



Informatica® Cloud Data Quality
May 2022

Introduction

Informatica Cloud Data Quality Introduction
May 2022
May 2022

© Copyright Informatica LLC 2019, 2023

Publication Date: 2023-02-23

Table of Contents

Chapter 1: Introduction to Data Quality.....	4
Informatica Intelligent Cloud Services.	4
Data Quality.	5
Data quality life cycle.	5
Data Integration.	6
Data Integration tasks	7
Data Integration Elastic.	7
Data Profiling.	7
Administrator.	8
Monitor.	9
Informatica Intelligent Cloud Services security.	10
 Chapter 2: Data Quality tools.....	 11
Projects.	11
Assets.	11
Explore page.	12
 Chapter 3: Informatica resources.....	 14
Informatica Intelligent Cloud Services web site.	14
Informatica Intelligent Cloud Services Communities.	14
Informatica Intelligent Cloud Services Marketplace.	14
Informatica Intelligent Cloud Services Trust Center.	14
Informatica Product Availability Matrices.	15
Informatica Documentation.	15
Data Integration connector documentation.	15
Informatica Knowledge Base.	15
Informatica Global Customer Support.	16
 Index.....	 17

CHAPTER 1

Introduction to Data Quality

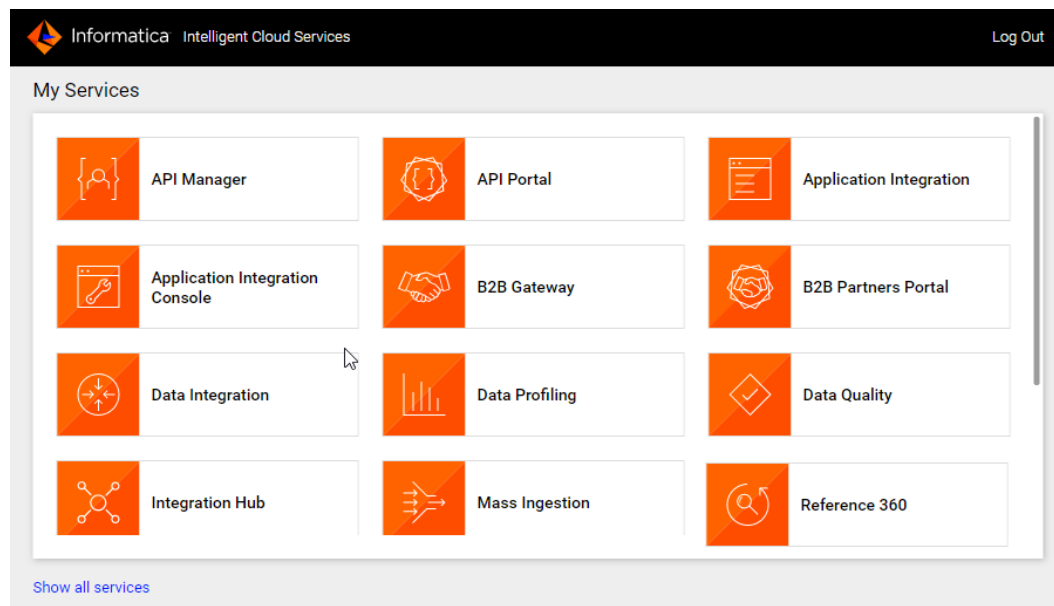
Data Quality is one of several services available in Informatica Intelligent Cloud Services. Data Quality is an on-demand subscription service that provides all the features that you need to verify, standardize, and improve the quality of your business data.

Informatica Intelligent Cloud Services

When you log in to Informatica Intelligent Cloud Services, the **My Services** page displays the services that apply to data integration. The **My Services** page might also include other services that you subscribe to and trial versions of other services.

The following services apply to data integration:

- **Data Integration.** Manage data integration projects, create assets such as mappings, tasks, and taskflows, and run tasks and taskflows.
- **Administrator.** Perform administrative tasks such as creating users and managing privileges, downloading Secure Agents and configuring runtime environments, configuring connections, managing licenses, and creating schedules.
- **Monitor.** Check the status of tasks that are in progress and tasks that have completed.

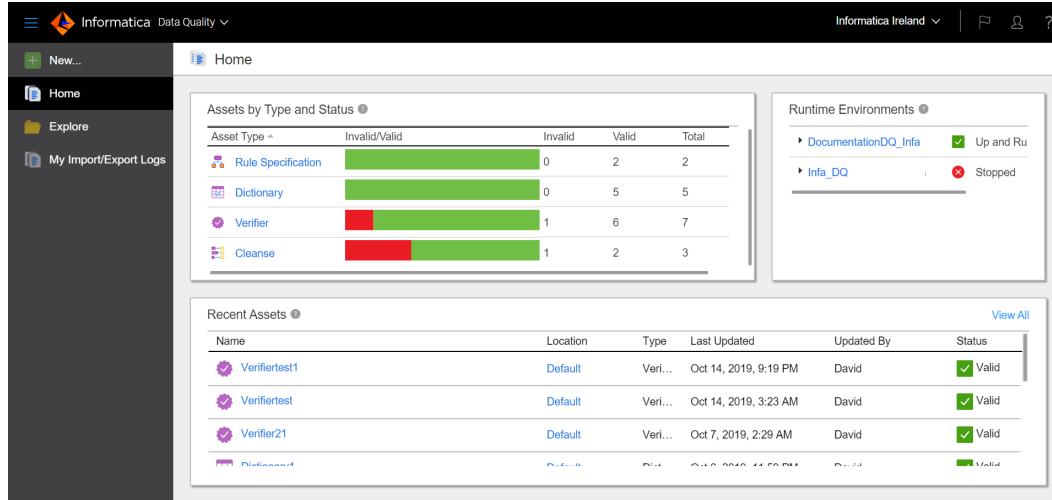


You might want to use other services with Data Integration such as DiscoveryIQ.

Data Quality

Use Data Quality to create data quality assets. Add the assets to transformations in a mapping in Data Integration.

When you select Data Quality from the **My Services** page, the **Home** page appears, as shown in the following image:



The **Home** page displays the following panels:

- Runtime Environments. The status of the runtime environments that are available to you.
- Assets by Type and Status. A summary of the assets that you can access.
- Recent Assets. The assets that you most recently modified.

You can access the following pages from the navigation bar for Data Quality:

- New. Create dictionaries, rule specifications, cleanse assets, deduplicate assets, parse assets, labeler assets, and verifier assets.
- Home. Return to the Home page.
- Explore. Manage data quality projects and assets.
- My Import/Export Logs. View the status of your imports and exports.

When you switch from Data Quality to another service, the panels and the options in the navigation bar change to suit the service.

Data quality life cycle

The assets that you configure in Data Quality constitute a set of operations that you can add to transformations in a mapping in Data Integration.

To understand and improve the quality of your data, you can move the data through the following stages:

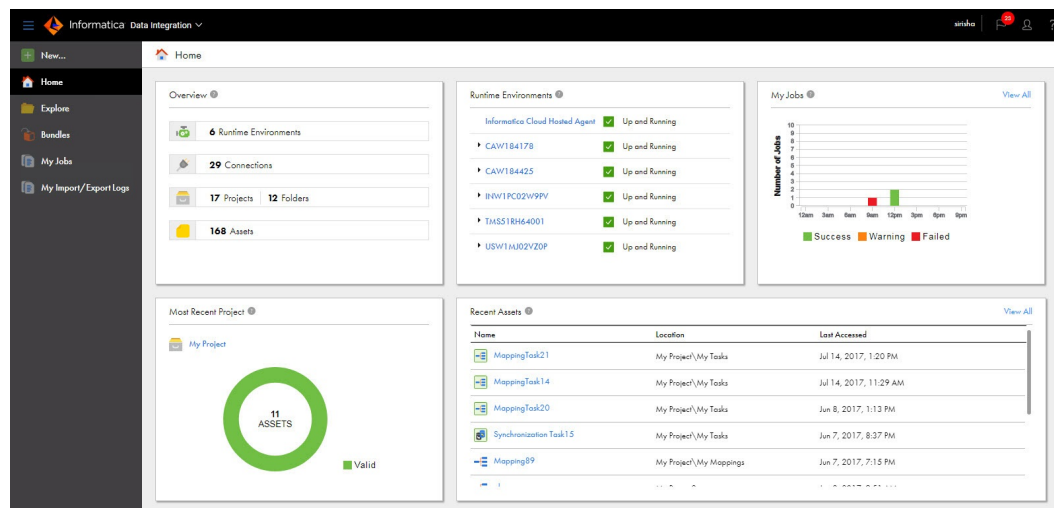
1. **Discover.** Analyze the content and structure of the source data.
2. **Design.** Create assets in Data Quality to address the issues that you find in the source data.
3. **Apply.** Add the assets to one or more mappings in Data Integration, and run the mappings on the data.
4. **Measure.** Review the results of the mappings. Optionally, update the assets in Data Quality and run the mappings again in Data Integration.

Note: You can run a profile to analyze the source data as part of step 1. Use the Data Profiling service to create and run a profile.

Data Integration

Use Data Integration to create data integration projects and run data integration tasks.

When you select Data Integration from the **My Services** page, the Data Integration **Home** page appears, as shown in the following image:



The **Home** page displays the following panels:

- Overview. A summary of projects, folders, connections, and assets available to your organization.
- Runtime Environments. The status of runtime environments that are available to you.
- My Jobs. A list of jobs that you ran in the last 24 hours.
- Most Recent Project. The last project that you created or that contains the most recently modified asset.
- Recent Assets. The assets that you most recently modified.

You can access the following pages from the navigation bar for Data Integration:

- New. Create new data integration assets such as mappings, tasks, and taskflows. You can create assets from scratch or use templates as a head start.
- Home. Return to the Home page.
- Explore. Create and manage projects and assets.
- Bundles. Create and manage bundles of assets to share with your organization or other organizations.
- My Jobs. View the status of your mapping, task, and taskflow instances.
- My Import/Export Logs. View the status of your imports and exports.

When you switch from Data Integration to another service such as Administrator or Monitor, the options in the navigation bar change to suit the service.

Data Integration tasks

A Data Integration task is a process that you configure to analyze, extract, transform, and load data. You can run individual tasks manually or set tasks to run on a schedule.

For example, you might schedule a mapping task to run a mapping that includes a Rule Specification transformation.

You can use multiple types of task to integrate data in Data Integration. For more information, see the Data Integration online help.

Note: Data Integration tasks that contain Cleanse, Deduplicate, Parse, Labeler, Rule Specification, or Verifier transformations are not certified for serverless runtime execution. If you run a Data Integration task that contains a data quality asset in a serverless runtime environment, the task can count towards your serverless compute hour usage and quota.

Data Integration Elastic

Use Data Integration Elastic to run data integration jobs on an elastic cluster that adapts to the number of jobs and the size of the data that jobs process.

Data Integration Elastic uses an elastic mapping to define data flow logic from a source to a target.

When you run a mapping task that is based on an elastic mapping, a Secure Agent creates an elastic cluster and the Serverless Spark engine runs on the cluster to process the underlying mapping. If the cluster becomes idle, the Secure Agent terminates the cluster so that you pay only for the resources that you use.

To use Data Integration Elastic, your organization must have the appropriate licenses.

Data Profiling

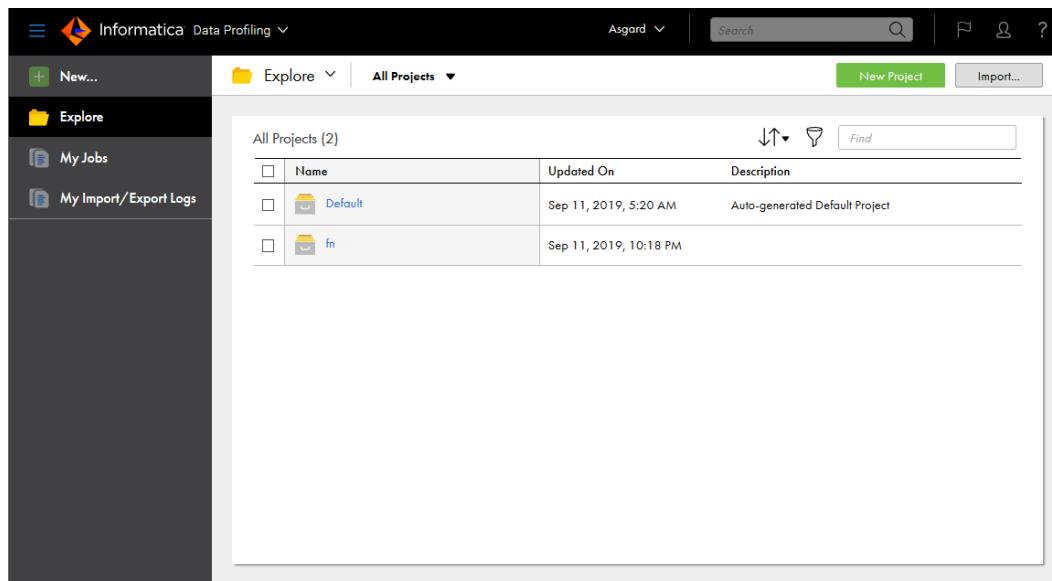
Use Data Profiling to create and run data profiling tasks. In Data Profiling, data profiling tasks are also called profiles.

When you run a profile on a source object, the results include the following column statistics:

- Number of distinct, non-distinct, and null values
- Percentage of distinct, non-distinct, and null values
- Documented and inferred data types
- Number of patterns
- Percentage of top pattern
- Maximum and minimum length of values
- Maximum and minimum values
- Average, sum, and standard deviation for numeric data types
- Value frequencies
- Outliers

You can run data profiling tasks on relational data sources and flat file data sources.

When you select Data Profiling from the **My Services** page, the **Explore** page appears as shown in the following image:



You can access the following pages from the navigation bar for Data Profiling:

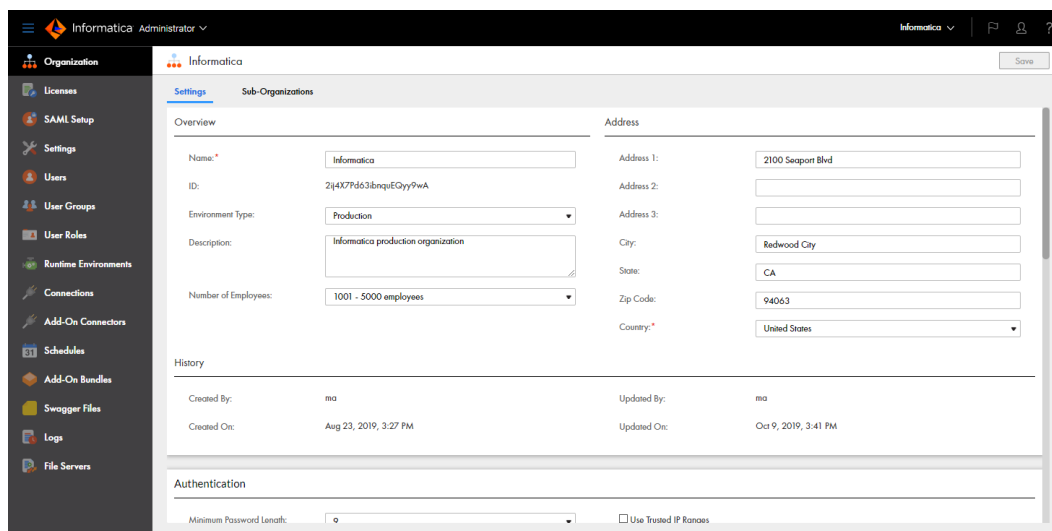
- New. Create data profiling tasks.
- Explore. Create and manage projects, folders, and assets.
- My Jobs. View the status of your data profiling jobs.
- My Import/Export Logs. View the status of your imports and exports.

Note: You can open a profiling task from the Explore page in Data Quality.

Administrator

Administrative tasks for Informatica Intelligent Cloud Services are performed in the Administrator service.

When you select Administrator from the **My Services** page, the navigation bar provides options for Informatica Intelligent Cloud Services administration, as shown in the following image:



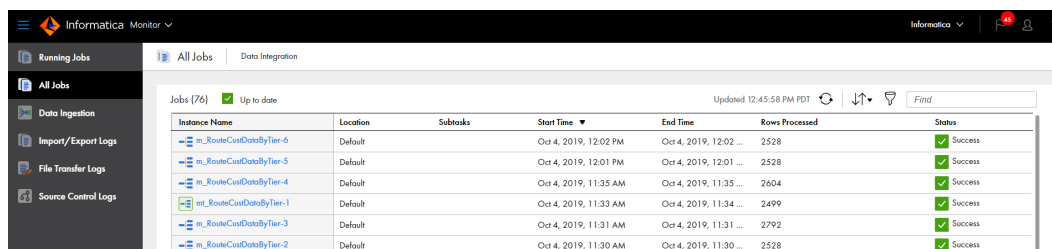
The following options are available based on your licenses:

- Organization. Manage your organization details and your sub-organizations.
- Licenses. View licenses for your organization and manage your sub-organization licenses.
- Ecosystem single sign-on. When you create your organization through Microsoft Azure, configure some single sign-on properties for Azure users.
- SAML Setup. Configure SAML single sign-on settings.
- Settings. Configure source control settings and upgrade settings for Secure Agent services.
- Users. Add new users. Configure user groups and roles for individual users.
- User Groups. Create user groups. Assign users to user groups.
- User Roles. Manage user roles for your organization. Configure the privileges associated with each role.
- Runtime Environments. Configure Secure Agents and Secure Agent groups.
- Connections. Create and manage connections.
- Add-On Connectors. Install add-on connectors.
- Schedules. Create and manage schedules for tasks.
- Add-On Bundles. Install, uninstall, upgrade, and copy bundles for your organization.
- Swagger Files. Generate swagger files for REST V2 connections.
- Logs. View asset and security logs.
- File servers. Configure file transfer servers and file transfer users to allow remote partners to send files to your organization.

Monitor

Use Monitor to monitor jobs, imports, and exports in your organization. A job is an instance of a mapping, task, or taskflow.

When you select Monitor from the **My Services** page, the navigation bar provides options for monitoring activity, as shown in the following image:



The screenshot shows the Informatica Monitor interface. On the left is a navigation bar with options: Running Jobs, All Jobs, Data Ingestion, Import/Export Logs, File Transfer Logs, and Source Control Logs. The main area displays a table of jobs. The table has columns for Instance Name, Location, Subtasks, Start Time, End Time, Rows Processed, and Status. The status for all jobs shown is 'Success'.

Instance Name	Location	Subtasks	Start Time	End Time	Rows Processed	Status
m_RouteCustDataByTier-6	Default		Oct 4, 2019, 12:02 PM	Oct 4, 2019, 12:02 ...	2528	Success
m_RouteCustDataByTier-5	Default		Oct 4, 2019, 12:01 PM	Oct 4, 2019, 12:01 ...	2528	Success
m_RouteCustDataByTier-4	Default		Oct 4, 2019, 11:35 AM	Oct 4, 2019, 11:35 ...	2604	Success
m_RouteCustDataByTier-1	Default		Oct 4, 2019, 11:33 AM	Oct 4, 2019, 11:34 ...	2499	Success
m_RouteCustDataByTier-3	Default		Oct 4, 2019, 11:31 AM	Oct 4, 2019, 11:31 ...	2792	Success
m_RouteCustDataByTier-2	Default		Oct 4, 2019, 11:30 AM	Oct 4, 2019, 11:30 ...	2528	Success

The navigation bar provides the following options:

- Running Jobs. Provides run-time details about the Data Integration jobs that are running or have completed within the last five minutes.
- All Jobs. Provides details about all Data Integration jobs in the organization.
- Data Ingestion. Provides details about the Mass Ingestion jobs in the organization.
- Import/Export Logs. Provides details about imports and exports that are running and that have completed.
- File Transfer Logs. Provides details about the file transfers in the organization.

- Source Control Logs. Provides a log of actions on source-controlled objects in the organization.

To view details about a specific job, import instance, or export instance, click the instance name.

Informatica Intelligent Cloud Services security

Informatica Intelligent Cloud Services uses authentication and encryption to ensure that data is secure and available only to users within the organization. Informatica Intelligent Cloud Services uses industry approved algorithms to encrypt sensitive information. All customer data is encrypted at rest using an AES-256 key. Data in transit is encrypted using the TLS 1.2 protocol.

For more information, see the [Informatica Intelligent Cloud Services Trust Center](#).

CHAPTER 2

Data Quality tools

Informatica Intelligent Cloud Services provides the following tools that you can use to satisfy your data integration business needs:

- Projects
- Assets
- Explore page

Projects

Informatica Intelligent Cloud Services assets are organized in projects. You can create multiple folders within a project to contain the assets that you need for your project.

For example, you might have a business need to provide monthly analyses of your company's sales activity. You create a project called Monthly Sales Analysis. You create folders that contain the assets that you use on a monthly basis to integrate data from various systems to perform your analysis.

The types of assets that you use in your projects depends on the services that you use. For example, a Data Integration project can include assets such as mappings, tasks, and taskflows.

Assets

In Data Quality, an asset is an object that defines part or all of a data quality operation.

In Data Integration, assets are mappings, taskflows, and tasks such as mapping tasks, synchronization tasks, and replication tasks. Data Integration assets also include components such as business services, mapplets, and hierarchical schemas.

You can create the following assets in Data Quality: cleanse, deduplicate, parse, labeler, rule specification, and verifier. Add the assets to a Cleanse transformation, Deduplicate transformation, Parse transformation, Labeler transformation, Rule Specification transformation, and Verifier transformation respectively when you design a mapping in Data Integration. You can add each transformation to a non-elastic mapping. You can add a Cleanse transformation, Parse transformation, Rule Specification transformation, and Verifier transformation to an elastic mapping.

You can also create dictionaries in Data Quality that other data quality assets can use. A dictionary is a table of reference data that a mapping can read to evaluate the accuracy or usefulness of source data.

Create assets to satisfy your data objectives. Templates are available for certain asset types. You can use a template as-is, or you can use it as a head start and customize it to meet your needs.

You can use assets in multiple projects. All assets must be part of at least one project. If you create an asset without specifying a project, the asset is created in the default project.

Explore page

Use the **Explore** page to work with your Data Quality projects and assets. The **Explore** page shows the assets that you can work with in Data Quality.

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.

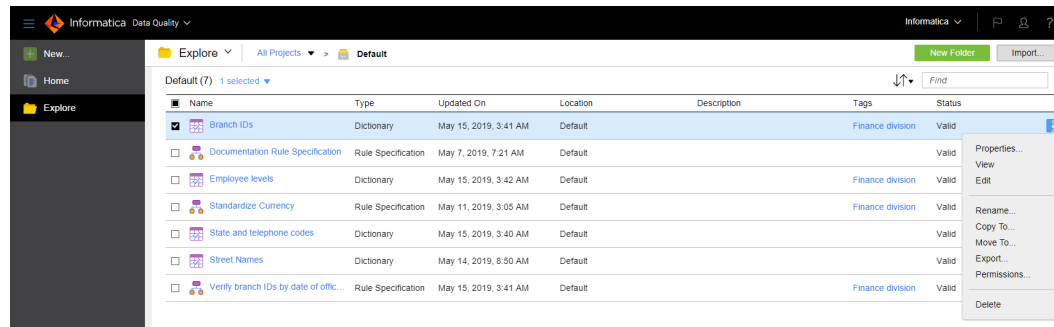
The following characters cannot be used on the **Explore** page:

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

Working with projects and assets on the Explore page

You can perform actions on projects and assets on the Explore page. To see the actions that you can perform on a project or asset, click **Actions** in the row that contains the project or asset.

The following image shows the menu of actions that you can perform on an asset:



The **Actions** menu lists the actions you can perform based on your user role privileges and the permissions specified for the selected asset.

You can also perform an action on multiple objects at one time. Select the check box to the left of each object, or select the Select All check box to select all of the objects that are displayed on the current page. After you select the objects, right-click the objects to open a menu and choose the action

Customizing the Explorer page

You can display, hide, or rearrange object properties on the Explore page. To display or hide properties, right-click the column heading area and check or uncheck the properties.

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

CHAPTER 3

Informatica resources

In addition to the online help, you can find information about Informatica Intelligent Cloud Services using the following resources.

Informatica Intelligent Cloud Services web site

You can access the Informatica Intelligent Cloud Services web site at <http://www.informatica.com/cloud>.

Informatica Intelligent Cloud Services Communities

Use the Informatica Intelligent Cloud Services Community to discuss and resolve technical issues. You can also find technical tips, documentation updates, and answers to frequently asked questions.

Access the Informatica Intelligent Cloud Services Community at:

<https://network.informatica.com/community/informatica-network/products/cloud-integration>

Developers can learn more and share tips at the Cloud Developer community:

<https://network.informatica.com/community/informatica-network/products/cloud-integration/cloud-developers>

Informatica Intelligent Cloud Services Marketplace

Visit the Informatica Marketplace to try and buy Data Integration Connectors, templates, and maplets:

<https://marketplace.informatica.com/>

Informatica Intelligent Cloud Services Trust Center

The Informatica Intelligent Cloud Services Trust Center provides information about Informatica security policies and real-time system availability.

You can access the trust center at <https://www.informatica.com/trust-center.html>.

Subscribe to the Informatica Intelligent Cloud Services Trust Center to receive upgrade, maintenance, and incident notifications. The [Informatica Intelligent Cloud Services Status](#) page displays the production status of all the Informatica cloud products. All maintenance updates are posted to this page, and during an outage, it will have the most current information. To ensure you are notified of updates and outages, you can subscribe to receive updates for a single component or all Informatica Intelligent Cloud Services components. Subscribing to all components is the best way to be certain you never miss an update.

To subscribe, go to <https://status.informatica.com/> and click **SUBSCRIBE TO UPDATES**. You can then choose to receive notifications sent as emails, SMS text messages, webhooks, RSS feeds, or any combination of the four.

Informatica Product Availability Matrices

Product Availability Matrices (PAMs) indicate the versions of the operating systems, databases, and types of data sources and targets that a product release supports. You can browse the Informatica PAMs at <https://network.informatica.com/community/informatica-network/product-availability-matrices>.

Informatica Documentation

Use the Informatica Documentation Portal to explore an extensive library of documentation for current and recent product releases. To explore the Documentation Portal, visit <https://docs.informatica.com>.

If you have questions, comments, or ideas about the product documentation, contact the Informatica Documentation team at infa_documentation@informatica.com.

Data Integration connector documentation

You can access documentation for Data Integration Connectors at the Documentation Portal. To explore the Documentation Portal, visit <https://docs.informatica.com>.

Informatica Knowledge Base

Use the Informatica Knowledge Base to find product resources such as how-to articles, best practices, video tutorials, and answers to frequently asked questions.

To search the Knowledge Base, visit <https://search.informatica.com>. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at KB_Feedback@informatica.com.

Informatica Global Customer Support

You can contact a Customer Support Center by telephone or online.

For online support, click **Submit Support Request** in Informatica Intelligent Cloud Services. You can also use Online Support to log a case. Online Support requires a login. You can request a login at <https://network.informatica.com/welcome>.

The telephone numbers for Informatica Global Customer Support are available from the Informatica web site at <https://www.informatica.com/services-and-training/support-services/contact-us.html>.

INDEX

A

assets [11](#), [12](#)

C

Cloud Application Integration community
URL [14](#)
Cloud Developer community
URL [14](#)

D

Data Integration community
URL [14](#)
Data Quality
overview [4](#)

E

Explore page [12](#)

I

Informatica Intelligent Cloud Services
web site [14](#)

M

maintenance outages [14](#)

P

project folders [11](#), [12](#)
projects [11](#), [12](#)

S

security [10](#)
status
Informatica Intelligent Cloud Services [14](#)
system status [14](#)

T

TLS
security [10](#)
trust site
description [14](#)

U

upgrade notifications [14](#)

W

web site [14](#)