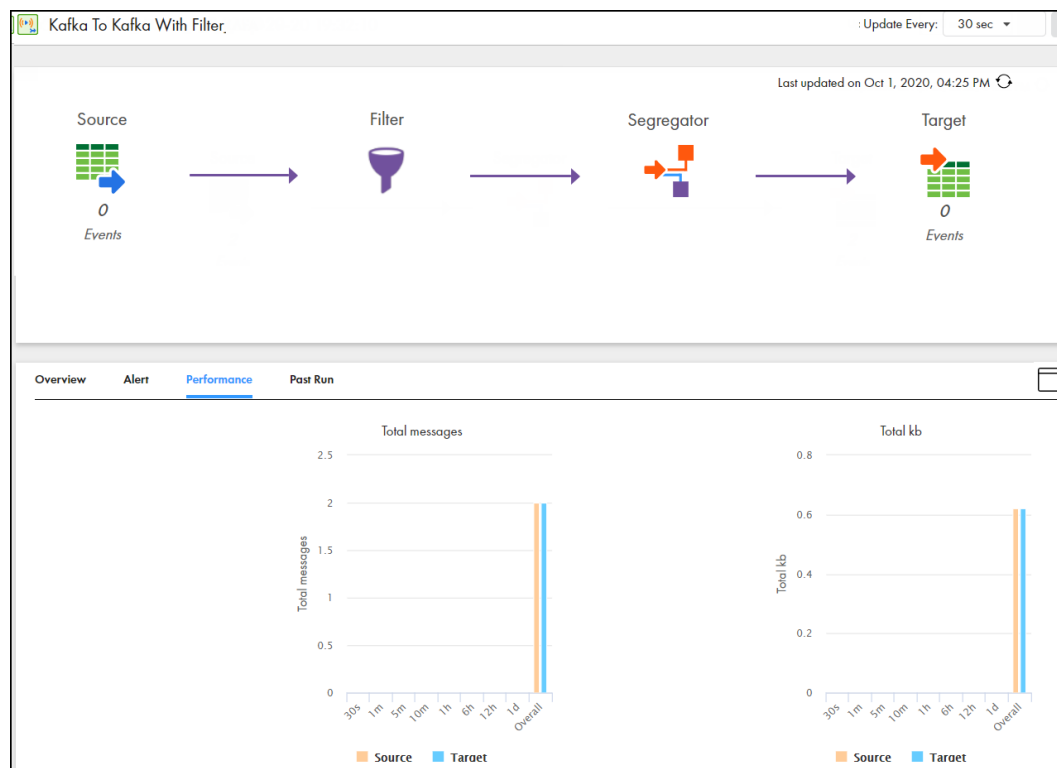
The following table describes the job alert properties:

Property	Description
Alert	The messages or a group of messages that the job returns when a deployed job encounters a warning.
Time	The date and time when the event occurred.

## Performance tab

The **Performance** tab displays graphs of throughput information for the source and target of the job.

The following image shows the **Performance** tab for a streaming ingestion job:



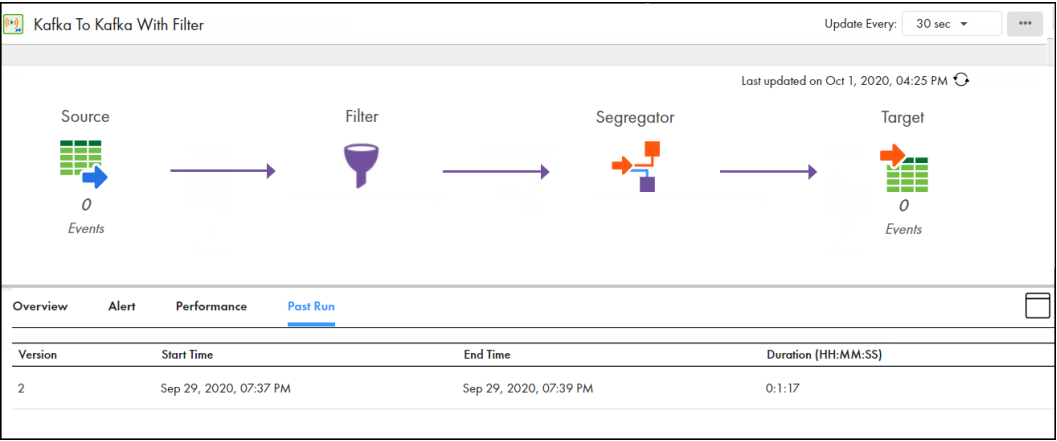
The following table describes the job performance properties:

Property	Description
Total messages	The average number of messages streamed per second.
Total kb	The average kilobits of messages streamed per second.

## Past run tab

The **Past Run** tab displays the statistics and status information related to the previous runs of a streaming ingestion job.

The following image shows the **Past Run** tab for a streaming ingestion job:



The following table describes the past run properties:

Column	Description
Version	The version number of the job.
Start Time	The date and time when the job was deployed.
End Time	The date and time when the job was undeployed.
Duration	Total time the job ran before it is undeployed. The total time is shown in hh:mm:ss format.

## CHAPTER 9

# Asset Management

You can manage Mass Ingestion assets such as projects, folders, and ingestion tasks from the **Explore** page.

You can perform the following management tasks depending on your user role and permissions:

- Edit ingestion tasks.
- Copy projects, folders, and ingestion tasks.
- Move folders and ingestion tasks.
- Rename projects, folders, and ingestion tasks.
- Delete projects, folders, and ingestion tasks.
- Apply tags so you can filter assets on the **Explore** page.
- Configure user permissions for projects, folders, or assets.
- Use source control to manage versions of projects, folders, and tasks.
- Migrate assets between organizations.

## Editing ingestion tasks

You can edit an ingestion task from the **Explore** page.

1. In Mass Ingestion, open the **Explore** page.
2. If a list of projects is displayed, select the project or project folder that contains the ingestion task that you want to edit.
3. In the list of tasks, select the row for the ingestion task that you want to edit.
4. In the Actions menu for the selected row, click **Edit**.  
The **Definition** page of the task wizard appears in edit mode.
5. Edit the available fields on the **Definition**, **Source**, **Target**, and **Runtime Options** pages.
6. When you are finished, click **Save**.

# Copying projects, folders, and tasks

You can copy projects, folders, and tasks on the **Explore** page. You might want to copy an object to use as a template, or you might want to create a backup copy.

When you copy a project, the new project contains all of the folders and tasks that were in the original project. Similarly, when you copy a folder, the new folder contains all of the tasks that were in the original folder.

When you copy a task within the folder where the task exists, you have the option to keep both tasks or cancel the operation. When you copy a task into a different folder that contains a task with the same name, you have the option to keep both tasks, overwrite the task in the folder, or cancel the operation. If you choose to keep both tasks, Informatica Intelligent Cloud Services appends the new task name with "Copy x" where x is the sequential copy number.

**Note:** To avoid naming conflicts with duplicate tasks, rename assets with a "Copy x" suffix.

1. On the **Explore** page, navigate to the object that you want to copy.
2. In the list of assets, select one or more rows for the objects that you want to copy. Then perform one of the following actions:
  - If you selected a single asset row, in the Actions menu for the row, click **Copy To**. Alternatively, right-click the row and click **Copy To**.
  - If you selected multiple asset rows, right-click a highlighted row and click **Copy To** to copy all of the assets.
3. Browse to the new location and click **Select**.

# Moving folders and tasks

You can move folders and tasks on the **Explore** page.

1. On the **Explore** page, navigate to the folder or task that you want to move.
2. In the list of assets, select one or more rows for the ingestion tasks or folders that you want to copy. Then perform one of the following actions:
  - If you selected a single row, in the Actions menu for the row, click **Move To**. Alternatively, right-click the row and click **Move To**.
  - If you selected multiple rows, right-click a highlighted row and click **Move To** to copy all of the assets.
3. Browse to the new location and click **Select**.

# Renaming projects and folders

You can rename projects and folders.

The following characters cannot be used on the **Explore** page:

# ? ' | { } " ^ & [ ] / \

Do not use these characters in project, folder, asset, or tag names.

1. On the **Explore** page, navigate to the project or folder that you want to rename.
2. In the row that contains the project or folder, click **Actions** and select **Properties**.
3. Enter the new name and click **Save**.

You cannot use special characters in a name or use the same name as another object that is in the same folder.

## Renaming database ingestion tasks

You can rename database ingestion tasks.

The following characters cannot be used on the **Explore** page:

# ? ' | { } " ^ & [ ] / \

Do not use these characters in project, folder, asset, or tag names.

1. On the **Explore** page, select the row for the ingestion task that you want to rename.
2. In the Actions menu for the selected row, click **Edit**.  
The **Definition** page of the task wizard appears in edit mode.
3. In the **Name** field, enter the new name and click **Save**.

If you rename a database ingestion task, the task name and job name will no longer be synchronized. If you want the task name to correspond to the job name, undeploy the job and then deploy the task again. Any log history will be lost.

## Renaming file ingestion tasks

You can rename file ingestion tasks.

The following characters cannot be used on the **Explore** page:

# ? ' | { } " ^ & [ ] / \

Do not use these characters in project, folder, asset, or tag names.

1. On the **Explore** page, navigate to the task that you want to rename.
2. To rename a task, in the row that contains the task, click **Actions** and select **Rename**.
3. Enter the new name and click **Save**.

You cannot use special characters in a name or use the same name of another task that is in the same folder.

# Renaming streaming ingestion tasks

You can rename streaming ingestion tasks.

The following characters cannot be used on the **Explore** page:

# ? ' | { } " ^ & [ ] / \

Do not use these characters in project, folder, asset, or tag names.

1. On the **Explore** page, navigate to the task that you want to rename.
2. To rename a task, in the row that contains the task, click **Actions** and select **Rename**.
3. Enter the new name and click **Save**.

You cannot use special characters in a name or use the same name as another task that is in the same folder.

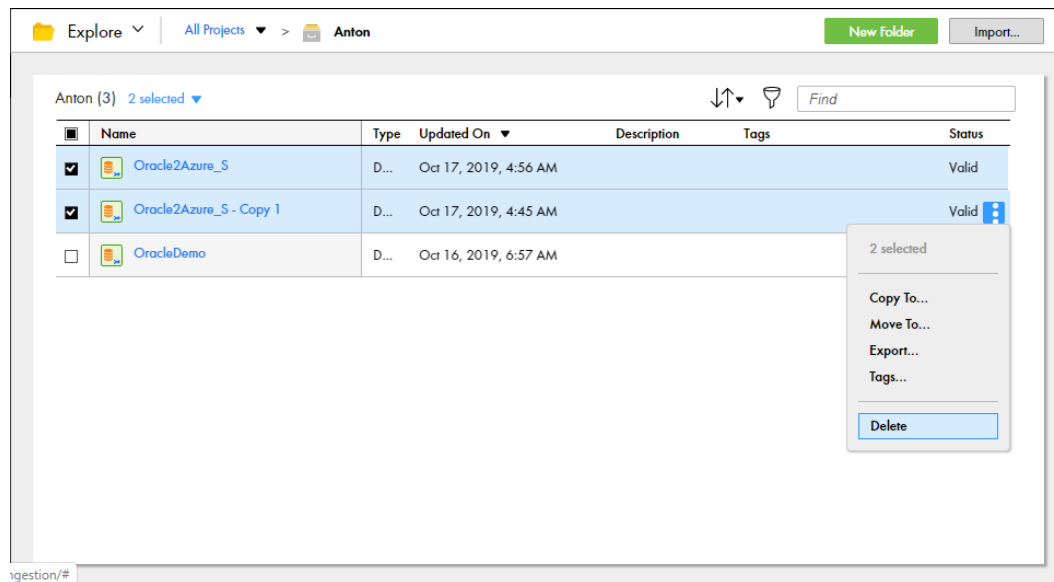
# Deleting projects, folders, and tasks

You can delete a project, folder, or task if you no longer need it. However, before you delete it, verify that no users in the organization plan to use it. You cannot retrieve projects, folders, or tasks after you delete them.

You cannot delete an asset in the following situations:

- The asset is a task that is associated with a deployed job.  
**Note:** In Mass Ingestion Databases, if you undeploy the job, you can then delete the task.
- The asset is a folder that contains tasks.

Delete a project, folder, or task from the **Explore** page, as shown in the following image:



1. To delete a project, folder, or task, on the **Explore** page, navigate to the object that you want to delete.
2. In the list of assets, select one or more rows for the ingestion tasks or folders that you want to delete. Then perform one of the following actions:

- If you selected a single row, in the Actions menu for the row, click **Delete**. Alternatively, right-click the row and click **Delete**.
- If you selected multiple rows, right-click a highlighted row and click **Delete** to delete all of the objects.

A confirmation message appears.

3. Click **Delete** again.

The selected objects are permanently deleted. You cannot use them again.

## Tags

A tag is an asset property that you can use to group assets. Create tags to filter for assets that share a common attribute on the **Explore** page.

For example, your regional offices manage the assets that only apply to their region. Each of your organization's assets includes a tag that identifies the asset by region. You want to view all of the assets that the Southwest regional office manages. On the **Explore** page, you explore by tag and then click the SW Region tag, as shown in the following image:

Explore ▾ | All Tags ▾ > SW Region

SW Region (2)					
<input type="checkbox"/>	Name	Description	Type	Location	Updated On
<input type="checkbox"/>	m_RegionTotalNew		Mapping	Accounts\Febbruary2018	Mar 29, 2018, 6:06 PM
<input type="checkbox"/>	m_TotalMonthly		Mapping	Accounts\Febbruary2018	Mar 29, 2018, 6:06 PM

You can assign tags to all asset types. An asset can have up to 64 tags.

You can find all of the assets that have a particular tag using one of the following methods:

- Click the name of the tag in the **Tags** column, in any row.
- Explore by tag, and then in the list of tags that shows on the page, click the name of the tag.

The following image shows an **Explore** page that lists all the tags created for the organization:

Explore ▾ | All Tags ▾

All Tags (3)			
<input type="checkbox"/>	Name	Asset Count	Updated On
<input type="checkbox"/>	NE Region	1	Mar 29, 2018, 6:44 PM
<input type="checkbox"/>	NW Region	2	Mar 29, 2018, 7:29 PM
<input type="checkbox"/>	SW Region	2	Mar 29, 2018, 6:48 PM

Click the name of a tag to see a list of all the assets associated with the tag.

## Creating tags

You can create multiple tags to assign to assets.

You can create tags that you want to use for an asset when you configure the asset properties, or you can create multiple tags to be available for future use. To create multiple tags for future use, you use an asset's Properties dialog box.

Follow this procedure if you want to create multiple tags without assigning them to an asset.

1. On the **Explore** page, browse by asset type.
2. In a row that contains an asset, click **Actions** and select **Properties**.
3. In the **Tags** field, enter the name of a tag that you want to create, and then press Enter.

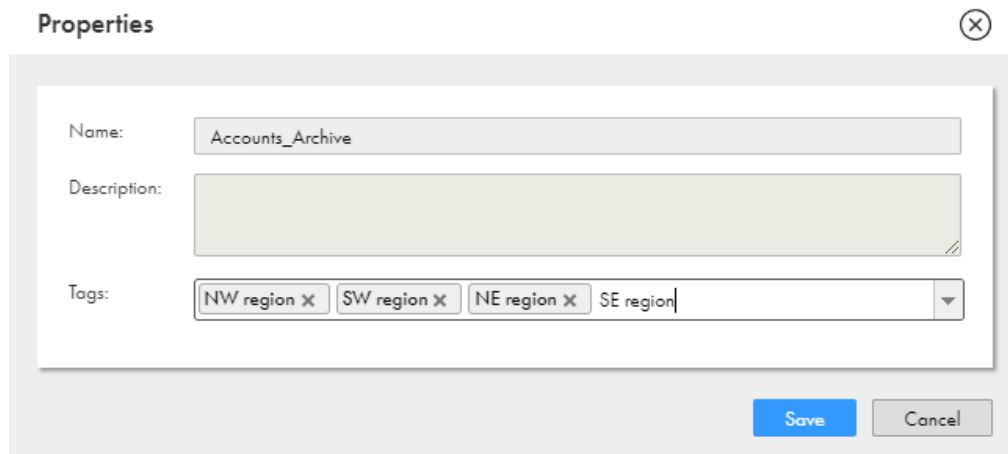
A tag can have a maximum of 255 characters.

The following characters cannot be used on the **Explore** page:

# ? ' | { } " ^ & [ ] / \

Do not use these characters in project, folder, asset, or tag names.

4. Continue to enter the desired tags. Press Enter after each tag name to add it to the tag list.



5. After you have entered the tags, delete the tags from the **Tags** field so that the asset does not become associated with the tags. The tags will still appear in the list of available tags.
6. Click **Save**.

## Assigning tags to an ingestion task

After you define an ingestion task, you can assign tags to the task. You can then filter the tasks on the **Explore** page based on one or more tag names.

1. In Mass Ingestion, open the **Explore** page.
2. If a list of projects is displayed, select the project or project folder that contains the ingestion task to which you want to assign one or more tags.
3. In the **All Assets** or **All Assets > ingestion task type** view, select the row for the ingestion task.
4. In the Actions menu for the selected row, click **Properties**. Alternatively, right-click the row and select **Properties**.

The **Properties** dialog box for the task appears.

5. In the **Tags** list, select one or more existing tags or enter a new tag name.
6. Click **Save**.

After you assign tags to ingestion tasks, you can find all ingestion tasks that have a particular tag name from the **Explore** page. Click **Explore > Tags**. Then in the list of tags, click a tag name to see a list of the tasks with that tag. You can drill down on a task name to see the task details.

Alternatively, in a list of task assets, click a tag name in the **Tags** column.



## Editing and deleting tags

You can edit or delete a tag on the **Explore** page.

Edit a tag name or description in the tag properties. When you edit a tag, the properties for associated assets update as well. For example, if your m\_sales asset has the NorthWest tag and you change the name of the tag to NW, the name of the tag changes to NW in the m\_sales asset properties.

If you delete a tag, the tag no longer appears in the asset properties.

1. On the **Explore** page, browse by tags.
2. In the row that contains the tag, perform one of the following tasks:
  - To edit a tag, click **Actions** and select **Edit**. After you make your changes, click **Save**.
  - To delete a tag, click **Actions** and select **Delete**.

## Asset dependencies

You can view object dependencies for an asset. You might want to view object dependencies before performing certain operations on an asset.

For example, you cannot delete an asset if another object depends on the asset. You must first delete the dependent objects and then delete the asset. You can find the dependent objects by viewing the asset dependencies.

You can view object dependencies for an asset on the **Explore** page. To view object dependencies for an asset, in the row that contains the asset, click **Actions** and select **Show Dependencies**. The **Dependencies** page opens showing the **Uses** tab by default.

The **Uses** tab lists the objects that the selected asset uses.

A Mass Ingestion service task uses connections and a runtime environment.

**Note:** If a database ingestion task was last saved prior to the Mass Ingestion Spring 2020 April release, you must save the task again before you try to view object dependencies for it the first time. You need to perform this action only once. If the Save button is not available, make a minor edit to the task

The **Used By** tab lists the objects that use the selected asset.

To drill down to the lowest level dependency, you can continue to show dependencies for each asset that appears on the **Dependencies** page. At the top of the **Dependencies** page, a breadcrumb shows the chain of dependencies.

The following image shows that the asset mt\_FilterArchCustRecords is dependent on m\_FilterCustRecords, which is dependent on FF\_USW1PF:



Uses		Used By		
Uses (1)				
<input type="checkbox"/>	Name	Type	Location	Updated By
<input type="checkbox"/>	m_FilterCustRecords	Runtime Environment	/tmp/mo	

If you have the appropriate permissions, you can perform actions on the **Dependencies** page such as viewing or deleting assets. To view or delete an asset, in the row that contains the asset, click **Actions** and select the action.

If you work with source controlled assets, you can view source control information such as the last pull time and the last check-in. To view source control information, you can add the following source control columns to the table:

- Last Pull Time
- Checked Out By
- Last Check in
- Git Hash

## Configuring user permissions on an ingestion task

You can configure permissions for an ingestion task if you are assigned a user role that has the **Set Permission** privilege for the asset type of Mass Ingestion Streaming Task.

Typically, the organization administrator assigns user roles to specific users of the Mass Ingestion service.

1. In Mass Ingestion, open the **Explore** page and navigate to the row for the ingestion task for which you want to set permissions.
2. In the Actions menu for the row, select **Permissions**.

The **Permissions** dialog box lists the users and user groups that have permissions set on the task. Other users cannot access the task.

If the **Permissions** dialog box lists no users or user groups, no permissions are configured for the task. In this case, any user can access the task without permission restrictions.

3. To add a user to the users list and grant permissions on the task to that user, perform the following steps:
  - a. On the **Users** tab, click **Add**.
  - b. In the **Add User** dialog box, select a user and click **Add**.
  - c. In the **Permissions** dialog box, select the permissions on the task that you want to grant to the user.
  - d. Click **Save**.
4. To add a user group to the groups list and grant permissions on the task to that user group, perform the following steps:
  - a. On the **Groups** tab, click **Add**.
  - b. In the **Add Group** dialog box, select a group and click **Add**.

If no groups are listed, no user groups are defined.
  - c. In the **Permissions** dialog box, select the permissions on the task that you want to grant to the group.
  - d. Click **Save**.
5. To edit the permissions that are set for a listed user or user group, on the **Users** or **Groups** tab in the **Permissions** dialog box, select or clear the permission check boxes for the user or group. Then click **Save**.
6. To remove all of the permissions that you set for one or more users or user groups, on the **Users** or **Groups** tab in the **Permissions** dialog box, select the users or groups from which you want to remove all permissions. Then click **Remove** and click **Save**.

# Asset migration

You can migrate Informatica Intelligent Cloud Services assets from one organization to another organization. To migrate assets, you export the assets from the source organization and then import the assets into the target organization.

You can import and export tasks and their dependent objects.

**Note:** If you want to export database ingestion tasks for which you cannot view object dependencies, you must save the tasks again before you try to export them. Otherwise, when you import the tasks into the target, any asset overrides that you specified for the import operation, such as an override connection or Secure Agent, will not be applied to the imported tasks. You need to perform this action only once for a task.

You can export single assets, groups of assets, or export all of the assets in a project. If you export a project or folder, the file structure remains intact so that when you perform the import in the target organization, you can duplicate the original structure.

Before you migrate an asset from one organization to another, ensure that you meet the following requirements:

- You have a user account in the source and target organizations with a role that has import and export privileges, such as the Admin or Designer role.
- The source and target organizations have the required license to import and export assets.
- The target organization has the required licenses for the assets that you want to import.
- The target organization uses the same version or a newer version of Informatica Intelligent Cloud Services. The versions might differ temporarily if the organizations aren't on the same POD (Point of Delivery) during an Informatica Intelligent Cloud Services upgrade.

To export or import assets in a sub-organization, log in to the sub-organization. If you have administrator privileges in the parent organization, you can also switch to the sub-organization and export or import assets.

## Dependent objects

Dependent objects are assets that are required by other assets.

When you set up an export, you have the option to include or exclude dependent objects in the export file. The dependent objects must exist either in the export file or in the target organization, else the import fails.

You might want to include dependent objects if they do not exist in the target organization. Or, you might want to include dependent objects if you want to replace the dependent objects in the target organization with updated versions from the source organization. If you choose to include dependent objects, the export file includes dependent objects for all of the assets that you include in the export. When you configure the import, you can choose which dependent assets to import.

You might want to exclude an asset's dependent objects if the objects exist in the target organization and you do not want to replace them.

**Note:** Schedules are not dependent objects and are not included when you export assets that use them.

## Runtime environments and connections

Runtime environments and connections are dependent objects.

If you configure an export to include dependent objects, you can use the source connections and runtime environments in the export file, or you can select connections and runtime environments in the target organization.

If you configure an export to exclude dependent objects, be sure that a suitable connection and runtime environment for the assets exists in the target organization. If a dependent connection or runtime environment does not exist in the target organization, during the import operation you must select a connection or runtime environment in the target organization.

When you select a connection or runtime environment in the target organization during the import, the connector type and version must be the same as the connector type and version that the asset used in the source organization.

If the target organization has connections or runtime environments with the same name as those in the export file, Informatica Intelligent Cloud Services uses the connections or runtime environments that exist in the target organization. Informatica Intelligent Cloud Services does not overwrite the connections or runtime environments in the target organization.

**Note:** An export or import cannot include a Cloud Hosted Agent or shared agent. If an asset uses a Cloud Hosted Agent or a shared agent, you can select a runtime environment to use for the asset during import.

## Schedules

You can migrate schedules from one organization to another organization. You might want to migrate a schedule if you migrate an asset that uses it.

When you export an asset that uses a saved schedule, the schedule is not included in the export file.

To migrate a schedule, you export the schedule from the source organization using Administrator and import the schedule into the target organization using the service that will use the schedule. For example, to migrate a schedule that's used for Data Integration mapping tasks, you export the schedule from the source organization using Administrator and import the schedule into the target organization using Data Integration.

For information about exporting schedules, see the Administrator help. For information about importing schedules, see ["Importing assets" on page 536](#).

## Asset export

When you export assets, Informatica Intelligent Cloud Services creates an export ZIP file that contains the assets that you selected for export.

You can select individual assets to export, or you can select an entire project or folder. When you export a project or folder, the export file includes all of the assets in the project or folder.

To export an asset, you need the following privileges and permissions:






- Your user role must have privileges to export assets.
- You must have read permission on the asset.

**Note:** Informatica recommends that you include no more than 1000 objects in an export file.

## Export Files



An export .zip file contains multiple subfolders and files.

When you open the export file, the following high-level structure is initially displayed:

<input type="checkbox"/> Name	Date modified	Type	Size
 Explore	3/11/2020 2:59 PM	File folder	
 SYS	3/11/2020 2:59 PM	File folder	
 ContentsofExportPackage_VProject-1583954369675.csv	3/11/2020 7:20 PM	Microsoft Excel C...	1 KB
 exportMetadata.v2.json	3/11/2020 7:20 PM	JSON File	7 KB
 exportPackage.chksum	3/11/2020 7:20 PM	CHKSUM File	2 KB





## Explore folder

The Explore folder contains a metadata .dat file for each exported task. If you export a project, a json file with metadata about the project also appears in the Explore folder, at the same level as the project folder. The following image shows an example for an exported project:

<input type="checkbox"/> Name	Date modified	Type	Size
 VProject	3/11/2020 2:59 PM	File folder	
 VProject.Project.json	3/11/2020 7:20 PM	JSON File	5 KB

**Note:** The Explore folder structure reflects how the objects appear on the source organization's **Explore** page.

Drill down on the project folder to view the .dat files that contain metadata for each task in the project:

<input type="checkbox"/> Name	Date modified	Type	Size
 RPS_SI2.SIDataflow.dat	3/11/2020 7:20 PM	DAT File	2 KB
 vp_ing2_incrementalb.DBMI_TASK.dat	3/11/2020 7:20 PM	DAT File	1 KB
 vp_ing3_combined.DBMI_TASK.dat	3/11/2020 7:20 PM	DAT File	1 KB
 vp_orazwhateverworks.DBMI_TASK.dat	3/11/2020 7:20 PM	DAT File	1 KB

An extension is appended to the .dat file names to indicate the asset type. The following table lists the Mass Ingestion asset types and their associated extensions:

Asset Type	Extension
database ingestion task	DBMI_TASK
streaming ingestion task	SIDataFlow

## SYS folder

The SYS folder contains .zip files for the associated connections and Agent groups. The following image shows an example of the contents of the SYS folder:

<input type="checkbox"/> Name	Date modified	Type	Size
DEMO-Snowflake.Connection.zip	3/11/2020 7:20 PM	WinRAR ZIP archive	2 KB
GBW1PF0V4FSE.AgentGroup.zip	3/11/2020 7:20 PM	WinRAR ZIP archive	1 KB
mhredhat5.informatica.com.AgentGroup.zip	3/11/2020 7:20 PM	WinRAR ZIP archive	1 KB
MHV19PWXQA01.AgentGroup.zip	3/11/2020 7:20 PM	WinRAR ZIP archive	1 KB
mp_Kafka_jotunheim.Connection.zip	3/11/2020 7:20 PM	WinRAR ZIP archive	2 KB
<input type="checkbox"/> rao_ora18rh_rh5.Connection.zip	3/11/2020 7:20 PM	WinRAR ZIP archive	3 KB

Each .zip file contains a json file and a metadata file for the asset.

## CSV file

The .csv file lists the objects in the export file. The following image is an example of a .csv file, as displayed as a spreadsheet in Excel, for an exported project, including database ingestion and streaming ingestion tasks and all dependent objects:

objectPath	objectName	objectType	id
/SYS	MHV19PWXQA01	AgentGroup	8mGSa2rWTCdjOO7AXzMJyk
/SYS	mhredhat5.informatica.com	AgentGroup	2mrI3AOINUPh9VIsWXvyQf
/Explore/VProject	vp_ing3_combined	DBMI_TASK	ay6A8Gek5gOcpHKLwmZ4KS
/SYS	rao_ora18rh_rh5	Connection	74Ah0AfpNVOI6D6e0zfGOR
/SYS	DEMO-Snowflake	Connection	7CcpypD9pIFb5ZgTYNZzq8
/Explore/VProject	RPS_SI2	SIDataflow	9mZmsWt8PSOg2mX0eIV6qY
/SYS	mp_Kafka_jotunheim	Connection	9dX4XEp8Swycz6XZRP5CI4
/Explore/VProject	vp_ing2_incrementalb	DBMI_TASK	64mhQOKB78VeJC4mAvARCT
/SYS	GBW1PF0V4FSE	AgentGroup	04OmYZTCxQEaZYfthOMi8v
/Explore	VProject	Project	bihoive8aAafGUdRSNJgw9

## Exporting assets

You can select a single asset, multiple assets, or a project to export.

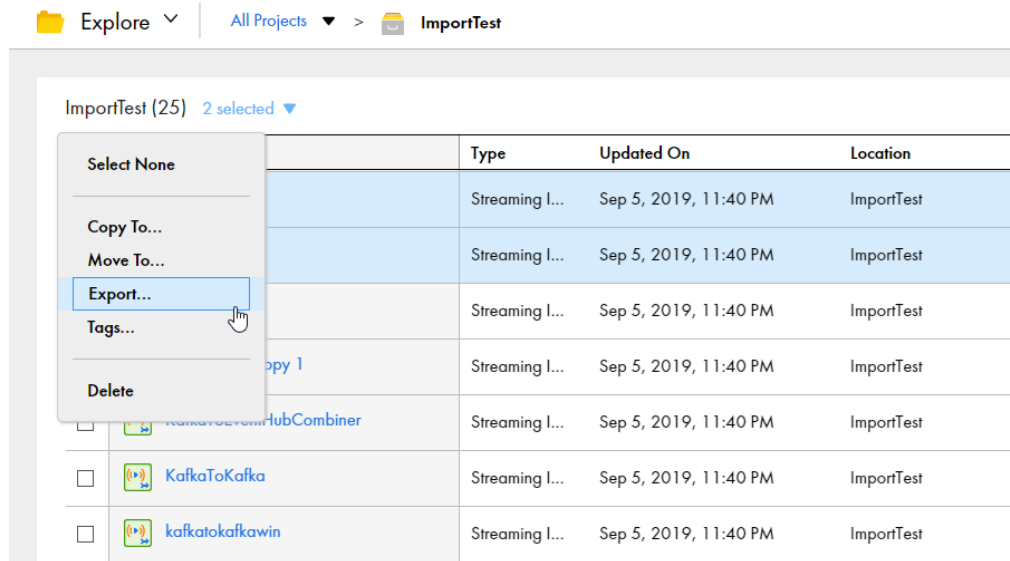
To include multiple assets, you can either select each asset within a folder or select a project or folder to export all of its assets. If you export a project, during import you can import the entire project or import only the assets that you select.

1. Log in to the source organization.
2. On the **Explore** page, navigate to the assets that you want to export.
3. Select the assets that you want to export.

To export a single asset or project, select the asset or project, and then click **Actions** and select **Export**.

To export multiple assets, select the check box to the left of each asset that you want to export. Or, select the check box for each project or folder that contains the assets that you want to export. From the selection menu, select **Export**.

The following image shows the selection menu with multiple assets selected:



4. On the **Export Assets** page, change the job name or leave the default.
5. Select whether to include dependent objects for the assets.
6. Click **Export**.
7. To see the progress of the job, select **My Import/Export Logs** from the navigation bar and then select **Export** from the menu at the top of the page. Click the name of the log to open the log details page.

## Asset import

You can import all of the assets in an export file or select the assets that you want to import.

When you import assets, you specify the following information:

- The assets in the export file that you want to import and the projects in which to import them.
- Whether to overwrite assets in the target project with assets in the export file when there is a name conflict.
- The connections and runtime environments to use for the imported assets.

To import an asset, you need the following privileges and permissions:

- Your user role must have privileges to import assets.
- If you import an asset into the target project as a new asset, you must have create, update, and read permissions on the asset.
- If you overwrite an asset in the target project, you must have update and read permissions on the asset.

Additionally, to overwrite a source-controlled asset in the target project, you must have the asset checked out.

The **Import Assets** page lists the assets that are in the export file. You can select which assets you want to import, and then specify which project to import the assets to. You can accept the default project, which is the same project name as the source project, or you can select a different project. If the project does not exist in the target organization, Informatica Intelligent Cloud Services creates it.

## Asset name conflicts

You can specify how Informatica Intelligent Cloud Services handles asset name conflicts when the export file contains assets with the same name as assets in the target project. You can choose whether to overwrite the assets in the target project or use the existing assets in the target project.

To see how the import handles any asset name conflicts before you start the import job, you can test the import on the **Import Assets** page before you import the assets. The import action displays in the **Status** column for each asset. You can filter the list of assets by asset name, asset type, or status.

The following image shows a list of assets and the import action to be performed when overwriting existing assets is enabled:

Import Assets

Start an import job, review the assets from the import file, and resolve any error related to location, connections, or runtime environments.

▼ Select Import File

Filename: \* DataIngestion-Import-Example.zip Choose File...

▼ Specify Import Job Details

Define the import behavior when an asset already exists, and provide a name that will be used to identify the import job on the My Jobs page.

Job Name: \* etstreject1 - Copy 2-1567692535909 ☒ Overwrite existing assets, excluding connections and runtime environments

▼ Select Assets

Assets (1) All selected

<input checked="" type="checkbox"/>	Name	Dependencies	Type	Location	Description	Status
<input checked="" type="checkbox"/>	etstreject...	0	Streaming Inge...	Jyothy		<input checked="" type="checkbox"/> Overwrite existing o...

If the target organization has connections or runtime environments with the same name as those in the export file, Data Ingestion uses the connections or runtime environments that exist in the target organization. Data Ingestion does not overwrite the connections or runtime environments in the target organization.

## Runtime environment and connection selection

If the export file contains dependent objects, the target connection and runtime environment fields show the connection and runtime environments from the export file as the default. You can accept the default or select a different connection or runtime environment.

If the export file doesn't include dependent objects and the connections or runtime environments that are used by the assets in the export file don't exist in the target organization, you must select a target connection or runtime environment.

When you select a connection or runtime environment that exists in the target organization, the connector type and version must be the same as the connector type and version that the asset used in the source organization.

## Importing assets

Import assets from an Informatica Intelligent Cloud Services export file.

1. Log in to the target organization.
2. On the **Explore** page, navigate to **All Projects** and click **Import**.
3. On the **Import Assets** page, navigate to the export file and click **Open**, or drag the zip file from the Downloads folder in Windows.

The **Import Assets** page lists the assets in the file.



4. Optionally, change the import job name.
5. Choose whether to overwrite existing assets with the assets in the import.
  - If you choose to overwrite existing assets, when an asset has the same name as an asset in the target project, the asset replaces the existing asset in the target project.
  - If you do not choose this option, if an asset with the same name exists in the target project, the asset is not imported.
6. Select the assets that you want to import.

If the export file contains a project and you want to import the entire project, select all of the assets. Informatica Intelligent Cloud Services creates the project in the source organization.
7. Select the target project or accept the default.
8. Click **Test** to see the potential results of the import.

In the Select Assets area, the status for each asset shows the action that the service performs when you import the files.
9. If necessary, revise your selections to resolve any issues in the test results.
10. Click **Import**.

You can see the progress of the import on the **Import** tab of the **My Import/Export Logs** page. When the import process is complete, a message appears in **Notifications**. Click the link in the message to open the log details page and see the results of the import.

## Post-import tasks

To complete the migration process you need to perform certain tasks based on the types of assets that you imported.

Perform the following tasks after you import assets:

- Configure connection passwords and security tokens. Informatica Intelligent Cloud Services does not include connection passwords and security tokens in imports for security reasons.

## Source control

You can use a GitHub or Azure DevOps Git source control repository with Informatica Intelligent Cloud Services to manage and track changes made to Informatica Intelligent Cloud Services objects such as projects, folders, and assets.

You can use source control to enable version management for the Informatica Intelligent Cloud Services objects that appear on the **Explore** page, except for Data Integration bundles. You cannot apply source control to objects that do not appear on the **Explore** page such as runtime environments or connections. The source control repository structure mirrors the structure in the organization, with **Explore** as the top level directory.

To use source control, the following prerequisites must be met:

- The organization has the appropriate Informatica Intelligent Cloud Services licenses to use source control.
- The organization administrator has configured a connection between the source control repository and the Informatica Intelligent Cloud Services organization.

- Your user role has privileges to use the Informatica Intelligent Cloud Services source control feature.
- You have entered your source control repository user credentials in Informatica Intelligent Cloud Services.

For information about configuring a connection between a GitHub or Azure DevOps Git source control repository and Informatica Intelligent Cloud Services, see the Administrator help.

**Note:** Informatica recommends that you include no more than 1000 objects in a container such as a project or folder when you use a source control repository with Informatica Intelligent Cloud Services.

## Source control actions

You can perform the following actions on source-controlled objects such as projects, folders, and assets:

### **Pull an object.**

Pull an object to add it to the organization or update a project with the version in the source control repository.

### **Check out an object.**

Check out an object that you want to work on. When you check out an object, the object locks so that other users cannot make changes to it.

### **Check in an object.**

Check in an object to add it to the source control repository or update the source control repository with the latest version of the object. When you check in an object, the lock releases.

### **Delete an object.**

Delete an object from the organization and the source control repository. Before you can delete an object, you must check it out.

### **Restore an object version.**

Restore an object to a previous version.

### **Undo a checkout.**

Undo a checkout if you don't want to save the changes you made to the object. When you undo a checkout, the object reverts to the last source control version.

### **Unlink an object.**

Unlink an object if you no longer want the object in the organization to stay in sync with the object in the source control repository.

**Note:** Some organizations do not have permission to update the source control repository. If your organization cannot update the repository, you can perform a pull action to get a specified version of the Informatica Intelligent Cloud Services objects. However, you cannot perform other source control actions such as checking out and checking in objects.

To check out, pull, unlink, or undo a checkout for an object, you must have update permissions on the object.

## Source control and the Git repository

If you usually work directly in a Git source control repository, you might notice a few differences between using source control in Informatica Intelligent Cloud Services and working directly in the repository.

Note the following differences:

- A pull action in Informatica Intelligent Cloud Services is the same as a Git pull command. However, a pull in Informatica Intelligent Cloud Services cannot merge changes.

- A checkout action in Informatica Intelligent Cloud Services locks an object so that no one else can check it out or change it.
- A check-in action in Informatica Intelligent Cloud Services is the same as a Git commit command and push command combined. Use a check-in to add Informatica Intelligent Cloud Services objects to the source control repository and to commit changes to the repository.

## Configuring repository access

To work with source controlled objects, specify your GitHub or Azure DevOps Git repository credentials in Informatica Intelligent Cloud Services.

Your credentials can include a user name and a personal access token.

If your administrator has configured the organization's repository for OAuth access, you can enable OAuth access instead of providing a personal access token.

Personal access tokens must be configured to enable full control of private repositories. For information about generating a personal access token, see the GitHub or Azure DevOps Git help.

In Informatica Intelligent Cloud Services, perform the following steps to configure access to the repository:

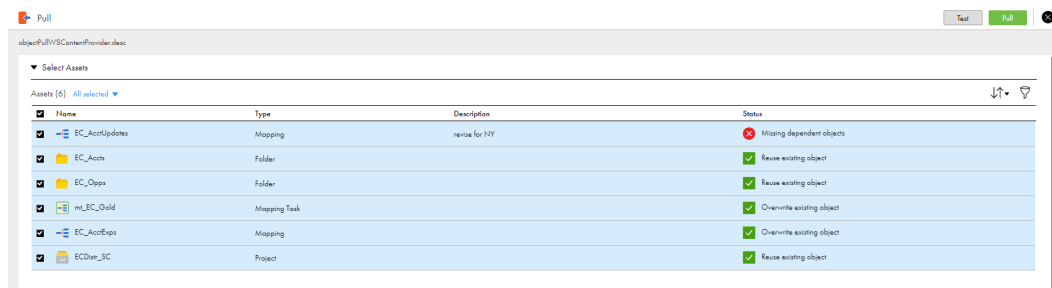
1. Click the **User** icon in the top right corner of the Informatica Intelligent Cloud Services window and then select **Settings**.
2. Perform one of the following tasks:
  - Enter your repository credentials. For GitHub, enter your user name and personal access token. For Azure DevOps Git, enter your personal access token.
  - Enable OAuth access to the repository. If you have not already authorized access, a Git access app appears. Select to authorize access for Informatica Intelligent Cloud Services.
3. Click **Save**.

## Source control pulls

Pull an object to add it to the organization or replace the version of the object in the organization with the version in the source control repository. You can pull a project, folder, or an individual object that isn't checked out by another user.

Before a pull takes place, you can review the potential results of the pull. You can cancel the action or select objects to exclude from the pull.

The following image shows a preview page with potential results in the **Status** column:



Name	Type	Description	Status
EC_AccelUpdones	Mapping	revise for NY	Missing dependent objects
EC_Accs	Folder		Reuse existing object
EC_Opex	Folder		Reuse existing object
mi_EC_Ould	Mapping Task		Overwrite existing object
EC_AccExp	Mapping		Overwrite existing object
ECDev_SC	Project		Reuse existing object

You can also test the pull to see if any errors might occur. If any errors appear, you can exclude the objects from the pull or close the page to cancel the pull.

A pull doesn't include dependent objects. For example, in Data Integration, if you pull a mapping that uses a saved query, the pull doesn't include the saved query.

A pull doesn't change the state of pulled objects in the organization. For example, if an object was checked out before the pull, it remains checked out after the pull.

Consider the following rules and guidelines:

- If you want to pull an object that uses a connection or runtime environment, be sure that the organization includes a connection and runtime environment before you perform the pull. When you select an object that uses a connection or runtime environment, you can select a runtime environment and connection to use on the test page.
- Informatica Intelligent Cloud Services doesn't consider capitalization in object names. As a result, you can't pull a project that contains multiple assets that have the same name with different capitalization. For example, if you try to pull a project that contains an asset named "sales" and an asset named "SALES", you receive an error because a project or folder can't contain multiple assets with the same name.
- You can pull objects regardless of whether your organization can or cannot update the source control repository.

## Project and folder pulls

You can pull one or more projects or folders.

A pull includes all of the objects within the project or folder. If an object in the project or folder is not source-controlled, the pull does not affect the object.

You specify the version of the objects that you want to pull. The versions of an object that you can select are based on the object's current location in Informatica Intelligent Cloud Services. For example, you check out version 5 of the m\_customers object and move it to a project called NewCustomers. When you pull the object, the available versions do not include the versions of the object from the previous location.

A pull updates the project or folder in the organization to be identical to the selected version of the project or folder in the repository. For example, if you pull an older version of a project and the project in the organization contains objects that were added in later versions, the pull deletes the objects.

## Pulling an object

Use the pull action to update the organization with objects in the source control repository.

To pull a single asset, you can select the asset and then click **Actions** and select **Pull**. To pull a project, folder, or multiple assets, perform the following steps:

1. On the **Explore** page, click **Pull from Git**.
2. Select the project that contains the objects you want to pull and click **Select**.  
The preview page appears with a list of available versions of the object. If the pull is for multiple objects, the preview page lists all of the versions in the repository.
3. Select the version that you want to pull and click **Pull**.
4. On the preview page, review the actions in the **Status** column for each object. These actions occur when the pull action is performed.
5. To exclude any objects from the pull, clear the check box by the object name.
6. Click **Test** to see if errors might occur as a result of the pull action.
7. If the test is successful, click **Pull**.  
The pull action generates a log showing details of this action. You can view the log on the **Source Control Logs** page in Monitor.

## Checking out and checking in objects

Check out an object so you can make changes to it. Check in the object when you want to update the source control repository with your changes. You can also check in objects to add them to the source control repository.

When you check out an object, the object locks so that other users can't make changes to it. When you check in an object, a new version of the object is created in the source control repository.

If you check in an object that's in a folder or project that isn't source controlled, the folder or project becomes source controlled. An asset can't reside in a source control repository unless it's in a container such as a project or folder.

**Note:** The size of a check-in cannot exceed 50 MB.

### Checking out an object

When you check out an object, the object is locked so that other users can't update it while you are making your changes.

You can check out individual objects, multiple objects, or a project or folder to check out all of the objects within the project or folder.

Before you check out an object, you might want to perform a pull to be sure you update the latest version of the object. For more information about the pull action, see ["Source control pulls" on page 539](#)

1. On the **Explore** page, navigate to the object you want to check out.
2. In the row that contains the object, click **Actions** and select **Check Out**.
3. If the checkout includes multiple objects, on the preview page, review the results in the **Status** column. If you want to exclude an object, clear the check box next to the object name.
4. Click **Check Out**.

### Checking in an object

Check in an object to add the object to the source control repository or to update the source control repository with the latest version of the object in the organization.

If no changes were made to the object, the check-in isn't reflected in the source control history and a new version of the object isn't created in the source control repository.

**Note:** The size of a check-in cannot exceed 50 MB.

1. On the **Explore** page, navigate to the object you want to check in.
2. In the row that contains the object, click **Actions** and select **Check In**.
3. If the check-in includes multiple objects, on the preview page, review the results in the **Status** column. If you want to exclude an object, clear the check box next to the object name. If you want to cancel the action, close the page.
4. Add a summary and optionally, a description.  
A summary is required and has a max length of 255 characters.  
A description is optional and has a max length of 500 characters.
5. Click **OK**.

## Deleting an object

To delete a source controlled object, you delete it from the organization and from the source control repository.

You must check out an object before you can delete it.

You can't delete an object that's checked out by another user or delete a project or folder recursively.

1. On the **Explore** page, navigate to the object that you want to delete.
2. On the row that contains the object, click **Actions** and then click **Delete**.
3. To confirm that you want to delete the asset from the organization, click **Delete**.
4. Add a summary that describes the reason for the delete action and optionally, a description.  
A summary is required and has a max length of 255 characters.  
A description is optional and has a max length of 500 characters.
5. To delete the object from the repository and complete the delete action, click **OK**.

## Reverting to an older version

You might want to revert to a previous version of an object if you want to discard changes that were made to the object.

To revert to a previous version, perform a pull action and select the version that you want to restore in the organization.

If you revert the version of a project or folder and the project or folder in Informatica Intelligent Cloud Services contains objects that are not in the repository's project or folder, the pull action deletes the additional objects if they are source controlled. If the objects are not source controlled, the action doesn't delete the additional objects.

For more information about the pull action, see [“Source control pulls” on page 539](#).

## Undoing a checkout

When you undo a checkout, the object reverts to the last version in the source control repository. The object's version history will not include a record of the checkout and undo checkout actions. The undo checkout releases the lock so that the object is available for checkout.

You can undo the checkout of individual objects, multiple objects, or a project or folder in a single checkout action.

You can undo the checkout of any object that you have checked out. You cannot undo the checkout of an object that has been checked out by another user unless you have the Admin role or your user role has the Force Undo Checkout feature privilege for the Administrator service.

If you undo the checkout of a project or folder, you can select which objects within the project or folder to include or exclude. By default, all of the objects are included.

**Note:** If an object was moved or renamed after it was checked out, undoing the checkout restores the object's name and location to its name and location before it was checked out.

1. On the **Explore** page, navigate to the object.
2. In the row that contains the object, click **Actions** and select **Undo Check Out**.
3. If the undo checkout includes a project or folder, on the preview page, select the objects within the project or folder to exclude from the undo checkout action.

## Unlinking an object

You can unlink an object so that it's no longer source controlled.

Unlinking an object doesn't delete the object from the source control repository or the organization, but you can no longer update the repository for any changes you make to the object in the organization. If you decide to link the object in the future, you can check in the object to reestablish the link. If the name of the object or the path to the object has not changed, the checked in object becomes a new version of the object in the source control repository.

The object must be checked in before you can unlink it. You can unlink an object that's checked out by another user if you have the Admin role or your user role has the Force Undo Checkout feature privilege for the Administrator service.

You can't unlink a project or folder that contains source controlled objects. To unlink a project or folder, unlink each object within the project or folder first.

1. On the **Explore** page, navigate to the object that you want to unlink.
2. On the row that contains the object, click **Actions** and select **Unlink**.

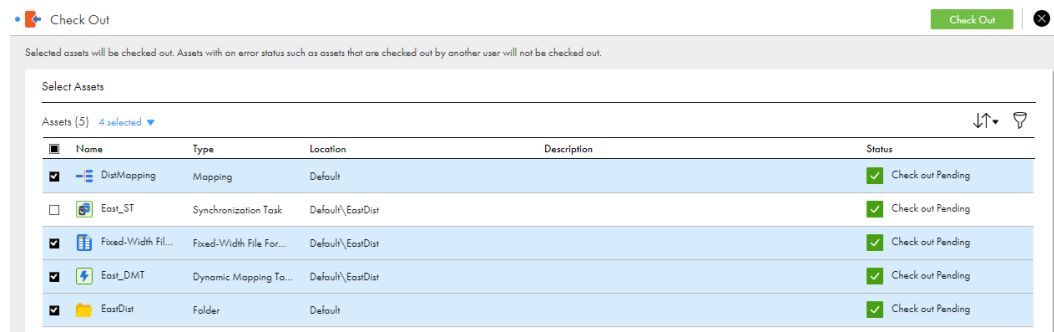
## Working with multiple objects

You can select multiple objects to check in, check out, undo checkout, pull, or unlink.

The total size of the selected objects cannot exceed 50 MB.

You can select a project or folder or select multiple objects within a project or folder. When you include a project or folder or multiple objects in a source control action, a preview page appears that shows you the expected results of the action if you proceed. If an object listed on the preview page is not source controlled, it will be ignored. If an object is checked out by another user or you do not have permission to update the object, the status on the preview page shows that the action will fail. You can opt to remove any of the objects before you continue.

For example, in the Default project, you select the EastDist folder and the DistMapping asset to check out. The preview page includes the DistMapping asset, the EastDist folder, and all of the objects in the EastDist folder. You don't want to check out the East\_ST asset so you clear its check box before you proceed with the checkout action. The following image shows the preview page:



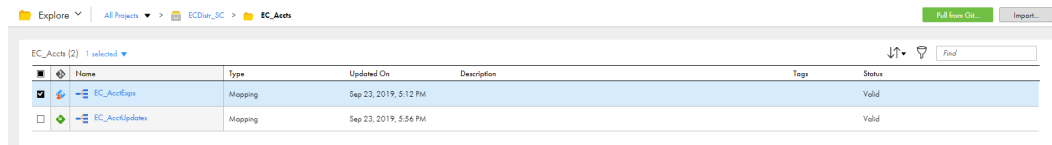
If you select multiple objects for a pull, the preview page lists all of the versions in the repository.

**Note:** You can't delete multiple objects in one transaction.

## Viewing source control columns on the Explore page

If your organization has any objects that are source controlled, the **Explore** page displays an additional column that indicates whether an object is checked in, checked out, or is not source controlled.

In the following image, the blue and red icon indicates that the EC\_AcctExps object is checked out:

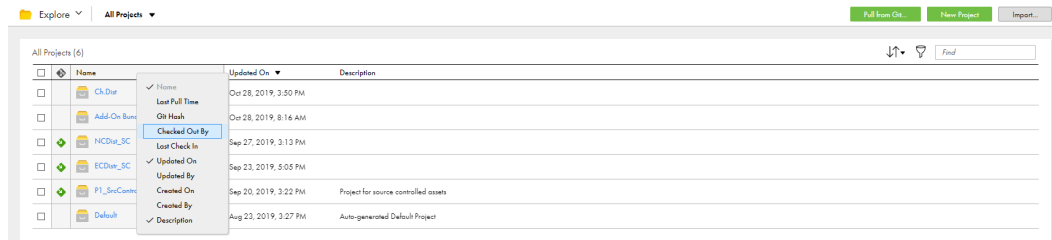


Additional source control columns are available but are not displayed by default. You might want to customize the **Explore** page to show all of the source control-related columns. This way you can easily identify source controlled objects, and you can use the columns to filter and sort source controlled objects on the page.

You can add the following source control-related columns to the **Explore** page:

- Last Pull Time
- Checked Out By
- Last Check In
- Git Hash

To display these columns, right-click a column heading and select the columns that you want to add, as shown in the following image:



## Source control best practices

To use source control effectively, use the following guidelines as best practices.

Adhere to the following guidelines as you develop and work with assets:

### Guidelines for managing dependencies

Use the following guidelines to manage assets with dependencies:

- Create connections and runtime environments before you pull assets from the repository.

When required connections and runtime environments exist in the target organization, you can run tasks immediately after you pull them from the repository.

- Ensure that reusable assets such as mappings and components are present in the repository before you use them.

Informatica Intelligent Cloud Services does not allow you to save an asset such as a mapping task when the dependent mapping does not exist in the organization.

- Avoid moving or renaming source-controlled assets that are used by other assets.

If you move or rename a source-controlled asset, references to the asset can break.



## Guidelines for checking in and checking out assets

Use the following guidelines when you check in and check out assets:

- Identify all dependencies before you check out reusable assets such as mappings, mapplets, and user-defined functions.

Source control operations such as check out, check in, and pull do not automatically include dependent assets.

- When you need to update a reusable asset such as a mapping or component, check out the asset and all dependent assets.

For example, when you need to update a mapping, check out the mapping and all mapping tasks that use it to ensure that changes to the mapping are propagated to the mapping tasks.

- Check in a reusable asset and all dependent assets in one operation.

This ensures that changes to the asset and dependent assets are committed to the source control repository at the same time. It also ensures that users get the latest versions of the dependent assets when they pull the assets.

- Enter comments when you check in assets.

When you check in assets, you might enter a release tag name in the **Summary** field and enter more descriptive comments in the **Description** field. When you do this, the **Git Summary** field in Informatica Intelligent Cloud Services shows the release tag that is associated with the asset.

- When you check in multiple assets at one time, limit the number of assets to 1000 or fewer.

Checking in more than 1000 assets at one time can degrade performance between Informatica Intelligent Cloud Services and the GitHub repository service.

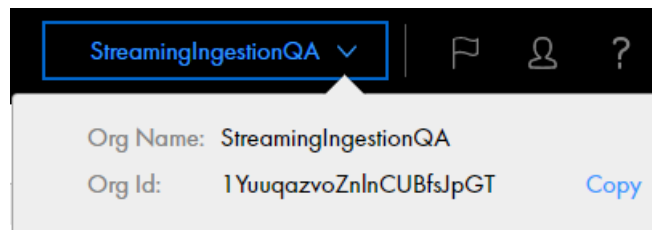
## CHAPTER 10

# Troubleshooting

Use the following sections to troubleshoot errors in Mass Ingestion.

**Note:** To get support for Mass Ingestion, you might need to give your organization ID to Informatica Global Customer Support. You can find your organization ID through the **Organization** menu in the upper right corner.

The following image shows the **Organization** menu:



To copy the organization ID, click the **Copy** option that appears when you hover the cursor to the right of the **Org ID** field.

You can also find your organization ID on the **Organization** page in Administrator.

## Troubleshooting an application ingestion task

**If you change the unsupported data type of a source field to a supported data type, the change might not be replicated to the target.**

This problem occurs when the **Modify field** schema drift option is set to **Replicate** and the **Add field** option is set to **Ignore**.

When you deploy an application ingestion job, Mass Ingestion Applications does not create target columns for the source fields of unsupported data types. If you change the unsupported data type to a supported data type for the source field, Mass Ingestion Applications processes the modification made to the source field but does not replicate the change to the target. When Mass Ingestion Applications tries to add a field with the supported data type to the target, the operation is ignored because the schema drift option **Add column** is set to **Ignore**.

To handle this situation, perform the following steps:

1. On the **Schedule and Runtime Options** page in the application ingestion task wizard, in the **Schema Drift Options** section, set the **Add field** option to **Replicate**.

2. Change the source field data type to a supported data type again so that the application ingestion job can detect the schema change.

**Note:** The application ingestion job does not propagate the field values that were added prior to changing the data type of the source field.

3. If you want to propagate all the values from the source field to the target, resynchronize the target table with the source.

**If you change a primary key constraint on the source, Mass Ingestion Applications stops processing the source object on which the DDL change occurred.**

This problem occurs if you add or drop a primary key constraint, or if you add or drop a field from an existing primary key.

To resume processing the source object for combined initial and incremental jobs, resynchronize the target table with the source.

To resume processing the source object for incremental jobs, perform the following steps:

1. On the **Source** page of the application ingestion task wizard, add an object selection rule to exclude the source object.
2. Redeploy the task.  
Mass Ingestion Applications deploys the modified task and deletes the information about the primary keys of the excluded object.
3. Edit the task again to delete the object selection rule that excluded the source object.
4. Redeploy the task.

**An application ingestion initial load job that processes many source objects and has a Google BigQuery target might fail with the following error:**

`The job has timed out on the server. Try increasing the timeout value.`

This problem occurs when the job is configured to process many source objects and the connection configured for the Google BigQuery target times out before completing the processing of the source objects. To resolve this issue, you must increase the timeout interval for the Google BigQuery V2 connection configured for the target.

To increase the timeout interval for the connection, perform the following steps:

1. In Administrator, open the Google BigQuery V2 connection associated with the application ingestion job in edit mode.
2. In the **Provide Optional Properties** field, set the timeout property to the required timeout interval in seconds. Use the following format:  

```
"timeout": "<timeout_interval_in_seconds>"
```
3. Save the connection.
4. Redeploy the application ingestion job.

## Troubleshooting a database ingestion task

**If you change an unsupported data type of a source column to a supported data type, the change might not be replicated to the target.**

This problem occurs when the **Modify column** schema drift option is set to **Replicate** and the **Add column** option is set to **Ignore**.

Mass Ingestion Databases does not create target columns for source columns that have unsupported data types when you deploy a task. If you change the unsupported data type to a supported data type for the source column later, Mass Ingestion Databases processes the modify column operation on the source but does not replicate the change to the target. When Mass Ingestion Databases tries to add a column with the supported data type to the target, the operation is ignored because the schema drift option **Add column** is set to **Ignore**.

To handle this situation, perform the following steps:

1. On the **Schedule and Runtime Options** page in the database ingestion task wizard, under **Schema Drift Options**, set the **Add column** option to **Replicate**.
2. Change the source column data type to a supported data type again so that the database ingestion job can detect this schema change.  
The database ingestion job processes the DDL operation and creates the new target column.  
**Note:** The database ingestion job does not propagate the column values that were added prior to changing the source column data type.
3. If you want to propagate all of the values from the source column to the target, resynchronize the target table with the source.

**If you change a primary key constraint on the source, Mass Ingestion Databases stops processing the source table on which the DDL change occurred.**

This problem occurs if you add or drop a primary key constraint, or if you add or drop a column from an existing primary key.

To resume processing the source table for combined initial and incremental jobs, resynchronize the target table with the source.

To resume processing the source table for incremental jobs, perform the following steps:

1. On the **Source** tab in the database ingestion task definition, add a table selection rule to exclude the source table.
2. Redeploy the task.  
Mass Ingestion Databases deploys the edited task and deletes the information about the primary keys of the excluded table.
3. Edit the task again to delete the table selection rule that excluded the source table.
4. Redeploy the task.

**If a DDL column-level change causes a source table subtask to stop or be in error and then you resume the database ingestion job, the expected change in the table state is delayed.**

If a DDL column-level change on a source table causes a table subtask to stop or be in error and then you resume the database ingestion job, the state of the table subtask might remain unchanged until a DML operation occurs on the table. For example, if you set a schema drift option to **Stop Table** for an incremental or initial and incremental database ingestion task and then deploy and run the job, when a DDL change occurs on a source table, the job monitoring details shows the table subtask to be in the Error state. If you stop the job and then resume it with a schema drift override to replicate the DDL change, the table subtask temporarily remains in the Error state until the first DML operation occurs on the source table.

**Mass Ingestion Databases failed to deploy a task that has a Snowflake target with the following error:**

```
Information schema query returned too much data. Please repeat query with more selective predicates.
```

This error occurs because of a known Snowflake issue related to schema queries. For more information, see the Snowflake documentation.

In Mass Ingestion Databases, the error can cause the deployment of a database ingestion task that has a Snowflake target to fail when a large number of source tables are selected.

To handle the deployment failure, drop the target tables. Then update the database ingestion task to select fewer source tables for generating the target tables. Then try to deploy the task again.

**A database ingestion job that runs on Linux ends abnormally with the following out-of-memory error:**

```
java.lang.OutOfMemoryError: unable to create new native thread
```

The maximum number of user processes that is set for the operating system might have been exceeded. If the Linux ulimit value for maximum user processes is not already set to **unlimited**, set it to **unlimited** or a higher value. Then resume the job.

**If you copy an asset to another location that already includes an asset of the same name, the operation might fail with one of the following errors:**

```
Operation succeeded on 1 artifacts, failed on 1 artifacts.  
Operation did not succeed on any of the artifacts.
```

If you try to copy an asset to another location that already has an asset of the same name, Mass Ingestion Databases displays a warning message that asks if you want to keep both assets, one with a suffix such as "- Copy 1". Note that when you choose to keep both assets, Mass Ingestion Databases validates the name length to ensure that it will not exceed the maximum length of 50 characters after the suffix is added. If the name length will exceed 50 characters, the copy operation will fail. In this case, you must copy the asset to another location, rename the copy, and then move the renamed asset back to the original location.

**Schema drift options on the Schedule and Runtime Options page do not match the options in the job log**

In the Mass Ingestion Databases Spring 2020 April release, the order of the **Schema Drift** options displayed on the **Schedule and Runtime Options** page changed but their values remained in the original order. If you use a database ingestion task that was created after the Spring 2020 April release and before the Fall 2020 December release, the schema drift options that the associated job uses at runtime, as reported under "schemaChangeRules" in the job log, might not match the **Schema Drift** options displayed on the **Runtime Options** page. In this case, reset each of the **Schema Drift** options on the **Runtime Options** page and save the task again.

**A Kafka consumer ends with one of the following errors:**

```
org.apache.avro.AvroTypeException: Invalid default for field meta_data: null not a  
{ "type": "array" ...  
org.apache.avro.AvroTypeException: Invalid default for field header: null not a  
{ "type": "record" ...
```

This error might occur because the consumer has been upgraded to a new Avro version but still uses the Avro schema files from the older version.

To resolve the problem, use the new Avro schema files that Mass Ingestion Databases provides.

**A database ingestion job that propagates incremental change data to a Kafka target that uses Confluent Schema Registry fails with the following error:**

```
io.confluent.kafka.schemaregistry.client.rest.exceptions.RestClientException:  
Register operation timed out; error code: 50002
```

This problem might occur when the job is processing many source tables, which requires Confluent Schema Registry to process many schemas. To resolve the problem, try increasing the value of the Confluent Schema Registry **kafkastore.timeout.ms** option. This option sets the timeout for an operation on the Kafka store. For more information, see the Confluent Schema Registry documentation.

**Subtasks of a database ingestion job that has a Google BigQuery target fail to complete initial load processing of source tables with the following error:**

```
The job has timed out on the server. Try increasing the timeout value.
```

This problem occurs when the job is configured to process many source tables and the Google BigQuery target connection times out before initial load processing of the source tables is complete. To resolve this problem, increase the timeout interval in the Google BigQuery V2 target connection properties.

1. In Administrator, open the Google BigQuery V2 connection that is associated with the database ingestion job in Edit mode.
2. In the **Provide Optional Properties** field, set the timeout property to the required timeout interval in seconds. Use the following format:

```
"timeout": "<timeout_interval_in_seconds>"
```

3. Save the connection.
4. Redeploy the database ingestion task.

**A database ingestion task with an Amazon Redshift target returns one of the following errors during deployment:**

```
Mass Ingestion Databases could not find target table 'table_name' which is mapped to source table 'table_name' when deploying the database ingestion task.
```

```
com.amazon.redshift.util.RedshiftException: ERROR: Relation "table_name" already exists
```

This problem occurs because Amazon Redshift reads table and column names as lowercase by default.

To resolve the problem and allow case-sensitive identifiers on Amazon Redshift targets, perform the following steps:

1. Disable the Amazon Redshift `downcase_delimited_identifier` parameter at the user level by using the following statement:

```
ALTER USER username SET downcase_delimited_identifier TO off;
```

2. Redeploy the database ingestion task.

To prevent these errors, you can set the `enable_case_sensitive_identifier` parameter to "true" when configuring the database parameter group.

**Deployment of a database ingestion task fails if the source table or column names include multibyte or special characters and the target is Databricks Delta.**

When a new Databricks Delta target table is created during deployment, an entry is added to the Hive metastore that Databricks Delta uses. The Hive metastore is typically a MySQL database. More specifically, column names are inserted into the `TABLE_PARAMS` field of the metastore. The charset collation of the `PARAM_VALUE` from `TABLE_PARAMS` is `latin1_bin`, and the charset is `latin1`. This charset does not support Japanese characters. To resolve the problem, create an external metastore with `UTF-8_bin` as the collation and `UTF-8` as the charset. For more information, see the Databricks Delta documentation at

<https://docs.microsoft.com/en-us/azure/databricks/kb/metastore/jpn-char-external-metastore> and <https://kb.databricks.com/metastore/jpn-char-external-metastore.html>.

## Troubleshooting a streaming ingestion task

**While deploying a streaming ingestion task with Amazon Kinesis source, data loss is encountered when you send data immediately after the task status changes to Up and Running.**

Workaround: After deploying the task to run as a job, wait for some time before sending the data.

**A streaming ingestion job with Kinesis Streams source which consumes streams from Amazon DynamoDB fails to read the data ingested from Kinesis Streams. The streaming ingestion job runs with Up and Running status, but does not return any error.**

Provide the following required permissions to the Amazon DynamoDB user:

```
dynamodb:CreateTable
- dynamodb:DescribeTable
- dynamodb:Scan
- dynamodb:PutItem
- dynamodb:GetItem
- dynamodb:UpdateItem
- dynamodb>DeleteItem
Resource:
- !Join ["", ["arn:aws:dynamodb:*:", !Ref 'AWS::AccountId', ":table/
*"]]
```

Provide the following permissions for Amazon CloudWatch:

```
"Action": "cloudwatch:DescribeAlarms"
"Action": "cloudwatch:PutMetricData"
```

**While creating a Kafka connection, if the SSL mode is set as disabled, any other additional connection property values declared are not considered.**

To override this issue, declare the additional connection properties in the **Additional Security Properties** field.

**When you try to ingest high volume of data to a Kafka target, the streaming ingestion job runs with the error:**

```
A message in the stream exceeds the maximum allowed message size of 1048576 byte.
```

You get this error message if the received message size is more than 1 MB, which is the maximum message size that a Kafka server can receive.

**When you try to ingest high volume of data to an Amazon Kinesis Firehose or Amazon Kinesis Streams target, the streaming ingestion job runs with the error:**

```
INFO - Mon Feb 04 07:01:44 UTC 2019PutKinesisStream[id=0421419e-b24f-4e3f-
ad19-9a1fbc7b0f3c] Failed to publish to kinesis records
StandardFlowFileRecord[uuid=36926fcd-dfed-46f9-ae41-
a6d32c3d5633,claim=StandardContentClaim
[resourceClaim=StandardResourceClaim[id=1549263211727-1, container=default,
section=1], offset=2758,
length=1067926],offset=0,name=testKinesis.2758-1070684.txt,size=1067926] because the
size was greater than 1024000 bytes
```

You get this error message if the received message size is more than 1 MB, which is maximum message buffer size of Amazon Kinesis. Any message of size greater than 1 MB never reaches the target and is lost in the transition.

**While deploying a streaming ingestion task with Amazon Kinesis source, data loss is encountered.**

Workaround: To override this issue, don't send data when the data flow is in Stopped, Deploying, Edit, or Redeploy state, after deploying the task to run as a job.

# INDEX

## A

- action
  - description [394](#)
- Advanced FTP V2 connections
  - properties [45](#)
- Advanced FTPS V2 connections
  - properties [47](#)
- Advanced SFTP V2 connections
  - properties [49](#)
- Amazon Kinesis
  - AWS credential profile [56](#)
- Amazon Kinesis connection
  - overview [52](#)
- Amazon MSK
  - target [460](#)
- Amazon Redshift
  - target considerations for database ingestion [211](#)
- Amazon Redshift targets
  - data types mapped to Amazon Redshift targets [309](#)
  - mappings with Db2 for i sources [276](#)
  - mappings with Db2 for Linux, UNIX, and Windows sources [294](#)
  - mappings with Db2 for z/OS sources [301](#)
  - mappings with Microsoft SQL Server sources [309](#)
  - mappings with MySQL sources [321](#)
  - mappings with Netezza sources [335](#)
  - mappings with Oracle sources [342](#)
  - mappings with PostgreSQL sources [355](#)
  - mappings with SAP HANA sources [367](#)
  - mappings with Teradata sources [375](#)
- Amazon Redshift V2
  - connection properties [50](#)
- Amazon Redshift V2 targets
  - file ingestion task [426](#)
  - properties [426](#)
- Amazon S3
  - flat file
    - target considerations for database ingestion [211](#)
  - Google Cloud Storage
    - target considerations for database ingestion [211](#)
  - Microsoft Azure Data Lake Storage
    - target considerations for database ingestion [211](#)
  - target considerations for database ingestion [211](#)
- Amazon S3 V2
  - connection properties [56](#)
- Amazon S3 V2 sources
  - file ingestion task [403](#)
  - properties [403](#)
- Amazon S3 V2 targets
  - file ingestion task [428](#)
  - properties [428](#)
- Apache Kafka
  - source properties [470](#)
  - target considerations for database ingestion [214](#)
- application ingestion tasks
  - configuring runtime options [172](#)

- application ingestion tasks (*continued*)
  - configuring the source [156](#)
  - configuring the target [160](#)
  - defining basic information for an application ingestion task [155](#)
- apply modes
  - Soft Deletes [230](#)
- asset name conflicts
  - importing [536](#)
- assets
  - creating tags [527](#)
  - exporting [532](#), [534](#)
  - importing [535](#), [536](#)
  - managing ingestion assets [523](#)
  - migrating between organizations [531](#)
  - source control [537](#)
  - tags [527](#)
- authentication
  - OAuth 2.0 authorization code [122](#), [124](#)
- Azure Data Lake Storage Gen2
  - connection properties [86](#)
- Azure DevOps user credentials [539](#)
- Azure Event Hub
  - target [461](#)
- Azure Event Hubs
  - target considerations for database ingestion [214](#)
- Azure Event Hubs Kafka
  - source [453](#)

## B

- browser [16](#)

## C

- Confluent Kafka
  - source [455](#)
  - target [460](#)
  - target considerations for database ingestion [214](#)
- connection
  - Amazon Kinesis Firehose
    - connection properties [53](#)
  - Amazon Kinesis Streams
    - connection properties [54](#)
- connections
  - Amazon Redshift V2 [50](#)
  - Amazon S3 V2 [56](#)
  - AMQP
    - connection properties [61](#)
  - Azure Data Lake Storage Gen2 [86](#)
  - Azure Event Hub
    - connection properties [88](#)
  - creating connections for ingestion tasks [44](#)
  - Databricks Delta [65](#)
  - Db2 for i Database Ingestion [62](#)



connections (*continued*)

- Db2 for LUW Database Ingestion connection [63](#)
- Db2 for zOS Database Ingestion [64](#)
- flat file [68](#)
- Google Analytics Mass Ingestion [69](#)
- Google BigQuery [69](#)
- Google Cloud Storage V2 [71](#)
- Google PubSub [75](#)
- importing [535](#)
- JDBC V2 [77](#)
  - JMS
    - connection properties [78](#)
  - Kafka
    - connection properties [79](#)
- Marketo V3 [85](#)
- Mass Ingestion connections overview [44](#)
- Microsoft Azure Blob Storage V3 [85](#)
- Microsoft Azure Synapse Analytics - Database Ingestion [88](#)
- Microsoft Azure Synapse SQL [90](#)
- Microsoft Dynamics 365 Mass Ingestion [91](#)
- Microsoft SQL Server [94](#)
- migrating [531](#)
- MongoDB Mass Ingestion [97](#)
  - MQTT
    - connection properties [98](#)
- MySQL [99](#)
- Netezza [100](#)
- NetSuite Mass Ingestion [100](#)
- OPC UA [101](#)
- Oracle Database Ingestion [103](#)
- Oracle Fusion Cloud Mass Ingestion [108](#)
- PostgreSQL [109](#)
- REST V2 [125](#)
- Salesforce Mass Ingestion [111](#)
- SAP HANA Database Ingestion [113](#)
- SAP ODP Extractor [115](#)
- ServiceNow Mass Ingestion [119](#)
- Snowflake Data Cloud [121](#)
- Teradata connection [127](#)
- testing connections for ingestion tasks [44](#)
- Workday Mass Ingestion [129](#)
- Zendesk Mass Ingestion [130](#)

connections Hadoop Files V2 [75](#)

connectors

- Data Ingestion connectors overview [38](#)

copying

- folders [524](#)
- projects [524](#)
- tasks [524](#)

creating

- tags [527](#)

## D

Data Ingestion connectors

- overview [38](#)

data type mappings

- customizing the default mappings [245](#)

Database Ingestion

- architecture [180](#)
- overview [178](#)
- troubleshooting [547](#)
- use cases [178](#)

database ingestion jobs

- overriding schema drift options [273](#)
- redeploying a job [274](#)
- resuming a job [272](#)

database ingestion jobs (*continued*)

- resynchronizing source and target objects [275](#)
- running a deployed job from monitoring [271](#)
- running jobs based on a schedule [271](#)
- stopping [271](#)
- stored procedure setup for CDC [186](#)
- undeploying a job [274](#)

database ingestion tasks

- Amazon Redshift target properties [246](#)
- Amazon S3 target properties [246](#)
- configuration task flow [230](#)
- configuring runtime options [267](#)
- configuring the source [232](#)
- configuring the target [243](#)
- connectors [40](#)
- Databricks Delta target properties [250](#)
- Db2 for i sources [183](#)
- Db2 for Linux, UNIX, and Windows sources [184](#)
- Db2 for z/OS sources [185](#)
- defining basic information for a task [231](#)
- deploying and ingestion task [270](#)
- example of table selection rules [242](#)
- flat file target properties [250](#)
- Google BigQuery target properties [253](#)
- Google Cloud Storage target properties [253](#)
- Kafka target properties [257](#)
- limitations [182](#)
- Microsoft Azure Data Lake Storage target properties [260](#)
- Microsoft Azure SQL Database source considerations [188](#)
- Microsoft Azure Synapse Analytics target properties [264](#)
- Microsoft SQL Server source considerations [188](#)
- MongoDB sources [190](#)
- MySQL sources [191](#)
- Netezza source considerations [192](#)
- Oracle source considerations [193](#)
- Oracle source privileges [197](#)
- Oracle target properties [264](#)
- PostgreSQL source considerations [205](#)
- rules for renaming target tables [244](#)
- SAP HANA sources [208](#)
- Snowflake target properties [265](#)
- source considerations [183](#)
- source types [179](#)
- target considerations [210](#)
- Teradata source considerations [210](#)

Databricks Delta

- connection properties [65](#)
- target [458](#)
- target considerations for database ingestion [212](#)
- target properties [477](#)

Databricks Delta sources

- file ingestion task [406](#)
- properties [406](#)

Databricks Delta targets

- file ingestion task [429](#)
- mappings with Db2 for i sources [277](#)
- mappings with DB2 for Linux, UNIX, and Windows sources [295](#)
- mappings with Db2 for zOS sources [302](#)
- mappings with Microsoft SQL Server sources [311](#)
- mappings with MySQL sources [323](#)
- mappings with Netezza sources [336](#)
- mappings with Oracle sources [344](#)
- mappings with PostgreSQL sources [357](#)
- mappings with SAP HANA sources [369](#)
- mappings with Teradata sources [377](#)
- properties [429](#)

Db2 for i Database Ingestion connections

- connection properties [62](#)

- Db2 for i sources
  - data types mapped to Amazon Redshift targets [276](#)
  - mappings with Databricks Delta targets [277](#)
  - mappings with Google BigQuery targets [278](#)
  - mappings with Microsoft Azure Synapse Analytics targets [290](#)
  - mappings with Oracle targets [291](#)
  - mappings with Snowflake targets [293](#)
- Db2 for Linux, UNIX, and Windows sources
  - data types mapped to Amazon Redshift targets [294](#)
- DB2 for Linux, UNIX, and Windows sources
  - mappings with Databricks Delta targets [295](#)
  - mappings with Google BigQuery targets [296](#)
  - mappings with Microsoft Azure Synapse Analytics targets [299](#)
  - mappings with Snowflake targets [300](#)
- Db2 for LUW Database Ingestion connection
  - connection properties [63](#)
- Db2 for z/OS sources
  - source preparation and usage considerations [185](#)
  - data types mapped to Amazon Redshift targets [301](#)
  - mappings with Microsoft Azure Synapse Analytics targets [307](#)
  - mappings with Snowflake targets [308](#)
  - stored procedure setup for CDC [186](#)
- Db2 for zOS Database Ingestion connections
  - connection properties [64](#)
- Db2 for zOS sources
  - mappings with Databricks Delta targets [302](#)
  - mappings with Google BigQuery targets [303](#)
- deleting
  - folders [526](#)
  - projects [526](#)
  - tags [529](#)
  - tasks [526](#)
- dependencies
  - viewing dependencies [529](#)
- dependent objects
  - in export files [531](#)
  - runtime environments and connections [531](#)
- directories
  - configuring Secure Agent login to access [20](#)

## E

- editing
  - tags [529](#)
- email addresses
  - for notification [37](#)
- Explore page
  - source control columns [544](#)
  - tags [527](#)
- exporting
  - assets [532](#), [534](#)
  - dependent objects [531](#)
  - export file structure and contents [533](#)
  - projects [534](#)
- exporting assets
  - overview [531](#)

## F

- file ingestion
  - schedule configuration [444](#), [445](#)
- file ingestion task
  - advanced FTP V2 source [397](#)
  - advanced FTP V2 target [425](#)
  - advanced FTPS V2 source [399](#)
  - advanced FTPS V2 target [425](#)

- file ingestion task (*continued*)
  - advanced SFTP V2 source [401](#)
  - advanced SFTP V2 target [426](#)
  - configuration [396](#)
  - defining [396](#)
  - description [392](#)
  - file listener [407](#)
  - local folder source [412](#)
  - local folder target [432](#)
  - running [445](#)
  - source configuration [397](#)
  - target configuration [424](#)
- file ingestion tasks
  - prerequisites [396](#)
  - sources [392](#)
  - targets [393](#)
- firewall
  - configuration [18](#), [22](#)
- flat file
  - connection properties [68](#)
- folders
  - copying [524](#)
  - deleting [526](#)
  - for creating in a project in Mass Ingestion [36](#)
  - importing [536](#)
  - moving [524](#)

## G

- Getting Started
  - in Mass Ingestion [16](#)
- GitHub user credentials [539](#)
- Google Analytics Mass Ingestion connections
  - connection properties [69](#)
- Google BigQuery
  - connection properties [69](#)
  - target considerations for database ingestion [213](#)
- Google BigQuery targets
  - mappings with Db2 for i sources [278](#)
  - mappings with DB2 for Linux, UNIX, and Windows sources [296](#)
  - mappings with Db2 for zOS sources [303](#)
  - mappings with Microsoft SQL Server sources [312](#)
  - mappings with MySQL sources [325](#)
  - mappings with Netezza sources [337](#)
  - mappings with Oracle sources [345](#)
  - mappings with PostgreSQL sources [358](#)
  - mappings with SAP HANA sources [370](#)
  - mappings with Teradata sources [378](#)
- Google BigQuery V2 targets
  - file ingestion task [430](#)
  - properties [430](#)
- Google Cloud Storage
  - target [459](#)
  - target properties [479](#)
- Google Cloud Storage V2
  - connection properties [71](#)
- Google Cloud Storage V2 sources
  - file ingestion task [408](#)
  - properties [408](#)
- Google Cloud Storage V2 targets
  - file ingestion task [431](#)
  - properties [431](#)
- Google PubSub
  - connection properties [75](#)
  - source [454](#)
  - source properties [469](#)
  - target [459](#)

Google PubSub (*continued*)  
target properties [480](#)

## H

Hadoop Files V2  
connection properties [75](#)  
Hadoop Files V2 sources  
file ingestion task [410](#)  
properties [410](#)  
Hadoop Files V2 targets  
file ingestion task [432](#)  
properties [432](#)

## I

importing  
asset name conflicts [536](#)  
assets [535](#)  
connections [535](#), [536](#)  
dependent objects [531](#)  
name conflicts [535](#)  
post-migration tasks [537](#)  
projects [535](#)  
runtime environments [536](#)  
importing assets  
overview [531](#)  
ingestion tasks  
editing tasks from Explore page [523](#)  
managing tasks [523](#)  
setting user permissions on a task [530](#)

## J

JDBC V2  
connection properties [77](#)  
JMS  
source [454](#)

## K

Kafka  
source [455](#)  
target [460](#)  
Kerberised Kafka  
prerequisites [81](#)

## L

Linux  
configuring proxy settings [23](#), [73](#)

## M

Marketo V3  
connection properties [85](#)  
Mass Ingestion  
My Services page [13](#)  
overview [12](#)  
Mass Ingestion connections  
overview [44](#)

Mass Ingestion Databases  
suborganization [16](#)  
Secure Agent group [16](#)  
System Configuration Details [16](#)  
Mass Ingestion Files  
suborganization [16](#)  
Secure Agent group [16](#)  
System Configuration Details [16](#)  
Mass Ingestion projects and assets [13](#)  
Mass Ingestion Streaming  
suborganization [16](#)  
overview [451](#)  
Secure Agent group [16](#)  
System Configuration Details [16](#)  
Transformations [461](#)  
Mass Streaming Ingestion  
use cases [451](#)  
mass streaming ingestion tasks  
AMQP [452](#)  
Azure Event Hubs Kafka [452](#)  
Google PubSub [452](#)  
Kinesis [452](#)  
OPC UA [452](#)  
REST V2 [452](#)  
source types  
Flat files [452](#)  
JMS [452](#)  
Kafka [452](#)  
MQTT [452](#)  
Microsoft Azure Blob Storage V3  
connection properties [85](#)  
Microsoft Azure Blob Storage V3 sources  
file ingestion task [414](#)  
properties [414](#)  
Microsoft Azure Blob Storage V3 targets  
file ingestion task [432](#)  
properties [432](#)  
Microsoft Azure Data Lake Storage Gen2  
target [461](#)  
Microsoft Azure Data Lake Storage Gen2 sources  
file ingestion task [416](#)  
properties [416](#)  
Microsoft Azure Data Lake Storage Gen2 targets  
file ingestion task [433](#)  
properties [433](#)  
Microsoft Azure Data Lake Store targets  
file ingestion task [433](#)  
properties [433](#)  
Microsoft Azure Data Lake Store V3 sources  
file ingestion task [418](#)  
properties [418](#)  
Microsoft Azure SQL Database sources  
data types mapped to Amazon Redshift targets [309](#)  
mappings with Microsoft Azure Synapse Analytics targets [317](#)  
mappings with Oracle targets [318](#)  
mappings with Snowflake targets [320](#)  
Microsoft Azure Synapse Analytics  
target considerations for database ingestion [215](#)  
Microsoft Azure Synapse Analytics Database Ingestion connections  
connection properties [88](#)  
Microsoft Azure Synapse Analytics targets  
mappings with Db2 for i sources [290](#)  
mappings with DB2 for Linux, UNIX, and Windows sources [299](#)  
mappings with Db2 for z/OS sources [307](#)  
mappings with Microsoft Azure SQL Database sources [317](#)  
mappings with Microsoft SQL Server sources [317](#)  
mappings with MySQL sources [332](#)  
mappings with Netezza sources [340](#)

- Microsoft Azure Synapse Analytics targets *(continued)*
  - mappings with Oracle sources [352](#)
  - mappings with PostgreSQL sources [363](#)
  - mappings with SAP HANA sources [372](#)
  - mappings with Teradata sources [387](#)
- Microsoft Azure Synapse SQL
  - connection properties [90](#)
- Microsoft Azure Synapse SQL Data targets
  - properties [434](#)
- Microsoft Azure Synapse SQL targets
  - file ingestion task [434](#)
- Microsoft Dynamics 365 Mass Ingestion connections
  - connection properties [91](#)
- Microsoft SQL Server
  - connection properties [94](#)
- Microsoft SQL Server sources
  - data types mapped to Amazon Redshift targets [309](#)
  - mappings with Databricks Delta targets [311](#)
  - mappings with Google BigQuery targets [312](#)
  - mappings with Microsoft Azure Synapse Analytics targets [317](#)
  - mappings with Oracle targets [318](#)
  - mappings with Snowflake targets [320](#)
- migrating
  - assets [534](#)
- migration
  - assets [536](#)
  - of assets between organizations [531](#)
- MongoDB Mass Ingestion
  - connection properties [97](#)
- monitoring ingestion jobs
  - database ingestion job details [511](#)
  - file ingestion job details [511](#)
  - Job Overview tab [511](#)
  - job properties in job lists [504](#)
  - monitoring all jobs from Monitor [501](#)
  - monitoring My Jobs in Mass Ingestion [500](#)
  - My Jobs and All Jobs pages [500](#)
  - Object Details tab [511](#)
  - streaming ingestion job details [511](#)
  - viewing job details [505](#)
- moving
  - tasks and folders [524](#)
- My Services page
  - Mass Ingestion box [13](#)
- MySQL
  - connection properties [99](#)
- MySQL sources
  - data types mapped to Amazon Redshift targets [321](#)
  - mappings with Databricks Delta targets [323](#)
  - mappings with Google BigQuery targets [325](#)
  - mappings with Microsoft Azure Synapse Analytics targets [332](#)
  - mappings with Snowflake targets [333](#)

## N

- Netezza
  - connection properties [100](#)
- Netezza sources
  - data types mapped to Amazon Redshift targets [335](#)
  - mappings with Databricks Delta targets [336](#)
  - mappings with Google BigQuery targets [337](#)
  - mappings with Microsoft Azure Synapse Analytics targets [340](#)
  - mappings with Snowflake targets [341](#)
- NetSuite Mass Ingestion connections
  - connection properties [100](#)

## O

- object migration [531](#)
- OPC UA
  - connection properties [101](#)
  - source [456](#)
  - source properties [472](#)
- Oracle
  - target considerations for database ingestion [216](#)
- Oracle Database Ingestion connections
  - connection properties [103](#)
- Oracle Fusion Cloud Mass Ingestion connections
  - connection properties [108](#)
- Oracle sources
  - data types mapped to Amazon Redshift targets [342](#)
  - mappings with Databricks Delta targets [344](#)
  - mappings with Google BigQuery targets [345](#)
  - mappings with Microsoft Azure Synapse Analytics targets [352](#)
  - mappings with Oracle targets [353](#)
  - mappings with Snowflake targets [354](#)
- Oracle targets
  - mappings with Db2 for i sources [291](#)
  - mappings with Microsoft Azure SQL Database sources [318](#)
  - mappings with Microsoft SQL Server sources [318](#)
  - mappings with Oracle sources [353](#)
- organization ID
  - finding [546](#)
- organizations
  - finding your organization ID [546](#)
- output files
  - custom directory structure [220](#)

## P

- passwords
  - changing [37](#)
- POD
  - how to identify [18](#), [22](#)
- PostgreSQL
  - connection properties [109](#)
- PostgreSQL sources
  - data types mapped to Amazon Redshift targets [355](#)
  - mappings with Databricks Delta targets [357](#)
  - mappings with Google BigQuery targets [358](#)
  - mappings with Microsoft Azure Synapse Analytics targets [363](#)
  - mappings with Snowflake targets [365](#)
- profiles
  - editing [37](#)
- project folders
  - creating in Mass Ingestion [36](#)
- projects
  - copying [524](#)
  - creating in Mass Ingestion [36](#)
  - deleting [526](#)
  - exporting [534](#)
  - importing [535](#), [536](#)
- proxy settings
  - configuring on Linux [23](#), [73](#)
  - configuring on Windows [20](#), [72](#)

## R

- renaming
  - folders [524](#)
  - projects [524](#)
  - tasks [525](#), [526](#)

- requirements
  - Secure Agent [18, 21](#)
- REST API
  - deploy streaming ingestion task [488](#)
  - details of a streaming ingestion jobs [490](#)
  - history [498](#)
  - history of a streaming ingestion job [498](#)
  - list of available streaming ingestion jobs [492](#)
  - MIJobs [490, 492](#)
  - start a streaming ingestion task [489](#)
  - statistics [496](#)
  - statistics of a streaming ingestion job [496](#)
  - status [494](#)
  - status of a streaming ingestion job [494](#)
  - stop a streaming ingestion task [490](#)
  - undeploy a streaming ingestion task [489](#)
- REST V2
  - authentication
    - standard [125](#)
  - connection properties [125](#)
- restart and recovery
  - database ingestion incremental load jobs [275](#)
- runtime environments
  - migrating [531](#)

## S

- Salesforce Mass Ingestion connections
  - connection properties [111](#)
- SAP HANA Database Ingestion connections
  - connection properties [113](#)
- SAP HANA sources
  - mappings with Amazon Redshift targets [367](#)
  - mappings with Databricks Delta targets [369](#)
  - mappings with Google BigQuery targets [370](#)
  - mappings with Microsoft Azure Synapse Analytics targets [372](#)
  - mappings with Snowflake targets [374](#)
- schedules
  - migrating [532](#)
- Secure Agent Manager
  - launching [17](#)
- Secure Agent services
  - CMI Streaming Agent
    - Offline Agent [32](#)
    - Offline Mode [32](#)
  - Database Ingestion agent environment variable [27](#)
  - Database Ingestion service properties [24](#)
- Secure Agents
  - communication port [18, 22](#)
  - configuring a Windows service login [20](#)
  - domains whitelist [18, 22](#)
  - File Ingestion configuration properties [28](#)
  - installing on Linux [22](#)
  - installing on Windows [19](#)
  - IP address whitelist [18, 22](#)
  - permissions on Linux [22](#)
  - permissions on Windows [18](#)
  - registering on Linux [22](#)
  - registering on Windows [19](#)
  - requirements on Linux [21](#)
  - requirements on Windows [18](#)
  - starting on Windows [17](#)
- security questions
  - editing [37](#)
- ServiceNow Mass Ingestion connections
  - connection properties [119](#)

- Snowflake
  - target considerations for database ingestion [217](#)
- Snowflake Cloud Data Warehouse V2 targets
  - properties [437](#)
- Snowflake Data Cloud
  - authentication
    - standard [121](#)
  - connection properties [121](#)
- Snowflake Data Cloud V2 targets
  - file ingestion task [437](#)
- Snowflake targets
  - mappings with Db2 for i sources [293](#)
  - mappings with DB2 for Linux, UNIX, and Windows sources [300](#)
  - mappings with Db2 for z/OS sources [308](#)
  - mappings with Microsoft Azure SQL Database sources [320](#)
  - mappings with Microsoft SQL Server sources [320](#)
  - mappings with MySQL sources [333](#)
  - mappings with Netezza sources [341](#)
  - mappings with Oracle sources [354](#)
  - mappings with PostgreSQL sources [365](#)
  - mappings with SAP HANA sources [374](#)
  - mappings with Teradata sources [389](#)
- soft deletes
  - apply mode [230](#)
- source
  - Amazon Kinesis Streams [452](#)
  - AMQP [453](#)
- source control
  - actions [538](#)
  - best practices [544](#)
  - checking in and checking out objects [541](#)
  - checking in objects [541](#)
  - checking out objects [541](#)
  - configuring access to the repository [539](#)
  - deleting objects [542](#)
  - Explore page columns [544](#)
  - Git commands [538](#)
  - pulling objects [539](#)
  - pulling objects from the repository [540](#)
  - pulling projects and folders [540](#)
  - reverting to previous versions [542](#)
  - selecting multiple objects [543](#)
  - supported objects [537](#)
  - supported source control systems [537](#)
  - undoing a checkout [542](#)
  - unlinking objects [543](#)
  - updating organization with repository versions [539](#)
- source properties
  - flat file
    - rolling filename pattern [468](#)
  - MQTT
    - Client ID [472](#)
    - Max Queue Size [472](#)
- streaming ingestion
  - data format
    - binary [462](#)
    - JSON [462](#)
    - XML [462](#)
  - source
    - Amazon MSK [455](#)
    - Azure Event Hubs Kafka [453](#)
    - Confluent Kafka [455](#)
    - flat file [454](#)
    - Google PubSub [454](#)
    - Google PubSub properties [469](#)
    - JMS [454](#)

- streaming ingestion (*continued*)
  - source (*continued*)
    - Kafka
      - Azure Event Hubs
        - Azure Event Hubs source [467](#), [470](#)
        - namespace [467](#), [470](#)
      - Kafka Azure Event Hub properties [467](#)
      - Kafka properties [470](#)
      - MQTT [455](#)
      - OPC UA [456](#)
      - OPC UA properties [472](#)
    - target
      - Databricks Delta properties [477](#)
      - flat file [458](#)
      - Google Cloud Storage properties [479](#)
      - Google PubSub properties [480](#)
    - Target
      - Amazon MSK [460](#)
      - Azure Event Hub [461](#)
      - Confluent Kafka [460](#)
      - Databricks Delta [458](#)
      - Google Cloud Storage [459](#)
      - Google PubSub [459](#)
      - Kafka [460](#)
      - Microsoft Azure Data Lake Storage Gen2 [461](#)
    - transformations
      - combiner transformation [462](#)
      - filter transformation [463](#)
      - Python transformation [463](#)
      - Splitter transformation [464](#)
  - Streaming ingestion
    - secure agent [30](#)
    - Target
      - Amazon S3 [458](#)
      - Amazon S3 properties [475](#)
  - Streaming Ingestion
    - target
      - Azure SQL Database [460](#)
      - JDBC V2 [460](#)
      - Kinesis Firehose [457](#)
    - troubleshooting [550](#)
  - streaming ingestion jobs
    - REST API [492](#), [494](#), [496](#), [498](#)
  - streaming ingestion task
    - add transformation [482](#), [483](#)
    - Agent Parameters [486](#)
    - defining [465](#)
    - deploy [487](#)
    - email addresses [486](#)
    - Log level [486](#)
    - purge [486](#)
    - redeploy [487](#)
    - reject directory [486](#)
    - rollover [487](#)
    - runtime options [486](#)
    - source configuration [465](#)
    - target configuration [474](#)
    - transformation configuration [482](#)
    - undeploy [487](#)
  - streaming ingestion tasks
    - Microsoft Azure Data Lake Storage Gen2 target properties [481](#)
    - prerequisites [465](#)
    - REST API [488](#)–[490](#)
  - streaming ingestion
    - resume [488](#)
    - stop [488](#)

## T

- tags
  - adding to an ingestion task [528](#)
  - creating [527](#)
  - deleting [529](#)
  - editing [529](#)
  - properties [529](#)
- target
  - Amazon Kinesis Streams [457](#)
- target properties
  - flat file
    - rolling filename pattern [478](#)
- target properties for database ingestion
  - Amazon Redshift target properties [246](#)
  - Amazon S3 target properties [246](#)
  - Databricks Delta target properties [250](#)
  - flat file target properties [250](#)
  - Google BigQuery target properties [253](#)
  - Google Cloud Storage target properties [253](#)
  - Kafka properties [257](#)
  - Microsoft Azure Data Lake Storage target properties [260](#)
  - Microsoft Azure Synapse Analytics target properties [264](#)
  - Oracle target properties [264](#)
  - Snowflake target properties [265](#)
- targets, database ingestion
  - Amazon S3 [211](#)
  - flat file [211](#)
  - Google Cloud Storage [211](#)
  - Microsoft Azure Data Lake Storage [211](#)
  - Microsoft Azure Synapse Analytics [215](#)
  - Snowflake [217](#)
- Targets, database ingestion
  - Apache Kafka [214](#)
  - Confluent Kafka [214](#)
  - Kafka-enabled Azure Event Hubs [214](#)
- tasks
  - copying [524](#)
  - deleting [526](#)
  - moving [524](#)
  - renaming [525](#), [526](#)
- Teradata connection
  - connection properties [127](#)
- Teradata sources
  - data types mapped to Amazon Redshift targets [375](#)
  - mappings with Databricks Delta targets [377](#)
  - mappings with Google BigQuery targets [378](#)
  - mappings with Microsoft Azure Synapse Analytics targets [387](#)
  - mappings with Snowflake targets [389](#)
- time zones
  - changing user profile [37](#)
- Transformations
  - combiner transformation [462](#)
  - filter transformation [463](#)
  - Format Converter transformation [464](#)
  - Python transformation [463](#)
  - Splitter transformation [464](#)
- Troubleshooting
  - database ingestion task [547](#)
  - streaming ingestion task [550](#)

## U

- user permissions
  - setting permissions on an ingestion task [530](#)
- user profiles
  - editing [37](#)

## W

### whitelist

- Secure Agent domains [18](#), [22](#)

- Secure Agent IP addresses [18](#), [22](#)

### Windows

- configuring proxy settings [20](#), [72](#)

### Windows service

- configuring Secure Agent login [20](#)

Workday Mass Ingestion connections  
connection properties [129](#)

## Z

Zendesk Mass Ingestion connections  
connection properties [130](#)