



Informatica®

9.6.1 HotFix 1

# 引用数据指南

## Informatica 引用数据指南

9.6.1 HotFix 1  
2014 年 9 月

© 版权所有 Informatica LLC 2013, 2018

本软件和文档包含 Informatica Corporation 的专有信息，基于一份包含使用及披露限制的许可协议而提供，并受版权法保护。禁止对本软件进行反向工程处理。未事先征得 Informatica Corporation 同意，不得以任何形式、通过任何手段（电子、影印、录制或其他手段）复制或传送本文档的任何部分。本软件受美国和/或国际专利以及其他正在申请的专利的保护。

美国政府使用、复制或披露本软件受适用的软件许可协议中规定的限制以及适用的 DFARS 227.7202-1(a) 和 227.7702-3(a) (1995)、DFARS 252.227-7013(C)(1)(ii) (OCT 1988)、FAR 12.212(a) (1995)、FAR 52.227-19 或 FAR 52.227-14 (ALT III) 中的限制之制约。

本产品或文档中的信息如有更改，恕不另行通知。如发现本产品或文档中有什么问题，请以书面形式向我们报告。

Informatica、Informatica Platform、Informatica Data Services、PowerCenter、PowerCenterRT、PowerCenter Connect、PowerCenter Data Analyzer、PowerExchange、PowerMart、Metadata Manager、Informatica Data Quality、Informatica Data Explorer、Informatica B2B Data Transformation、Informatica B2B Data Exchange Informatica On Demand、Informatica Identity Resolution、Informatica Application Information Lifecycle Management、Informatica Complex Event Processing、Ultra Messaging 和 Informatica Master Data Management 是 Informatica Corporation 在美国和全球其他辖区的商标或注册商标。所有其他公司和产品名称可能是其各自所有者的商业名称或商标。

本软件和/或文档的某些部分受第三方版权制约，包括但不限于：版权所有 DataDirect Technologies。保留所有权利。版权所有 (C) Sun Microsystems。保留所有权利。版权所有 (C) RSA Security Inc. 保留所有权利。版权所有 (C) Ordinal Technology Corp. 保留所有权利。版权所有 (C) Aandacht c.v. 保留所有权利。版权所有 Genivia, Inc. 保留所有权利。版权所有 Isomorphic Software。保留所有权利。版权所有 (C) Meta Integration Technology, Inc. 保留所有权利。版权所有 (C) Intalio。保留所有权利。版权所有 (C) Oracle。保留所有权利。版权所有 (C) Adobe Systems Incorporated。保留所有权利。版权所有 (C) DataArt, Inc. 保留所有权利。版权所有 (C) ComponentSource。保留所有权利。版权所有 (C) Microsoft Corporation。保留所有权利。版权所有 (C) Rogue Wave Software, Inc. 保留所有权利。版权所有 (C) Teradata Corporation。保留所有权利。版权所有 (C) Yahoo! Inc. 保留所有权利。版权所有 (C) Glyph & Cog, LLC。保留所有权利。版权所有 (C) Thinkmap, Inc. 保留所有权利。版权所有 (C) Clearpace Software Limited。保留所有权利。版权所有 (C) Information Builders, Inc. 保留所有权利。版权所有 (C) OSS Nokalva, Inc. 保留所有权利。版权所有 Edifecs, Inc. 保留所有权利。版权所有 Cleo Communications, Inc. 保留所有权利。版权所有 (C) International Organization for Standardization 1986。保留所有权利。版权所有 (C) ej-technologies GmbH。保留所有权利。版权所有 (C) Jaspersoft Corporation。保留所有权利。版权所有 (C) International Business Machines Corporation。保留所有权利。版权所有 (C) yWorks GmbH。保留所有权利。版权所有 (C) Lucent Technologies。保留所有权利。版权所有 (C) University of Toronto。保留所有权利。版权所有 (C) Daniel Veillard。保留所有权利。版权所有 (C) Unicode, Inc. 版权所有 IBM Corp. 保留所有权利。版权所有 (C) MicroQuill Software Publishing, Inc. 保留所有权利。版权所有 (C) PassMark Software Pty Ltd. 保留所有权利。版权所有 (C) LogiXML, Inc. 保留所有权利。版权所有 (C) 2003-2010 Lorenzi Davide, 保留所有权利。版权所有 (C) Red Hat, Inc. 保留所有权利。版权所有 (C) The Board of Trustees of the Leland Stanford Junior University。保留所有权利。版权所有 (C) EMC Corporation。保留所有权利。版权所有 (C) Flexera Software。保留所有权利。版权所有 (C) Jinfonet Software。保留所有权利。版权所有 (C) Apple Inc. 保留所有权利。版权所有 (C) Telerik Inc. 保留所有权利。版权所有 (C) BEA Systems。保留所有权利。版权所有 (C) PDFlib GmbH。保留所有权利。版权所有 (C) Orientation in Objects GmbH。保留所有权利。版权所有 (C) Tanuki Software, Ltd. 保留所有权利。版权所有 (C) Ricebridge。保留所有权利。版权所有 (C) Sencha, Inc. 保留所有权利。版权所有 (C) Scalable Systems, Inc. 保留所有权利。

本产品包括由 Apache Software Foundation (<http://www.apache.org/>) 开发的软件和/或在不同 Apache 许可证版本（以下简称“许可证”）下许可的其他软件。您可从 <http://www.apache.org/licenses/> 获取这些许可证的副本。除非适用法律要求或者有相应书面协议，否则依据这些“许可证”分发的软件以“原样”提供，不附带任何明示或暗示的担保或条件。请参阅“许可证”中规定的具体语言管理权限和限制。

本产品包括由 Mozilla (<http://www.mozilla.org/>) 开发的软件、由 JBoss Group, LLC 开发的软件（版权所有 JBoss Group, LLC 保留所有权利）、由 Bruno Lowagie 和 Paulo Soares 开发的软件（版权所有 (C) 1999-2006 Bruno Lowagie 和 Paulo Soares）以及在 <http://www.gnu.org/licenses/lgpl.html> 网站上的不同版本 GNU Lesser General 公共许可协议下许可的软件。这些材料由 Informatica 按“原样”免费提供，不附带任何明示或暗示的担保，包括但不限于适销性和特定用途适用性的暗示担保。

本产品包括 ACE(TM) 和 TAO(TM) 软件，这些软件版权归 Douglas C. Schmidt 及其在华盛顿大学、加利福尼亚大学欧芬分校以及范德堡大学的研发团队所有（版权所有 (C) 1993-2006，保留所有权利）。

本产品包括由 OpenSSL Project 开发并在 OpenSSL Toolkit（版权所有 OpenSSL Project。保留所有权利）中使用的软件，该软件的再分发受 <http://www.openssl.org> 和 <http://www.openssl.org/source/license.html> 上规定条款之制约。

本产品包括 Curl 软件，版权所有 1996-2013, Daniel Stenberg <[daniel@haxx.se](mailto:daniel@haxx.se)>。保留所有权利。有关该软件的权限和限制受 <http://curl.haxx.se/docs/copyright.html> 上规定条款之制约。允许出于任何目的以免费或收费形式使用、复制、修改和分发该软件，但前提是所有副本均应注明上述版权声明以及本许可声明。

本产品包括由 MetaStuff, Ltd. 开发的软件，版权所有 2001-2005 ((C)) MetaStuff, Ltd. 保留所有权利。有关该软件的权限和限制受 <http://www.dom4j.org/license.html> 上规定条款之制约。

本产品包括由 Dojo Foundation 开发的软件，版权所有 (C) 2004-2007, Dojo Foundation。保留所有权利。有关该软件的权限和限制受 <http://dojotoolkit.org/license> 上规定条款之制约。

本产品包括 ICU 软件，版权所有 International Business Machines Corporation 和其他方。保留所有权利。有关该软件的权限和限制受 <http://source.icu-project.org/repos/icu/icu/trunk/license.html> 上规定条款之制约。

本产品包括由 Per Bothner 开发的软件，版权所有 (C) 1996-2006 Per Bothner。保留所有权利。<http://www.gnu.org/software/kawa/Software-License.html> 上的许可证中规定了您使用这些材料的权利。

本产品包括 OSSP UUID 软件，版权所有 (C) 2002 Ralf S. Engelschall，版权所有 (C) 2002 OSSP Project，版权所有 (C) 2002 Cable & Wireless Deutschland。有关该软件的权限和限制受 <http://www.opensource.org/licenses/mit-license.php> 上规定条款之制约。

本产品包括由 Boost (<http://www.boost.org/>) 开发的软件或在 Boost 软件许可证下许可的软件。有关该软件的权限和限制受 [http://www.boost.org/LICENSE\\_1\\_0.txt](http://www.boost.org/LICENSE_1_0.txt) 上规定条款之制约。

本产品包括由 University of Cambridge 开发的软件，版权所有 (C) 1997-2007 University of Cambridge。有关该软件的权限和限制受 <http://www.pcre.org/license.txt> 上规定条款之制约。

本产品包括由 The Eclipse Foundation 开发的软件，版权所有 (C) 2007 The Eclipse Foundation。保留所有权利。有关该软件的权限和限制受 <http://www.eclipse.org/org/documents/epl-v10.php> 和 <http://www.eclipse.org/org/documents/edl-v10.php> 上规定条款之制约。

本产品包括在 <http://www.tcl.tk/software/tcltk/license.html>、<http://www.bosrup.com/web/overlib/?License>、<http://www.stlport.org/doc/license.html>、<http://asm.ow2.org/license.html>、<http://www.cryptix.org/LICENSE.TXT>、<http://hsqldb.org/web/hsqLicense.html>、<http://httpunit.sourceforge.net/doc/license.html>、<http://jung.sourceforge.net/license.txt>、[http://www.gzip.org/zlib/zlib\\_license.html](http://www.gzip.org/zlib/zlib_license.html)、<http://www.openldap.org/software/release/license.html>、<http://www.libssh2.org>、<http://slf4j.org/license.html>、<http://www.sente.ch/software/OpenSourceLicense.html>、<http://fusesource.com/downloads/license-agreements/fuse-message-broker-v-5-3-license-agreement>、<http://antlr.org/license.html>、<http://aopalliance.sourceforge.net/>、<http://www.bouncycastle.org/licence.html>、<http://www.jgraph.com/jgraphdownload.html>、<http://www.jcraft.com/jsch/LICENSE.txt>、[http://jotm.objectweb.org/bsd\\_license.html](http://jotm.objectweb.org/bsd_license.html)；<http://www.w3.org/Consortium/Legal/2002/copyright-software-20021231>；<http://www.slf4j.org/license.html>；<http://nanoxml.sourceforge.net/orig/copyright.html>；<http://www.json.org/license.html>；<http://forge.ow2.org/projects/javaservice/>，<http://www.postgresql.org/about/licence.html>，<http://www.sqlite.org/copyright.html>，<http://www.tcl.tk/software/tcltk/license.html>，<http://www.jaxen.org/faq.html>，<http://www.jdom.org/docs/faq.html>，<http://www.slf4j.org/>

license.html; <http://www.iodbc.org/dataspace/iodbc/wiki/iodbc/License>; <http://www.keplerproject.org/md5/license.html>; <http://www.toedter.com/en/jcalendar/license.html>; <http://www.edankert.com/bounce/index.html>; <http://www.net-snmp.org/about/license.html>; <http://www.openmdx.org/#FAQ>; [http://www.php.net/license/3\\_01.txt](http://www.php.net/license/3_01.txt); <http://srp.stanford.edu/license.txt>; <http://www.schneier.com/blowfish.html>; <http://www.jmock.org/license.html>; <http://xsom.java.net>; <http://benalman.com/about/license/>; <https://github.com/CreateJS/EaselJS/blob/master/src/easeljs/display/Bitmap.js>; <http://www.h2database.com/html/license.html#summary>; <http://jsoncpp.sourceforge.net/LICENSE>; <http://jdbc.postgresql.org/license.html>; <http://protobuf.googlecode.com/svn/trunk/src/google/protobuf/descriptor.proto>; <https://github.com/rantav/hector/blob/master/LICENSE>; <http://web.mit.edu/Kerberos/krb5-current/doc/mitK5license.html>; <http://jibx.sourceforge.net/jibx-license.html>; <https://github.com/lyokato/libgeohash/blob/master/LICENSE>; and <https://github.com/hjiang/jsonxx/blob/master/LICENSE>.

本产品包括在 Academic 免费许可证 (<http://www.opensource.org/licenses/afl-3.0.php>)、通用开发和分发许可证 (<http://www.opensource.org/licenses/cddl1.php>)、通用公共许可证 (<http://www.opensource.org/licenses/cpl1.0.php>)、Sun Binary Code 许可协议补充许可条款、BSD 许可证 (<http://www.opensource.org/licenses/bsd-license.php>)、新 BSD 许可证 (<http://opensource.org/licenses/BSD-3-Clause>)、MIT 许可证 (<http://www.opensource.org/licenses/mit-license.php>)、Artistic 许可证 (<http://www.opensource.org/licenses/artistic-license-1.0>) 以及原始开发者公共许可证版本 1.0 (<http://www.firebirds.org/en/initial-developer-s-public-license-version-1-0/>) 下许可的软件。

本产品包括由 Joe Walnes 和 XStream Committers 开发的软件，版权所有 (C) 2003-2006 Joe Walnes，2006-2007 XStream Committers。保留所有权利。有关该软件的权限和限制受 <http://xstream.codehaus.org/license.html> 上规定条款之制约。本产品包括由 Indiana University Extreme! Lab 开发的软件。有关详细信息，请访问 <http://www.extreme.indiana.edu/>。

本产品包括软件版权所有 (c) 2013 Frank Balluffi 和 Markus Moeller。保留所有权利。有关此软件的权限和限制受 MIT 许可证上规定条款之制约。

本软件受美国专利号 5,794,246、6,014,670、6,016,501、6,029,178、6,032,158、6,035,307、6,044,374、6,092,086、6,208,990、6,339,775、6,640,226、6,789,096、6,823,373、6,850,947、6,895,471、7,117,215、7,162,643、7,243,110、7,254,590、7,281,001、7,421,458、7,496,588、7,523,121、7,584,422、7,676,516、7,720,842、7,721,270、7,774,791、8,065,266、8,150,803、8,166,048、8,166,071、8,200,622、8,224,873、8,271,477、8,327,419、8,386,435、8,392,460、8,453,159、8,458,230 和 RE44,478、国际专利及其他正在申请的专利保护。

免责声明：Informatica Corporation 以“原样”提供本文档，不附带任何明示或暗示的担保，包括但不限于非侵权、适销性或特定用途适用性的暗示担保。Informatica Corporation 不保证本软件或文档中没有错误。本软件或文档中提供的信息可能包括技术上的不准确性或排字错误。本软件和文档中包含的信息随时可能更改，恕不另行通知。

## 声明

本 Informatica 产品（以下称“软件”）包括由 Progress Software Corporation 的运营公司 DataDirect Technologies（以下称“DataDirect”）提供的某些驱动程序（以下称“DataDirect 驱动程序”），受以下条款和条件制约：

1. DataDirect 驱动程序以“原样”提供，不附带任何明示或暗示的担保，包括但不限于适销性、特定用途适用性以及非侵权的暗示担保。
2. 在任何情况下，DataDirect 或其第三方供应商均不对最终用户客户承担因使用 ODBC 驱动程序而引起的任何直接、间接、偶然、特殊、继发或其他损害赔偿的责任，无论是否已提前告知该种损害的可能性。这些限制适用于所有诉因，包括但不限于违反合同、违反担保、过失、严格责任、虚假陈述以及其他侵权行为。

本软件和文档包含 Informatica Corporation 的专有信息，基于一份包含使用及披露限制的许可协议而提供，并受版权法保护。禁止对本软件进行反向工程处理。未事先征得 Informatica Corporation 同意，不得以任何形式、通过任何手段（电子、影印、录制或其他手段）复制或传送本文档的任何部分。本软件受美国和/或国际专利以及其他正在申请的专利的保护。

美国政府使用、复制或披露本软件受适用的软件许可协议中规定的限制以及适用的 DFARS 227.7202-1(a) 和 227.7702-3(a) (1995)、DFARS 252.227-7013(C)(1)(ii) (OCT 1988)、FAR 12.212(a) (1995)、FAR 52.227-19 或 FAR 52.227-14 (ALT III) 中的限制之制约。

本产品或文档中的信息如有更改，恕不另行通知。如发现本产品或文档中有什么问题，请以书面形式向我们报告。

Informatica、Informatica Platform、Informatica Data Services、PowerCenter、PowerCenterRT、PowerCenter Connect、PowerCenter Data Analyzer、PowerExchange、PowerMart、Metadata Manager、Informatica Data Quality、Informatica Data Explorer、Informatica B2B Data Transformation、Informatica B2B Data Exchange Informatica On Demand、Informatica Identity Resolution、Informatica Application Information Lifecycle Management、Informatica Complex Event Processing、Ultra Messaging 和 Informatica Master Data Management 是 Informatica Corporation 在美国和全球其他辖区的商标或注册商标。所有其他公司和产品名称可能是其各自所有者的商业名称或商标。

本软件和/或文档的某些部分受第三方版权制约，包括但不限于：版权所有 DataDirect Technologies。保留所有权利。版权所有 (C) Sun Microsystems。保留所有权利。版权所有 (C) RSA Security Inc. 保留所有权利。版权所有 (C) Ordinal Technology Corp. 保留所有权利。版权所有 (C) Aandacht c.v. 保留所有权利。版权所有 Genivia, Inc. 保留所有权利。版权所有 Isomorphic Software。保留所有权利。版权所有 (C) Meta Integration Technology, Inc. 保留所有权利。版权所有 (C) Intalio。保留所有权利。版权所有 (C) Oracle。保留所有权利。版权所有 (C) Adobe Systems Incorporated。保留所有权利。版权所有 (C) DataArt, Inc. 保留所有权利。版权所有 (C) ComponentSource。保留所有权利。版权所有 (C) Microsoft Corporation。保留所有权利。版权所有 (C) Rogue Wave Software, Inc. 保留所有权利。版权所有 (C) Teradata Corporation。保留所有权利。版权所有 (C) Yahoo! Inc. 保留所有权利。版权所有 (C) Glyph & Cog, LLC。保留所有权利。版权所有 (C) Thinkmap, Inc. 保留所有权利。版权所有 (C) Clearpace Software Limited。保留所有权利。版权所有 (C) Information Builders, Inc. 保留所有权利。版权所有 (C) OSS Nokalva, Inc. 保留所有权利。版权所有 Edifecs, Inc. 保留所有权利。版权所有 Cleo Communications, Inc. 保留所有权利。版权所有 (C) International Organization for Standardization 1986。保留所有权利。版权所有 (C) ej-technologies GmbH。保留所有权利。版权所有 (C) Jaspersoft Corporation。保留所有权利。版权所有 (C) International Business Machines Corporation。保留所有权利。版权所有 (C) yWorks GmbH。保留所有权利。版权所有 (C) Lucent Technologies。保留所有权利。版权所有 (C) University of Toronto。保留所有权利。版权所有 (C) Daniel Veillard。保留所有权利。版权所有 (C) Unicode, Inc. 版权所有 IBM Corp. 保留所有权利。版权所有 (C) MicroQuill Software Publishing, Inc. 保留所有权利。版权所有 (C) PassMark Software Pty Ltd. 保留所有权利。版权所有 (C) LogiXML, Inc. 保留所有权利。版权所有 (C) 2003-2010 Lorenzi Davide。保留所有权利。版权所有 (C) Red Hat, Inc. 保留所有权利。版权所有 (C) The Board of Trustees of the Leland Stanford Junior University。保留所有权利。版权所有 (C) EMC Corporation。保留所有权利。版权所有 (C) Flexera Software。保留所有权利。版权所有 (C) Jinfonet Software。保留所有权利。版权所有 (C) Apple Inc. 保留所有权利。版权所有 (C) Telerik Inc. 保留所有权利。版权所有 (C) BEA Systems。保留所有权利。版权所有 (C) PDFlib GmbH。保留所有权利。版权所有 (C) Orientation in Objects GmbH。保留所有权利。版权所有 (C) Tanuki Software, Ltd. 保留所有权利。版权所有 (C) Ricebridge。保留所有权利。版权所有 (C) Sencha, Inc. 保留所有权利。版权所有 (C) Scalable Systems, Inc. 保留所有权利。

本产品包括由 Apache Software Foundation (<http://www.apache.org/>) 开发的软件 and/或在不同 Apache 许可证版本（以下简称“许可证”）下许可的其他软件。您可从 <http://www.apache.org/licenses/> 获取这些许可证的副本。除非适用法律要求或者有相应书面协议，否则依据这些“许可证”分发的软件以“原样”提供，不附带任何明示或暗示的担保或条件。请参阅“许可证”中规定的具体语言管理权限和限制。

本产品包括由 Mozilla (<http://www.mozilla.org/>) 开发的软件、由 JBoss Group, LLC 开发的软件（版权所有 JBoss Group, LLC 保留所有权利）、由 Bruno Lowagie 和 Paulo Soares 开发的软件（版权所有 (C) 1999-2006 Bruno Lowagie 和 Paulo Soares）以及在 <http://www.gnu.org/licenses/lgpl.html> 网站上的不同版本 GNU Lesser General 公共许可协议下许可的软件。这些材料由 Informatica 按“原样”免费提供，不附带任何明示或暗示的担保，包括但不限于适销性和特定用途适用性的暗示担保。

本产品包括 ACE(TM) 和 TAO(TM) 软件，这些软件版权归 Douglas C. Schmidt 及其在华盛顿大学、加利福尼亚大学欧芬分校以及范德堡大学的研发团队所有（版权所有 ((C)) 1993-2006，保留所有权利）。

本产品包括由 OpenSSL Project 开发并在 OpenSSL Toolkit（版权所有 OpenSSL Project。保留所有权利）中使用的软件，该软件的再分发受 <http://www.openssl.org> 和 <http://www.openssl.org/source/license.html> 上规定条款之制约。

本产品包括 Curl 软件，版权所有 1996-2013，Daniel Stenberg <[daniel@haxx.se](mailto:daniel@haxx.se)>。保留所有权利。有关该软件的权限和限制受 <http://curl.haxx.se/docs/copyright.html> 上规定条款之制约。允许出于任何目的以免费或收费形式使用、复制、修改和分发该软件，但前提是所有副本均应注明上述版权声明以及本许可声明。

本产品包括由 MetaStuff, Ltd. 开发的软件，版权所有 2001-2005 ((C)) MetaStuff, Ltd. 保留所有权利。有关该软件的权限和限制受 <http://www.dom4j.org/license.html> 上规定条款之制约。

本产品包括由 Dojo Foundation 开发的软件，版权所有 (C) 2004-2007, Dojo Foundation。保留所有权利。有关该软件的权限和限制受 <http://dojotoolkit.org/license> 上规定条款之制约。

本产品包括 ICU 软件，版权所有 International Business Machines Corporation 和其他方。保留所有权利。有关该软件的权限和限制受 <http://source.icu-project.org/repos/icu/icu/trunk/license.html> 上规定条款之制约。

本产品包括由 Per Bothner 开发的软件，版权所有 (C) 1996-2006 Per Bothner。保留所有权利。<http://www.gnu.org/software/kawa/Software-License.html> 上的许可证中规定了您使用这些材料的权利。

本产品包括 OSSP UUID 软件，版权所有 (C) 2002 Ralf S. Engelschall，版权所有 (C) 2002 OSSP Project，版权所有 (C) 2002 Cable & Wireless Deutschland。有关该软件的权限和限制受 <http://www.opensource.org/licenses/mit-license.php> 上规定条款之制约。

本产品包括由 Boost (<http://www.boost.org/>) 开发的软件或在 Boost 软件许可证下许可的软件。有关该软件的权限和限制受 [http://www.boost.org/LICENSE\\_1\\_0.txt](http://www.boost.org/LICENSE_1_0.txt) 上规定条款之制约。

本产品包括由 University of Cambridge 开发的软件，版权所有 (C) 1997-2007 University of Cambridge。有关该软件的权限和限制受 <http://www.pcre.org/license.txt> 上规定条款之制约。

本产品包括由 The Eclipse Foundation 开发的软件，版权所有 (C) 2007 The Eclipse Foundation。保留所有权利。有关该软件的权限和限制受 <http://www.eclipse.org/org/documents/epl-v10.php> 和 <http://www.eclipse.org/org/documents/edl-v10.php> 上规定条款之制约。

本产品包括在 <http://www.tcl.tk/software/tcltk/license.html>、<http://www.bosrup.com/web/overlib/?License>、<http://www.stlport.org/doc/license.html>、<http://asm.ow2.org/license.html>、<http://www.cryptix.org/LICENSE.TXT>、<http://hsqldb.org/web/hsqllicense.html>、<http://httpunit.sourceforge.net/doc/license.html>、<http://jung.sourceforge.net/license.txt>、[http://www.gzip.org/zlib/zlib\\_license.html](http://www.gzip.org/zlib/zlib_license.html)、<http://www.openldap.org/software/release/license.html>、<http://www.libssh2.org>、<http://slf4j.org/license.html>、<http://www.sente.ch/software/OpenSourceLicense.html>、<http://fusesource.com/downloads/license-agreements/fuse-message-broker-v-5-3-license-agreement>、<http://antlr.org/license.html>、<http://aopalliance.sourceforge.net/>、<http://www.bouncycastle.org/license.html>、<http://www.jgraph.com/jgraphdownload.html>、<http://www.jcraft.com/jsch/LICENSE.txt>、[http://jotm.objectweb.org/bsd\\_license.html](http://jotm.objectweb.org/bsd_license.html)、<http://www.w3.org/Consortium/Legal/2002/copyright-software-20021231>、<http://www.slf4j.org/license.html>、<http://nanoxml.sourceforge.net/orig/copyright.html>、<http://www.json.org/license.html>、<http://forge.ow2.org/projects/javaservice/>、<http://www.postgresql.org/about/licence.html>、<http://www.sqlite.org/copyright.html>、<http://www.tcl.tk/software/tcltk/license.html>、<http://www.jaxen.org/faq.html>、<http://www.jdom.org/docs/faq.html>、<http://www.slf4j.org/license.html>、<http://www.iodbc.org/dataspace/iodbc/wiki/IODBC/License>、<http://www.keplerproject.org/md5/license.html>、<http://www.toedter.com/en/jcalendar/license.html>、<http://www.edankert.com/bounce/index.html>、<http://www.net-snmp.org/about/license.html>、<http://www.openmdx.org/#FAQ>、[http://www.php.net/license/3\\_01.txt](http://www.php.net/license/3_01.txt)、<http://srp.stanford.edu/license.txt>、<http://www.schneier.com/blowfish.html>、<http://www.jmock.org/license.html>、<http://xsom.java.net>、<http://benalman.com/about/license/>、<https://github.com/CreateJS/EaselJS/blob/master/src/easeljs/display/Bitmap.js>、<http://www.h2database.com/html/license.html#summary>、<http://jsoncpp.sourceforge.net/LICENSE>、<http://jdbc.postgresql.org/license.html>、<http://protobuf.googlecode.com/svn/trunk/src/google/protobuf/descriptor.proto>、<https://github.com/rantav/hector/blob/master/LICENSE>、<http://web.mit.edu/Kerberos/krb5-current/doc/mitK5license.html>、<http://jibx.sourceforge.net/jibx-license.html>、<https://github.com/lyokato/libgeohash/blob/master/LICENSE>；and <https://github.com/hjiang/jsonxx/blob/master/LICENSE>。

本产品包括在 Academic 免费许可证 (<http://www.opensource.org/licenses/afl-3.0.php>)、通用开发和分发许可证 (<http://www.opensource.org/licenses/cddl1.php>)、通用公共许可证 (<http://www.opensource.org/licenses/cpl1.0.php>)、Sun Binary Code 许可协议补充许可条款、BSD 许可证 (<http://www.opensource.org/licenses/bsd-license.php>)、新 BSD 许可证 (<http://opensource.org/licenses/BSD-3-Clause>)、MIT 许可证 (<http://www.opensource.org/licenses/mit-license.php>)、Artistic 许可证 (<http://www.opensource.org/licenses/artistic-license-1.0>) 以及原始开发者公共许可证版本 1.0 (<http://www.firebirdsql.org/en/initial-developer-s-public-license-version-1-0/>) 下许可的软件。

本产品包括由 Joe Walnes 和 XStream Committers 开发的软件，版权所有 (C) 2003-2006 Joe Walnes，2006-2007 XStream Committers。保留所有权利。有关该软件的权限和限制受 <http://xstream.codehaus.org/license.html> 上规定条款之制约。本产品包括由 Indiana University Extreme! Lab 开发的软件。有关详细信息，请访问 <http://www.extreme.indiana.edu/>。

本产品包括软件版权所有 (c) 2013 Frank Balluffi 和 Markus Moeller。保留所有权利。有关此软件的权限和限制受 MIT 许可证上规定条款之制约。

本软件受美国专利号 5,794,246、6,014,670、6,016,501、6,029,178、6,032,158、6,035,307、6,044,374、6,092,086、6,208,990、6,339,775、6,640,226、6,789,096、6,823,373、6,850,947、6,895,471、7,117,215、7,162,643、7,243,110、7,254,590、7,281,001、7,421,458、7,496,588、7,523,121、7,584,422、7,676,516、7,720,842、7,721,270、7,774,791、8,065,266、8,150,803、8,166,048、8,166,071、8,200,622、8,224,873、8,271,477、8,327,419、8,386,435、8,392,460、8,453,159、8,458,230 和 RE44,478、国际专利及其他正在申请的专利保护。

免责声明：Informatica Corporation 以“原样”提供本文档，不附带任何明示或暗示的担保，包括但不限于非侵权、适销性或特定用途适用性的暗示担保。Informatica Corporation 不保证本软件或文档中没有错误。本软件或文档中提供的信息可能包括技术上的不准确性或排字错误。本软件和文档中包含的信息随时可能更改，恕不另行通知。

## 声明

本 Informatica 产品（以下称“软件”）包括由 Progress Software Corporation 的运营公司 DataDirect Technologies（以下称“DataDirect”）提供的某些驱动程序（以下称“DataDirect 驱动程序”），受以下条款和条件制约：

1. DataDirect 驱动程序以“原样”提供，不附带任何明示或暗示的担保，包括但不限于适销性、特定用途适用性以及非侵权的暗示担保。
2. 在任何情况下，DataDirect 或其第三方供应商均不对最终用户客户承担因使用 ODBC 驱动程序而引起的任何直接、间接、偶发、特殊、继发或其他损害赔偿的责任，无论是否已提前告知该种损害的可能性。这些限制适用于所有诉讼，包括但不限于违反合同、违反担保、过失、严格责任、虚假陈述以及其他侵权行为。

发布日期: 2018-05-16

# 目录

前言 .....	8
Informatica 资源 .....	8
Informatica 我的支持门户 .....	8
Informatica 文档 .....	8
Informatica 网站 .....	8
Informatica 入门知识库 .....	8
Informatica 知识库 .....	9
Informatica 支持 YouTube 频道 .....	9
Informatica Marketplace .....	9
Informatica Velocity .....	9
Informatica 全球客户支持部门 .....	9
第 1 章： 引用数据简介 .....	10
引用数据概览 .....	10
Informatica 引用数据 .....	11
用户定义的引用数据 .....	11
引用表 .....	12
引用表结构 .....	12
受管和非受管引用表 .....	12
第 2 章： Analyst 工具中的引用表 .....	13
Analyst 工具引用表概览 .....	13
引用表属性 .....	13
引用表列属性 .....	14
在引用表编辑器中创建引用表 .....	14
从配置文件数据创建引用表 .....	15
从配置文件列创建引用表 .....	15
从列值创建引用表 .....	16
从列模式创建引用表 .....	16
从平面文件创建引用表 .....	17
Analyst 工具平面文件属性 .....	17
从平面文件创建引用表 .....	18
从数据库表创建引用表 .....	18
从数据库表创建引用表 .....	19
引用表更新 .....	19
管理列 .....	20
管理行 .....	20
查找和替换值 .....	21
导出引用表 .....	21
启用和禁用对非受管引用表的编辑 .....	22

刷新引用表值. . . . .	22
审计跟踪事件. . . . .	22
查看审计跟踪事件. . . . .	23
引用表的规则和准则. . . . .	23
<b>第 3 章：Developer 工具中的引用数据. . . . .</b>	<b>25</b>
Developer 工具引用表概览. . . . .	25
引用数据和转换. . . . .	25
引用表. . . . .	26
引用表数据属性. . . . .	26
创建引用表对象. . . . .	27
从平面文件创建引用表. . . . .	28
从关系源创建引用表. . . . .	29
内容集. . . . .	29
字符集. . . . .	30
分类器模型. . . . .	30
模式集. . . . .	31
概率模型. . . . .	31
正则表达式. . . . .	31
标志集. . . . .	32
创建内容集. . . . .	33
在内容集中创建引用数据对象. . . . .	33
从中游配置文件生成引用数据. . . . .	34
<b>第 4 章：分类器模型. . . . .</b>	<b>35</b>
分类器模型概览. . . . .	35
分类器转换示例. . . . .	35
分类器模型结构. . . . .	36
分类器模型视图. . . . .	36
分类器模型引用数据. . . . .	37
将数据和标签值添加到分类器模型. . . . .	38
从分类器模型中删除数据值. . . . .	38
分类器模型标签数据. . . . .	38
将标签添加到分类器模型. . . . .	38
为分类器模型数据分配标签. . . . .	39
从分类器模型中删除标签. . . . .	39
编译分类器模型. . . . .	39
分类器得分. . . . .	40
分类器模型筛选器. . . . .	40
在引用数据中查找值. . . . .	40
查找不具有标签的数据行. . . . .	40
在数据行中查找数据. . . . .	41
通过数据对象创建分类器模型. . . . .	41

复制和粘贴操作. . . . .	42
将分类器模型复制到其他内容集. . . . .	42
从其他内容集导入分类器模型. . . . .	42
<b>第 5 章： 概率模型. . . . .</b>	<b>43</b>
概率模型概览. . . . .	43
概率模型结构. . . . .	44
概率模型视图. . . . .	44
标签创建器转换示例. . . . .	45
解析器转换示例. . . . .	46
概率模型引用数据. . . . .	47
将引用数据字符串添加到概率模型. . . . .	47
在概率模型中查找数据值. . . . .	47
概率模型标签数据. . . . .	47
溢出标签. . . . .	48
为概率模型数据分配标签. . . . .	48
将标签添加到概率模型. . . . .	48
从概率模型中删除标签. . . . .	49
编译概率模型. . . . .	49
概率模型属性. . . . .	49
从数据对象创建概率模型. . . . .	50
复制和粘贴操作. . . . .	50
将概率模型复制到其他内容集. . . . .	51
从其他内容集导入概率模型. . . . .	51
<b>索引. . . . .</b>	<b>52</b>

# 前言

《引用数据指南》包含有关 Informatica Data Quality 中引用数据对象的信息。该指南是专为在组织内协作完成项目的业务用户（如分析师和数据管理者）编写的。该指南介绍 Developer 工具、Analyst 工具和模型存储库中的引用数据对象。

## Informatica 资源

### Informatica 我的支持门户

Informatica 客户可以访问 Informatica 我的支持门户，网址为 <http://mysupport.informatica.com>。

该站点包含产品信息、用户组信息、业务通讯、Informatica 入门知识库、Informatica 知识库、Informatica 产品文档，并能够访问 Informatica 客户支持案例管理系统 (ATLAS) 和 Informatica 用户社区。

### Informatica 文档

Informatica 文档团队致力于创建各种准确、实用的文档。如果您对此文档有任何疑问、意见或建议，请与 Informatica 文档团队联系，电子邮件地址为 [infa\\_documentation@informatica.com](mailto:infa_documentation@informatica.com)。我们将根据您的反馈改进我们的文档。请告知我们是否可以就您的意见与您进行联系。

文档团队将根据需要更新文档。要获取适用于您的产品的最新文档，请导航至 <http://mysupport.informatica.com> 上的“Product Documentation”（产品文档）部分。

### Informatica 网站

您可以访问 Informatica 公司网站，网址为 <http://www.informatica.com>。该网址包含有关 Informatica 公司、企业背景、近期活动和营业点的相关信息。您还可以查看产品和合作伙伴信息。该网站的业务范围包括有关技术支持、培训和教育以及实施服务的重要信息。

### Informatica 入门知识库

作为 Informatica 的客户，您可以访问 Informatica 入门知识库，网址为 <http://mysupport.informatica.com>。入门知识库汇集了大量的资源，可以帮助您了解更多有关 Informatica 产品和功能的信息。其中包括各种文章和交互演示，这些文章和交互演示用于为常见问题提供解决方案，比较各种功能和行为，并指导您执行实际的任务。



## Informatica 知识库

作为 Informatica 的客户，您可以访问 Informatica 知识库，网址为 <http://mysupport.informatica.com>。通过该知识库可以搜索为 Informatica 产品的已知技术问题提供的解决方案。此外，还可以查看对常见问题的解答、技术白皮书以及技术知识。如果您对知识库有任何疑问、意见或建议，请与 Informatica 知识库团队联系，电子邮件地址为 [KB\\_Feedback@informatica.com](mailto:KB_Feedback@informatica.com)。

## Informatica 支持 YouTube 频道

您可以访问 Informatica 支持 YouTube 频道，网址为 <http://www.youtube.com/user/INFASupport>。Informatica 支持 YouTube 频道包含可指导您执行特定任务的相关解决方案的视频。如果您对 Informatica 支持 YouTube 频道有任何疑问、意见或建议，请通过电子邮件 [supportvideos@informatica.com](mailto:supportvideos@informatica.com) 与支持 YouTube 团队联系，或者通过推特联系 @INFASupport。

## Informatica Marketplace

Informatica Marketplace 是一个论坛，开发人员和合作伙伴可以在其中分享能够提高、扩展或增强数据集成实现的解决方案。您可以利用 Marketplace 提供的数以百计的解决方案中的任何方案，提高生产率，加快项目的实现时间。您可以访问 Informatica Marketplace，网址为 <http://www.informaticamarketplace.com>。

## Informatica Velocity

您可以访问 Informatica Velocity，网址为 <http://mysupport.informatica.com>。Informatica Velocity 基于数以百计的数据管理项目的实际经验而开发，其相当于我们曾在世界各地组织就职的顾问在规划、开发、部署和维护成功的数据管理解决方案方面的知识的集合。如果您对 Informatica Velocity 有任何疑问、意见或建议，请通过与 Informatica Professional Services 联系 [ips@informatica.com](mailto:ips@informatica.com)。

## Informatica 全球客户支持部门

您可以通过电话或联机支持与客户支持中心联系。

联机支持需要用户名和密码。您可以从 <http://mysupport.informatica.com> 请求用户名和密码。

可从 Informatica 网站中获取 Informatica 全球客户支持部门的电话号码，网址为 <http://www.informatica.com/us/services-and-training/support-services/global-support-centers/>。

# 第 1 章

## 引用数据简介

本章包括以下主题：

- [引用数据概览, 10](#)
- [Informatica 引用数据, 11](#)
- [用户定义的引用数据, 11](#)
- [引用表, 12](#)

## 引用数据概览

引用数据对象标识一组您在为数据质量操作配置转换时可以选择的数据值。可以在 Developer 工具和 Analyst 工具中创建引用数据对象。还可以将引用数据对象导入模型存储库和文件系统。可以使用 Data Quality Content 安装程序导入引用数据对象以及将引用数据安装到文件系统中。

可以创建和编辑以下类型的引用数据：

### 引用表

引用表包含一组数据值的标准版本和替代版本。将引用表添加到 Developer 工具中的某个转换，以确认源数据值是准确的且已正确设置格式。

数据库表至少包含两个列。一列包含标准或首选版本的字符串，其他列包含替代版本。将引用表添加到转换时，该转换将在输入端口数据中搜索同时显示在引用表中的值。可以使用任何对所处理的数据项目有用的数据来创建引用表。

### 内容集

内容集是在模型存储库或文件中指定引用数据值的模型存储库对象。将内容集添加到某个转换中时，该转换在输入数据中搜索与内容集中的数据模式匹配的值。

Data Quality Content 安装程序可以安装以下类型的引用数据：

### Informatica 引用表

Informatica 开发的存储库对象和数据文件。将加速器对象导入到模型存储库中时，将导入 Informatica 引用表。引用信息类型包括电话区号、邮政编码格式、名、社会保障号格式、职业和首字母缩略词。可以编辑 Informatica 引用表。

### Informatica 内容集

Informatica 开发的存储库对象和数据文件。将加速器对象导入到模型存储库中时，将导入内容集。内容集包含不同类型的引用数据，您可以将其用于对数据质量转换执行搜索操作。

### 地址引用数据文件

包含某个国家/地区的可投递地址数据的引用数据文件。地址验证器转换将读取该引用数据。无法创建或编辑地址引用数据文件。

地址引用数据对于某个定义的时间段是最新的，因此您必须定期（例如每个季度一次）刷新数据。无法查看或编辑地址引用数据。

### 标识填充文件

包含有关个人、家庭和公司标识的信息的引用数据文件。匹配转换和比较转换使用填充文件在输入数据中查找潜在标识。无法创建或编辑地址标识填充文件。

## Informatica 引用数据

可以从 Informatica 购买并下载地址引用数据和标识填充数据。

可以购买某个国家/地区的地址数据年度订阅，可以在订阅期间随时从 Informatica 下载最新的地址数据。

内容安装程序用户从应用程序单独下载和安装引用数据。有关您的系统上安装的引用数据，请联系用户管理员。

## 用户定义的引用数据

可以使用数据对象中的值创建引用数据对象。

例如，可以选择包含特定于某个项目或组织的值的数据对象或配置文件列。从列值创建自定义引用数据对象。

可以从数据列构建引用数据对象以确认以下内容：

- 列中的数据行包含同一类型的信息。
- 源值有效。引用对象可能包含一列有效值，引用对象也可能包含一列无效值。

下表列出了可以包含引用数据的项目数据列的常见示例：

信息	引用数据示例
库存单位 (SKU) 代码	使用 SKU 列创建组织的有效 SKU 代码的引用表。使用引用表在数据集中查找正确或错误的 SKU 代码。
员工代码	使用员工代码或员工 ID 列创建有效员工代码的引用表。使用引用表在员工数据中查找错误。
客户帐号	对客户帐户列运行配置文件，以标识帐号模式。使用配置文件创建错误数据模式的标志集。使用该标志集查找不符合正确的帐号结构的帐号。
客户名称	当客户名称列包含名、中间名和姓时，可以创建用于定义列中字符串所需的结构的概率模型。使用概率模型查找不属于该列的数据字符串。

# 引用表

将引用表添加到 Developer 工具中的某个转换。配置转换以在输入数据中查找引用表值并写入引用表中的对应有效值作为输出。

要在 Developer 工具中创建引用表，请使用以下方法之一：

- 创建空引用表并输入数据值。
- 从平面文件中的数据创建引用表。
- 从其他数据库表中的数据创建引用表。

## 引用表结构

大多数引用表至少包含两个列。一个列包含数据值的正确或所需的版本。其他列包含这些值的其他版本，包括可能显示在源数据中的替代版本。

包含正确或所需值的列称为有效列。当某个转换读取映射中的引用表时，该转换在无效列中查找值。当该转换找到无效值时，将返回有效列中的对应值。还可以将转换配置为返回单个通用值，而不是有效值。

有效列可以包含形式正确的数据，如邮政编码。有效列可以包含与项目相关的数据，如某个组织特有的库存单位 (SKU) 编号。还可以从错误数据（如包含您要搜索的已知数据错误的值）创建有效列。

例如，创建包含一系列某零售组织中有效 SKU 编号的引用表。将引用表添加到某个标签创建器转换并使用该转换创建映射。使用产品数据库表运行映射。当该映射运行时，标签创建器创建一个用于标识不包含有效 SKU 编号的产品记录。

### 引用表和解析器转换

当需要在基于模式的解析中使用表数据时，使用单个列创建引用表。将解析器转换配置为执行基于模式的解析，并将数据导入转换配置。

## 受管和非受管引用表

引用表在模型存储库中存储元数据。引用表可以在引用数据仓库或其他数据库存储列数据。

当引用数据仓库存储列数据时，Informatica 服务将该表标识为受管引用表。当其他数据库存储列数据时，Informatica 服务将该表标识为非受管引用表。内容管理服务为引用数据仓库使用数据库连接。

要维护引用表，您必须在内容管理服务中拥有以下特权：

- 创建引用表
- 编辑引用表数据
- 编辑引用表元数据

在 Administrator 工具中使用“安全”选项查看或更新服务特权。要在非受管引用表中编辑数据，还需要确认您已将引用表对象配置为允许编辑。

**注意:** 如果您编辑数据库应用程序中非受管引用表的元数据，请使用 Analyst 工具将模型存储库与数据库表同步。在 Developer 工具中使用非受管引用表之前，必须同步模型存储库和数据库表。

## 第 2 章

# Analyst 工具中的引用表

本章包括以下主题：

- [Analyst 工具引用表概览, 13](#)
- [引用表属性, 13](#)
- [在引用表编辑器中创建引用表, 14](#)
- [从配置文件数据创建引用表, 15](#)
- [从平面文件创建引用表, 17](#)
- [从数据库表创建引用表, 18](#)
- [引用表更新, 19](#)
- [审计跟踪事件, 22](#)
- [引用表的规则和准则, 23](#)

## Analyst 工具引用表概览

在 Analyst 工具的设计工作区中创建引用表。

可以从平面文件、模型存储库中的数据源和其他数据库中的表创建引用表。

可以从配置文件列或配置文件列中数据的子集创建引用表。还可以从您从配置文件中选择的列模式创建引用表。

创建或更新引用表时，将配置该表及其包含的数据列上的属性。

## 引用表属性

可以在 Analyst 工具中查看和编辑引用表属性。引用表显示用于描述存储库对象的常规属性和用于描述列数据的列属性。

要查看这些属性，请打开引用表并选择**属性**面板。

要编辑这些属性，请打开引用表并单击**操作 > 编辑表**。

# 引用表列属性

列属性包含有关列元数据的信息。

下表介绍了列属性：

属性	说明
名称	每个列的名称。
数据类型	每个列中数据的数据类型。可以选择以下数据类型之一： <ul style="list-style-type: none"><li>- 长整型</li><li>- 日期/时间型</li><li>- 小数</li><li>- 双精度型</li><li>- 整型</li><li>- 字符串型</li></ul> 创建空引用表或从平面文件创建引用表时，无法选择双精度数据类型。
精度	每个列的精度。精度是列可以容纳的最大位数或最大字符数。 您配置的精度值取决于数据类型。
小数位数	每个列的小数位数。小数位数是列可以在小数点右侧容纳的最大位数。适用于小数列。 您配置的小数位数取决于数据类型。
说明	每个列的可选说明。
可空	指示列是否可以包含空值。

# 在引用表编辑器中创建引用表

在引用表编辑器中定义表结构并将数据添加到引用表中。

1. 单击**设计 > 应用表**。
2. 选择**使用引用表编辑器**选项。
3. 单击**下一步**。
4. 单击**添加新列**。  
Analyst 工具将添加用于输入列属性的行。
5. 对于要包含在引用表中的每个列，单击**添加新列**图标并配置每个列的属性。  
**注意：**可以将列重新排序或删除列。
6. （可选）输入表的审计说明。  
审计说明显示在审计跟踪日志中。
7. 输入引用表的名称和位置。
8. 单击**完成**。

# 从配置文件数据创建引用表

可以使用配置文件数据创建与配置文件中的源数据相关的引用表。使用引用表查找源数据中不同类型的信息。

可以使用配置文件通过以下方法创建或更新引用表：

- 在配置文件中选择一个列并将其添加到引用表。
- 浏览某个配置文件列并将该列的一个子集添加到引用表。
- 在该配置文件中选择一个列并将该列的模式值添加到引用表。

相关主题：

- [“从中游配置文件生成引用数据” 页面上 34](#)

## 从配置文件列创建引用表

可以从配置文件列创建引用表。可以将配置文件列添加到现有引用表。**新建引用表**向导将列添加到引用表。

1. 在 Analyst 工具中打开**库**工作区。
2. 选择**配置文件**资产。  
此时将显示配置文件列表。
3. 找到具有您要添加到引用表的列的配置文件。
4. 单击**操作 > 打开**。
5. 选择您要添加到引用表的列。
6. 单击**操作 > 添加到引用表**。
7. 选择**新建引用表**选项。  
(可选) 选择**添加到现有引用表**。
8. 单击**下一步**。  
列名称默认显示为表名称。
9. (可选) 输入其他表名称、说明和默认值。  
Analyst 工具为任何不包含值的表记录使用该默认值。
10. 单击**下一步**。
11. 在**列属性**面板中，配置列的列属性。
12. (可选) 选择为引用表中的行创建说明列。  
输入该列的名称和精度。
13. 在**预览**面板中预览列值。
14. 单击**下一步**。  
列名称默认显示为表名称。(可选) 输入其他表名称和说明。
15. 在**保存位置**面板中，选择要创建引用表的位置。  
**引用表**面板将在您选择的位置列出引用表。
16. (可选) 输入审计说明。
17. 单击**完成**。

## 从列值创建引用表

可以从配置文件列中的列值创建引用表。在配置文件中选择一个列并选择要添加到引用表的列值或创建一个引用表以添加列值。

1. 在 Analyst 工具中打开**库**工作区。
2. 选择**配置文件**资产。  
此时将显示配置文件列表。
3. 选择某个配置文件以将其打开并单击**操作 > 打开**。
4. 选择您要添加到引用表的列。
5. 在**值**面板中，选择要添加的列值。使用 Ctrl 或 Shift 键选择多个值。
6. 单击**操作 > 添加到引用表**。  
此时将显示**新建引用表**向导。
7. 选择**新建引用表**选项。  
(可选) 选择**添加到现有引用表**，然后单击**下一步**。在项目或文件夹中导航到该引用表，预览该引用表数据，然后单击**下一步**。选择要添加的列并单击**完成**。
8. 单击**下一步**。
9. 列名称默认显示为表名称。(可选) 输入其他表名称、说明和默认值。  
Analyst 工具为任何不包含值的表记录使用该默认值。
10. 单击**下一步**。
11. 在**列属性**面板中，配置列的列属性。
12. (可选) 选择为引用表中的行创建说明列。  
输入该列的名称和精度。
13. 在**预览**面板中预览列值。
14. 单击**下一步**。
15. 列名称默认显示为表名称。(可选) 输入其他表名称和说明。
16. 在**保存位置**面板中，选择要创建引用表的位置。  
**引用表**: 面板将列出您选择的位置中的引用表。
17. (可选) 输入审计说明。
18. 单击**完成**。

## 从列模式创建引用表

可以从配置文件列中的列模式创建引用表。在配置文件中选择一个列并选择要添加到引用表的模式值或创建一个引用表以添加模式值。

1. 在 Analyst 工具中打开**库**工作区。
2. 选择**配置文件**资产。  
此时将显示配置文件列表。
3. 选择您要添加到引用表的列。
4. 在**模式**面板中，选择要添加的列模式。使用 Ctrl 或 Shift 键选择多个值。
5. 单击**操作 > 添加到引用表**。  
此时将显示**新建引用表**向导。



- 6. 选择**新建引用表**选项。  
(可选) 选择**添加到现有引用表**，然后单击**下一步**。在项目或文件夹中导航到该引用表，预览该引用表数据，然后单击**下一步**。选择要添加的列并单击**完成**。
- 7. 单击**下一步**。
- 8. 列名称默认显示为表名称。(可选) 输入其他表名称、说明和默认值。  
Analyst 工具为任何不包含值的表记录使用该默认值。
- 9. 单击**下一步**。
- 10. 在**列属性**面板中，配置列的列属性。
- 11. (可选) 选择为引用表中的行创建说明列。  
输入该列的名称和精度。
- 12. 在**预览**面板中预览列值。
- 13. 单击**下一步**。
- 14. 列名称默认显示为表名称。(可选) 输入其他表名称和说明。
- 15. 在**保存位置**面板中，选择要创建引用表的位置。  
**引用表:** 面板将列出您选择的位置中的引用表。
- 16. (可选) 输入审计说明。
- 17. 单击**完成**。

## 从平面文件创建引用表

可以从 CSV 文件导入引用数据。使用**新建引用表**向导导入文件数据。  
必须为您用于创建引用表的每个平面文件配置属性。

### Analyst 工具平面文件属性

将某个平面文件作为引用表导入时，必须为该文件中的每个列配置属性。您配置的选项决定 Analyst 工具从该文件读取数据的方式。  
下表介绍了当您为引用表导入文件数据时可以配置的属性：

属性	说明
分隔符	用于分隔数据列的字符。使用“其他”字段输入其他分隔符。 分隔符必须是可打印字符，并且必须不同于转义符和引号字符（如果已选择）。 不能选择非打印多字节字符作为分隔符。
文本限定符	定义文本字符串边界的引号字符。 选择“无引号”、“单引号”或“双引号”。 如果选择引号字符，向导将忽略引号对内的分隔符。

属性	说明
列名称	从第一行导入列名称。如果列名称显示在第一行中，请选择该选项。 向导在预览中使用第一行中的数据作为列名称。 默认情况下不启用。
值	选择从某一行开始进行值导入。当向导导入文件时，指示在预览中向导开始读取的行号。

## 从平面文件创建引用表

从平面文件创建引用表数据时，该表使用文件的列结构并导入文件数据。

1. 在 Analyst 工具中单击 **设计 > 引用表**。  
此时将显示 **新建引用表** 向导。
2. 选择 **导入平面文件** 选项。
3. 单击 **下一步**。
4. 单击 **浏览** 选择平面文件。
5. 单击 **上传** 将该文件上载到 Analyst 工具可以访问的 Informatica 服务安装目录中的某个目录。
6. 输入表名称。（可选）输入说明和默认值。  
Analyst 工具为任何不包含值的表记录使用该默认值。
7. 选择与平面文件中的数据相匹配的代码页。
8. 在 **文件的预览** 面板中预览数据。
9. 单击 **下一步**。
10. 配置平面文件属性。
11. 在 **预览** 窗格中，单击 **显示更新预览**。
12. 单击 **下一步**。
13. 在 **列属性** 面板中，确认或编辑每个列的列属性。
14. （可选）为引用表中的行创建说明列。输入该列的名称和精度。
15. （可选）输入表的审计说明。
16. 单击 **完成**。

## 从数据库表创建引用表

从某个数据库表创建引用表时，应连接到该数据库并导入表数据。

在从某个数据库表导入引用表数据之前，请确认 Informatica 域包含到该数据库的连接。如果域不包含到该数据库的连接，则可以在 Analyst 工具中定义一个连接。

要定义数据库连接，请单击 **管理 > 连接**。

## 从数据库表创建引用表

要创建引用表，请连接到数据库并导入数据。

1. 在 Analyst 工具中单击**设计 > 引用表**。  
此时将显示**新建引用表**向导。
2. 选择**连接到关系表**选项。
3. 选择**非受管表**在引用数据仓库中创建不存储数据的表。  
要对非受管引用表执行编辑操作，请选择**可编辑**选项。  
单击**下一步**。
4. 从已建立的连接列表中选择数据库连接。  
单击**下一步**。
5. 在**表**面板中，选择一个表。
6. 确认**属性**面板中的表属性。  
(可选) 单击**数据预览**查看表数据。  
单击**下一步**。
7. 在**列属性**面板中，选择有效的列。  
(可选) 指定当用户更新引用表时要写入审计跟踪的审计注释。  
如果您创建受管引用表，则可以在**列属性**面板上执行以下操作：
  - 编辑引用表列名称。
  - 添加用于行级说明的列。
8. 单击**下一步**。
9. 输入引用表的名称和说明 (可选)。
10. 在**文件夹**面板中，指示要存储引用表的项目和文件夹。  
**引用表**面板将列出您选择的文件夹中的引用表。
11. 单击**完成**。

## 引用表更新

引用表包含的业务数据可能随时间而变化。在引用表中查看并更新数据和元数据，以确认该表包含准确的信息。在 Analyst 工具中更新引用表。可以在受管引用表和非受管引用表中更新数据和元数据。

可以对引用表数据和元数据执行以下操作：

### 管理列

可以添加列，删除列和编辑列属性。

### 管理行

可以将数据行添加到引用表。

### 编辑引用数据值

可以编辑引用数据值。

### 替换数据值

使用**查找和替换**选项替换不再准确或与组织相关的数据值。可以查找列中的某个值并将其替换为其他值。可以使用单个值替换列中的所有值。

### 导出引用表

将引用表导出到逗号分隔的值 (CSV) 文件、字典文件或 Excel 文件。

### 启用或禁用对非受管表的编辑

更新非受管引用表，以便启用或禁用对表数据和元数据的编辑。

### 刷新引用表数据

将引用表数据重新加载到 Analyst 工具中以查看对数据的最新更改。

## 管理列

可以更改引用表中的列。还可以设置引用表属性。

1. 单击**设计 > 打开**。
2. 选择“引用表”资产。
3. 单击引用表名称以在**引用表面板**中将其打开。
4. 单击**操作 > 编辑表**。  
此时将显示**编辑列属性**对话框。使用对话框选项执行以下操作：
  - 更改表中的有效列。
  - 从表中删除列。
  - 更改列名称。
  - 更新列的描述性文本。
  - 更新引用表的可编辑状态。
  - 更新表的审计说明。对于您在**编辑列属性**窗口中执行的任何操作，审计说明显示在审计日志中。
5. 完成操作后，单击**确定**。

## 管理行

可以添加、编辑或删除引用表中的行。

1. 单击**设计 > 打开**。
2. 选择“引用表”资产。
3. 单击引用表名称以在**引用表面板**中将其打开。
4. 编辑数据行。可以通过以下方法编辑数据行：
  - 要添加行，请选择**操作 > 添加行**。  
在**添加行**窗口中，输入每个列的值。（可选）输入审计说明。  
单击**确定**应用更改。
  - 要编辑某个数据值，请在引用表中双击该值并更新该值。  
编辑数据后，使用行级选项接受或拒绝编辑。
  - 要编辑多个行，请选择要编辑的行并选择**操作 > 编辑**。  
在**编辑多行**窗口中，输入行中每个列的值。（可选）输入审计说明。

单击**确定**为每个行应用更改。

- 要删除行，请选择要删除的行并单击**操作 > 删除**。

在**删除行**窗口中，输入审计说明（可选）。

单击**确定**删除数据。

**注意:** 使用 Developer 工具编辑大型引用表中的行数据。例如，如果引用表包含 500 个以上的行，则在 Developer 工具中编辑该表。

## 查找和替换值

可以在引用表中查找和替换数据值。当表包含您必须更新的数据值的一个或多个实例时，使用查找和替换选项。

1. 单击**设计 > 打开**。
2. 选择“引用表”资产。
3. 单击引用表名称以在**引用表**面板中将其打开。
4. 单击**操作 > 查找和替换**。  
此时将显示**查找和替换**工具栏。
5. 在该工具栏中输入搜索条件：
  - 在**查找**字段中输入一个数据值。
  - 选择要搜索的列。默认情况下，该操作会搜索所有列。
  - 在**替换为**字段中输入一个数据值。
6. 在您选择的列中搜索**查找**字段中的数据值。  
使用以下选项逐个替换值或替换所有值：
  - 使用**下一个**和**上一个**选项逐个查找值。
  - 要替换某个值，请选择**替换**。
  - 使用**突出显示全部**选项显示该值的所有实例。
  - 要替换该值的所有实例，请选择**全部替换**。

## 导出引用表

将引用表导出到逗号分隔的文件、字典文件或 Microsoft Excel 文件。

1. 单击**设计 > 打开**。
2. 选择“引用表”资产。
3. 单击引用表名称以在**引用表**面板中将其打开。
4. 单击**操作 > 导出数据**。

此时将显示**将数据导出到文件**窗口。

下表介绍了**将数据导出到文件**窗口中的选项：

选项	描述
文件名	导出的数据的文件名。

选项	描述
文件格式	导出的文件的格式。可以选择以下格式： <ul style="list-style-type: none"><li>• csv.逗号分隔的文件。</li><li>• xls.Microsoft Excel 文件。</li><li>• dic.Informatica 字典文件。</li></ul>
将字段名称作为第一行导出	列名称选项。选择该选项可指示文件的第一行包含列名称。
代码页	引用数据的代码页。默认代码页为 UTF-8。

5. 单击**确定**导出文件。

## 启用和禁用对非受管引用表的编辑

可以对非受管引用表进行配置，以便启用或禁用对表中数据值和列的更新。

在更改引用表的可编辑状态之前，保存该表。

1. 单击**设计 > 打开**。
2. 选择“引用表”资产。
3. 单击引用表名称以在**引用表**面板中将其打开。
4. 单击**操作 > 编辑表**。
5. 选中或清除**可编辑**选项。

此时将显示**编辑列属性**窗口。

更改引用表的可编辑状态时，属性对话框将关闭。

## 刷新引用表值

在 Analyst 工具中，使用**刷新**选项可将引用表更新至最新值。

您可能需要刷新 Analyst 工具中显示的引用表值。例如，某人可能会在 Developer 工具中更改引用表，而您需要查看最新更改。使用**刷新**选项重新加载引用表数据。

要重新加载引用表值，请单击**操作 > 刷新**。

## 审计跟踪事件

您可以查看您对引用表进行的更改的审计跟踪。对引用表进行更改并输入审计跟踪说明时，Analyst 工具将创建审计跟踪日志事件。

审计跟踪日志事件提供有关您管理的引用表的信息。可以在“审计跟踪”选项卡中配置查询选项，以筛选您查看的日志事件。可以针对日期范围、类型、用户名和状态配置筛选器。

下表介绍了您查看审计跟踪日志事件时配置的筛选器选项：

选项	描述
日期	要搜索的日志事件的开始日期和结束日期。使用日历可选择这些日期。
类型	审计跟踪事件的类型。可以筛选并查看以下事件类型： <ul style="list-style-type: none"><li>- 数据。与引用表中的数据相关的事件。事件包括创建、编辑、删除和替换所有行。</li><li>- 元数据。与引用表元数据相关的事件。事件包括创建引用表，添加、删除和编辑列，以及更新有效列。</li></ul>
用户	编辑引用表并输入审计跟踪注释的用户。Analyst 工具从在 Administrator 工具中配置的 Analyst 工具用户生成用户列表。
状态	审计跟踪日志事件状态。状态与在引用表编辑器中执行的操作相对应。

审计跟踪日志事件还包括审计跟踪注释和插入、更新或删除的列值。

## 查看审计跟踪事件

查看审计跟踪日志事件，以获取有关对引用表进行的更改的更多信息。

1. 单击**设计 > 打开**并选择**引用表**。
2. 单击某个引用表名称以在面板中将其打开。此时将显示**引用表**选项卡。
3. 单击**审计跟踪**。
4. 配置筛选选项。
5. 单击**显示**  
此时将显示指定查询选项的日志事件。

## 引用表的规则和准则

在 Analyst 工具中使用引用表时，使用以下规则和准则：

- 从 Oracle、IBM DB2、IBM DB2/zOS、IBM DB2/iOS 或 Microsoft SQL Server 数据库导入引用表时，如果表、视图、架构、同义词或列名称包含混合大小写字符或小写字符，则 Analyst 工具无法显示预览。

如果您正在查看的引用表在其他客户端中发生更改，则使用**刷新**选项重新加载引用表并在 Analyst 工具中查看更改。

要预览位于区分大小写的数据库中表中的数据，请将“支持混合大小写标识符”属性设置为 true。在 Developer 工具或 Administrator 工具中，将 Oracle、IBM DB2、IBM DB2/zOS、IBM DB2/iOS 和 Microsoft SQL Server 数据库连接中的该属性设置为 true。

- 从使用一种格式的推理的列模式创建引用表时，Analyst 工具将通过使用另一种格式的列模式填充引用表。  
例如，为列模式 X(5) 创建引用表时，Analyst 工具将针对引用表中的列模式显示以下格式：XXXXX。
- 导入 Oracle 数据库表时，确认表中任何 VARCHAR2 列的长度。Analyst 工具无法导入包含长度大于 1000 的 VARCHAR2 列的 Oracle 数据库表。

- 要读取引用表，您需要在到存储表数据值的数据库的连接上具有执行权限。例如，如果引用数据仓库存储这些数据值，您需要在到该引用数据仓库的连接上具有执行权限。您需要执行权限才能以读取或写入模式访问引用表。数据库连接权限适用于数据库中的所有引用数据。



## 第 3 章

# Developer 工具中的引用数据

本章包括以下主题：

- [Developer 工具引用表概览, 25](#)
- [引用数据和转换, 25](#)
- [引用表, 26](#)
- [内容集, 29](#)

## Developer 工具引用表概览

可以在 Developer 工具中创建、更新和查看引用数据对象的配置属性。

使用 Developer 工具创建和更新以下类型的对象：

### 引用表

引用表包含一组数据值的标准版本和替代版本。将引用表添加到 Developer 工具中的某个转换，以确认源数据值是准确的且已正确设置格式。

### 内容集

内容集是在模型存储库或文件中指定引用数据值的模型存储库对象。内容集包含不同类型的引用数据，您可以将其用于在数据质量转换中执行搜索操作。

还可以在 Developer 工具中使用地址引用数据文件和标识填充文件。配置地址验证器转换时，选择地址引用数据文件。为标识匹配分析配置匹配转换时，选择标识填充文件。

## 引用数据和转换

多个转换读取引用数据以执行数据质量任务。

以下转换可以读取引用数据：

- 地址验证器。读取地址引用数据以确认地址的准确性。
- 大小写转换器。读取引用数据表以确定必须更改大小写的字符串。
- 分类器。读取内容集数据以确定字符串中信息的类型。
- 比较。在重复分析过程中读取标识填充数据。
- 标签创建器。读取内容集数据以标识字符串并为其添加标签。

- 匹配。在重复分析过程中读取标识填充数据。
- 解析器。读取内容集数据以根据字符串包含的信息解析字符串。
- 标准创建器。读取引用数据表以按照通用格式标准化字符串。

数据质量内容安装程序文件集包含您可以导入的 Informatica 引用数据对象。

# 引用表

将引用表添加到 Developer 工具中的某个转换。配置转换以在输入数据中查找引用表值并写入引用表中的对应有效值作为输出。

要在 Developer 工具中创建引用表，请使用以下方法之一：

- 创建空引用表并输入数据值。
- 从平面文件中的数据创建引用表。
- 从其他数据库表中的数据创建引用表。

## 引用表数据属性

可以在 Developer 工具中查看引用表数据和元数据的属性。从模型存储库打开引用表时，Developer 工具将显示这些属性。

引用表显示常规属性和列属性。可以在 Developer 工具中查看引用表属性。可以在 Analyst 工具中查看和编辑引用表属性。

下表介绍了引用表的常规属性：

属性	说明
名称	引用表的名称。
说明	引用表的可选说明。

下表介绍了引用表的列属性：

属性	说明
有效	确定包含有效引用数据的列。
名称	每个列的名称。
数据类型	每个列中数据的数据类型。
精度	每个列的精度。
小数位数	每个列的小数位数。
说明	列的内容说明。创建引用表时，可以选择添加说明。

属性	说明
包含用于行级说明的列	指示引用表包含用于列数据说明的列。
默认值	列中字段的默认值。创建引用表时，可以选择添加默认值。
连接名称	到包含引用表数据值的数据库的连接名称。

## 创建引用表对象

需要创建空引用表并手动添加值时选择该选项。

1. 从 Developer 工具菜单中选择 **文件 > 新建 > 引用表**。
2. 在新建表向导中，选择 **空引用表**。
3. 输入表的名称。
4. 选择用于存储表元数据的项目。

在“位置”字段中，单击 **浏览**。此时将打开 **选择位置** 对话框并显示存储库中的项目。选择所需的项目。  
单击 **下一步**。

5. 将两个或更多列添加到表中。单击 **新建** 选项创建列。

下表介绍了每个列的属性：

属性	默认值
名称	列
数据类型	字符串型
精度	10
小数位数	0
说明	空。可选属性。

6. 选择包含有效值的列。可以更改所创建的列的顺序。
7. 下表介绍了可选属性：

属性	默认值
包含用于行级说明的列	已清除
审计说明	空
默认值	空

单击 **完成**。

此时将在 Developer 工具工作区中打开引用表。

# 从平面文件创建引用表

可以从平面文件中存储的数据创建引用表。

- 1. 从 Developer 工具菜单中选择**文件 > 新建 > 引用表**。
- 2. 在新建表向导中，选择**从平面文件创建的引用表**。
- 3. 浏览至您要用做表的数据源的文件。
- 4. 输入表的名称。
- 5. 选择用于存储表元数据的项目。  
在“位置”字段中，单击**浏览**。此时将打开**选择位置**对话框并显示存储库中的项目。选择所需的项目。  
单击**下一步**。
- 6. 将 UTF-8 设置为代码页。
- 7. 指定平面文件使用的分隔符。
- 8. 如果平面文件包含列名称，则选择用于从文件的第一行导入列名称的选项。
- 9. 下表介绍了可选表属性：

属性	默认值
文本限定符	无引号
导入起始行	第 1 行
行分隔符	\012 LF (\n)
将连续分隔符视为一个整体	已清除
转义符	空
在数据中保留转义符	已清除
要预览的最多行数	500

单击**下一步**。

- 10. 选择包含有效值的列。
- 11. 下表介绍了可选属性：

属性	默认值
包含用于行级说明的列	已清除
审计说明	空
默认值	空
要预览的最多行数	500

单击**完成**。

此时将在 Developer 工具工作区中打开引用表。

# 从关系源创建引用表

可以使用数据库源创建受管或非受管引用表。

**注意:** 可以在连接浏览器中配置数据库连接。如果 Developer 工具不显示连接浏览器，则从 Developer 工具菜单中选择窗口 > 显示视图 > 连接浏览器。

1. 从 Developer 工具菜单中选择文件 > 新建 > 引用表。
2. 在新建表向导中，选择来自关系源的引用表。单击下一步。
3. 选择数据库连接。  
在“连接”字段中，单击浏览。此时将打开选择连接对话框并显示可用的数据库连接。  
要创建受管引用表，请连接至引用数据仓库。要创建非受管引用表，请连接至不同的数据库。  
选择连接时单击确定。
4. 如果您选择的数据库连接未指定引用数据仓库，请选择非受管表。  
如果您要对非受管引用表执行编辑操作，请选择可编辑选项。
5. 选择数据库表。  
在“资源”字段中，单击浏览。此时将打开选择资源对话框并显示数据库连接上的资源。浏览数据库并选择数据库表。
6. 输入表的名称。
7. 选择用于存储引用表对象的项目。  
在“位置”字段中，单击浏览。此时将打开选择位置对话框并显示存储库中的项目。  
选择项目并单击下一步。
8. 选择包含有效值的列。
9. 下表介绍了可以编辑的可选属性：

属性	默认值
包含用于行级说明的列	已清除
审计说明	空
默认值	空
要预览的最多行数	500

单击完成。

# 内容集

内容集是存储其他引用数据对象的数据或元数据的模型存储库。内容集可以包含字符集、模式集、标志集、正则表达式、概率模型和分类器模型。使用内容集定义和组织与单个项目、信息类型或业务目的相关的引用数据对象。

Developer 工具包含系统定义的不显示在模型存储库中的字符集和标志集。要查看和使用系统定义的对象，请在标签创建器转换、解析器转换或标准创建器转换中配置策略。

# 字符集

字符集包含可标识特定字符和字符范围的表达式。可以在使用为“字符添加标签”模式的标签创建器转换中使用字符集。

字符范围指定字符代码的顺序范围。例如，字符范围 “[A-C]” 与大写字符 “A”、“B” 和 “C” 相匹配。该字符范围与小写字符 “a”、“b” 或 “c” 不匹配。

使用字符集将特定字符或字符范围标识为“添加标签”操作的一部分。例如，可以在包含电话号码的列中为所有数字添加标签。为所有数字添加标签后，可以使用解析器转换标识模式，并将有问题的模式写入单独的输出口。

## 字符集属性

为字符集配置用于确定“为字符添加标签”操作的属性。

下表介绍了用户定义的字符集的属性：

属性	说明
标签	定义标签创建器转换应用至与字符集匹配的数据的标签。
标准模式	启用包含开始范围和结束范围字段的简单编辑视图。
开始范围	指定字符范围中的第一个字符。
结束范围	指定字符范围中的最后一个字符。对于具有单个字符的范围，将该字段留空。
高级模式	启用高级编辑视图，您可以在该视图中使用范围字符和分隔符字符手动输入字符范围。
范围字符	临时更改表示字符范围的符号。当您关闭字符集时，范围字符将还原为默认字符。
分隔符字符	临时更改分隔字符范围的符号。当您关闭字符集时，分隔符字符将还原为默认字符。

# 分类器模型

分类器模型分析输入字符串并确定其包含的信息的类型。在分类器转换中使用分类器模型。

当输入字符串包含大量数据时使用分类器模型。例如，可以使用分类器模型标识一组文档中的主题。导出每个文档中的文本，然后在单个数据列中将每个文档存储为单独的字段。分类器转换读取该数据并根据分类器模型中定义的标签对每个字段中的主题进行分类。

分类器模型包含以下各列：

“数据” 列

该列包含输入数据中可能存在的单词和短语。该转换会将输入数据与该列中的数据进行比较。

“标签” 列

该列包含一些描述性标签，可用于定义数据中的信息。该转换会在输出中返回该列中的某个标签。

分类器模型还包含编译数据，分类器转换可使用这些数据来计算输入数据的正确信息类型。

您可以在 Developer 工具中创建分类器模型。模型存储库用于存储分类器模型对象的元数据。列数据和编译数据存储在 Informatica 目录结构中的一个文件中。

# 模式集

模式集包含表达式，可用于标识“为标志添加标签”操作的输出中的数据模式。可以使用模式集分析标志化的数据输出端口，并将匹配的字符串写入一个或多个输出端口。在使用模式解析模式的解析器转换中使用模式集。

例如，可以将解析器转换配置为使用可标识姓名和首字母的模式集。在“为标志添加标签”模式下，该转换使用模式集分析标签转换的输出。可以将解析器转换配置为将输出中的姓名和首字母写入单独的端口。

## 模式集属性

配置用于确定模式集中的模式的属性。

下表介绍了用户定义的模式集的属性：

属性	说明
模式	定义模式解析器搜索的模式。可以为一个模式集输入多个模式。可以输入从通配符、字符和字符串的组合构造的模式。

# 概率模型

概率模型确定数据值表示的信息的类型。可以将概率模型与标签创建器转换和解析器转换一起使用。

概率模型包含一系列数据行和一系列相关的标签值。标签值描述数据行中的数据值包含的信息类型。为每个行中的每个值分配一个标签。编译概率模型时，Developer 工具在数据值和您指定的标签之间创建关联。

概率模型还包含编译数据，转换可使用这些数据来计算输入数据的正确信息类型。在 Developer 工具中编译模型时，将更新模型逻辑。

您可以在 Developer 工具中创建概率模型。模型存储库用于存储概率模型对象的元数据。Developer 工具将数据行、标签和编译数据写入一个文件。

# 正则表达式

在内容集的上下文中，正则表达式是可以在解析和添加标签操作中使用的表达式。使用正则表达式标识输入数据中的一个或多个字符串。可以在使用标志解析模式的解析器转换中使用正则表达式。还可以在使用“为标志添加标签”模式的标签创建器转换中使用正则表达式。

解析器转换使用正则表达式匹配输入数据中的模式，并将所有匹配的字符串解析到一个或多个输出。例如，您可以使用正则表达式标识输入数据中的所有电子邮件地址，并将每个电子邮件地址组件解析到不同的输出。

标签创建器转换使用正则表达式匹配输入模式并创建单个标签。包含多个输出的正则表达式不会生成多个标签。

## 正则表达式属性

配置确定正则表达式如何标识和写入输出字符串的属性。

下表介绍了用户定义的正则表达式的属性：

属性	说明
输出数	定义正则表达式写入的输出端口的数量。
正则表达式	定义解析器转换用于匹配字符串的模式。

属性	说明
测试表达式	包含您输入以测试正则表达式的数据。在该字段中键入数据后，该字段将突出显示与正则表达式匹配的字符串。
下一个表达式	移动至下一个与正则表达式匹配的字符串并将该字符串的字体更改为粗体。
上一个表达式	移动至上一个与正则表达式匹配的字符串并将该字符串的字体更改为粗体。

## 标志集

标志集包含能够标识特定标志的表达式。您可以在使用“为标志添加标签”模式的标签创建器转换中使用标志集。还可以在使用标志解析模式的解析器转换中使用标志集。

使用标志集可以将特定标志标识为添加标签和解析操作的一部分。例如，您可以使用标志集为所有采用“AccountName@DomainName”格式的电子邮件地址添加标签。为标志添加标签后，您可以使用解析器转换将这些电子邮件地址写入指定的输出端口。

### 标志集属性

为标志集配置用于确定添加标签操作的属性。

下表介绍了用户定义的字符集的属性：

属性	标志集模式	说明
名称	N/A	定义标志集的名称。
说明	N/A	描述标志集。
标志集选项	N/A	定义标志集使用正则表达式模式还是字符模式。
标签	正则表达式	定义标签创建器转换应用至与标志集匹配的数据的标签。
正则表达式	正则表达式	定义标签创建器转换用于匹配字符串的模式。
测试表达式	正则表达式	包含您输入以测试正则表达式的数据。在该字段中键入数据后，该字段将突出显示与正则表达式匹配的字符串。
下一个表达式	正则表达式	移动至下一个与正则表达式匹配的字符串并将该字符串的字体更改为粗体。
上一个表达式	正则表达式	移动至上一个与正则表达式匹配的字符串并将该字符串的字体更改为粗体。



属性	标志集模式	说明
标签	字符	定义标签创建器转换应用至与字符集匹配的数据的标签。
标准模式	字符	启用包含开始范围和结束范围字段的简单编辑视图。
开始范围	字符	指定字符范围中的第一个字符。
结束范围	字符	指定字符范围中的最后一个字符。对于单个字符范围，请将该字段留空。
高级模式	字符	启用高级编辑视图，您可以在该视图中使用范围字符和分隔符字符手动输入字符范围。
范围字符	字符	临时更改表示字符范围的符号。当您关闭字符集时，范围字符将还原为默认字符。
分隔符字符	字符	临时更改分隔字符范围的符号。当您关闭字符集时，分隔符字符将还原为默认字符。

## 创建内容集

创建内容集，以管理引用单个项目、信息类型或业务目的的引用数据对象。

1. 在**对象浏览器**视图中，选择用于存储内容集的项目或文件夹。
2. 单击**文件 > 新建 > 内容集**。
3. 输入内容集的名称。
4. （可选）选择**浏览**以更改内容集的模型存储库位置。
5. 单击**完成**。

## 在内容集中创建引用数据对象

可以在内容集中创建字符集、模式集、标志集、正则表达式、概率模型和分类器模型。

1. 在编辑器中打开内容集并选择**内容**视图。
2. 选择引用数据对象类型。
3. 单击**添加**。
4. 输入引用数据对象的名称。  
（可选）输入对象的说明。
5. 配置引用数据对象属性。
6. 单击**完成**。

**提示:** 可以将一个内容集中的引用数据对象复制到另一个内容集中。使用**复制到**和**粘贴自**选项可以创建内容集中某个对象的副本。使用 **Ctrl** 键可以选择多个内容集对象。

## 从中游配置文件生成引用数据

中游配置文件是您对映射中的转换运行的配置文件。可以运行中游配置文件，以便为引用数据对象生成数据。

例如，对您连接到标签创建器转换或解析器转换上的输入端口的转换运行配置文件。将配置文件数据添加到某个概率模型，然后将该概率模型应用至标签创建器转换或解析器转换。

1. 打开包含要剖析的转换的映射。
2. 选择该转换。  
例如，选择您连接到标签创建器转换或解析器转换上的输入端口的转换。
3. 单击**立即剖析**。
4. 在配置文件中选择“结果”选项卡，然后查看配置文件结果。
5. 在“列剖析”下，选择某个列。
6. 在“详细信息”下，选择用于显示配置文件值的选项。  
编辑器将显示您选择的列中的数据值。可以选择该列中的所有值或这些值的子集。
7. 将列数据导出到一个文件中。
  - 要导出所有列值，请单击**将值频率导出到文件**选项。
  - 要导出列值的子集，请右键单击这些值并选择**发送到 > 将结果导出到文件**。
8. 保存包含列数据的文件。可以将该文件保存在 Informatica 服务计算机或 Developer 工具计算机中。  
可以将该文件用作引用数据对象的数据源。例如，从该文件创建概率模型。

### 相关主题：

- [“从配置文件数据创建引用表” 页面上 15](#)

## 第 4 章

# 分类器模型

本章包括以下主题：

- [分类器模型概览, 35](#)
- [分类器模型结构, 36](#)
- [分类器模型视图, 36](#)
- [分类器模型引用数据, 37](#)
- [分类器模型标签数据, 38](#)
- [分类器得分, 40](#)
- [分类器模型筛选器, 40](#)
- [通过数据对象创建分类器模型, 41](#)
- [复制和粘贴操作, 42](#)

## 分类器模型概览

分类器模型是内容集中的引用数据对象。使用分类器模型分析包含多个值的长文本字符串。分类器模型确定每个字符串中最常见的信息类型。

将分类器模型添加到分类器转换中。转换搜索分类器模型数据和每个输入行中的数据之间的共有值。转换使用这些共有值对每一行代表的信息类型进行分类。

当输入数据具有以下特性时使用分类器模型：

- 输入数据包含文本。分类器模型将自然语言处理应用至文本数据以确定文本中信息的类型。自然语言处理检测输入字符串中的相关单词。自然语言处理忽略不相关的单词。
- 输入数据字符串包含多个值。例如，可以创建在每个字段中包含电子邮件内容的数据列。

分类器转换读取字符串数据类型。转换对输入字符串的长度没有限制。

在 Developer 工具中编译分类器模型。编译模型时，将在模型中的相似数据值之间创建关联。分类器转换使用编译的数据搜索输入数据中的信息。

## 分类器转换示例

可以使用分类器模型和分类器转换根据电子邮件包含的文本对电子邮件进行分类。

例如，您在客户支持中心工作，需要查看组织从客户处收到的电子邮件。组织的客户遍布多个国家/地区，会收到使用多种语言撰写的电子邮件。您需要按语言对这些电子邮件进行排序，从而可以将每封电子邮件发送到可以对客户进行最佳回复的中心。

完成以下步骤，以便对电子邮件进行排序：

1. 将电子邮件写入单个文件或一个数据库表。
2. 创建包含每种语言的示例文本的分类器模型。  
**注意：**可以使用电子邮件数据中的示例数据作为模型的源数据。将电子邮件文本复制到一个文件或数据库表，并从模型存储库中的文件或表创建数据源。
3. 将分类器模型添加到分类器转换中。
4. 将转换添加到映射，并将转换端口连接到数据源和数据目标。为每一种语言创建一个数据目标。

运行映射时，分类器转换分析电子邮件并将电子邮件文本写入正确的数据目标。可以与相应支持中心中的团队成员共享数据目标。

## 分类器模型结构

分类器模型包含一个引用数据值列和一个为引用数据值的每一行指定标签的列。当分类器转换将输入数据和模型数据进行对比时，转换返回最能够描述输入数据的标签。

分类器模型还包含编译数据。分类器转换使用此编译数据来计算引用数据与输入数据之间的相似度。编译模型时，可以创建或更新编译数据。

模型存储库存储分类器模型对象。创建或更新分类器模型时，将引用数据和编译的元数据写入 Developer 工具计算机上的某个文件。该文件是只读的。可以在 Developer 工具中的分类器模型中读取该文件路径。

## 分类器模型视图

可以使用默认视图和详细视图更新分类器模型中的数据。默认视图在一个表中显示标签值和数据值。详细视图在一系列文本框中显示数据值。

使用默认视图查看和更新每一行上的标签。可以选择一行、多行和所有行。默认视图可以显示大约 100 个字符的行数据。详细视图可以显示每一行中的所有数据。单击“分类器模型数据”选项可以在这些视图之间进行切换。使用详细视图查看和更新单个行中的数据。

可以在每个视图中添加数据，筛选数据行以及将标签添加到行。在详细视图中，可以搜索单个行中的数据值。

下图显示了包含语言分类的数据的分类器模型默认视图。



1. 在数据行中查找值。
2. 插入一个数据行。
3. 管理标签。
4. 编译分类器模型。

要为某数据行分配标签，请选择该数据行并单击标签名称。

## 分类器模型引用数据

分类器模型包含引用数据列，该列可以包含文本句子、段落或页面。引用数据代表分类器转换可以在映射中读取的不同类型的文本输入。创建模型时，确认引用数据包含您希望在运行映射时查找的文本类型。

可以使用映射源数据创建分类器模型。选择一个源数据示例并将该数据示例复制到模型中。

在使用分类器模型引用数据时，请考虑以下规则和准则：

- 引用数据字段可以具有任何长度。可以将文本页面输入每个数据字段。
- 从数据对象导入引用数据。
- 无法编辑引用数据值。但是，可以删除数据行。

## 将数据和标签值添加到分类器模型

使用数据源将数据附加到分类器模型。可以添加数据值和标签值。

1. 打开包含该模型的内容集。
2. 选择模型名称并单击**编辑**。
3. 单击**附加数据**。

此时将打开分类器模型向导。

4. 浏览模型存储库并选择您要使用的数据对象。单击**下一步**。

**注意：**请勿选择社交媒体数据对象作为数据源。

5. 查看数据对象上的列，并选择要作为模型的数据列或标签列进行添加的列。可以在同一操作中添加引用数据列和标签列。

- 要将某数据源列用作模型中的引用数据列，请选择该列名称并单击**数据**。

可以选择多个数据列。分类器模型将您选择的列的内容合并为单个列。

- 要将某数据源列用作模型的标签列，请选择该列名称并单击**标签**。

单击**下一步**。

6. 选择要从数据源复制的行数。

选择所有行，或输入要复制的行数。如果输入数字，则模型会从数据集的开头对行进行计数。

7. 单击**完成并保存模型**。

附加数据之后，确认您添加的数据行包含标签值。

## 从分类器模型中删除数据值

可以在分类器模型的默认视图和详细视图中删除数据值。要删除所有数据值，请使用默认视图。

1. 在 Developer 工具中打开分类器模型。要打开模型，请在内容集中选择模型名称并单击**编辑**。
2. 选择包含您要删除的数据的行。  
可以选择单个行、多行和所有行。
3. 单击**删除**。

## 分类器模型标签数据

分类器模型包含引用数据字段中信息类型的描述性标签。创建模型时，将为每个数据字段分配一个标签。

将数据源添加到分类器模型时，可以选择一个列作为标签数据列。还可以向模型输入标签名称。可以将模型中的任何标签分配给任何引用数据字段。

标签独立于其描述的引用数据值。如果删除包含所选标签的所有表行，不会从模型中删除该标签。如果删除某个标签，不会删除与该标签关联的引用数据值。

## 将标签添加到分类器模型

如果将不具有标签的数据值添加到分类器模型，则单独添加标签。

1. 打开包含该模型的内容集。
2. 选择模型名称并单击**编辑**。

3. 选择**新建标签**选项添加标签。  
此时将在模型中显示标签。
4. 在**新建标签**对话框中键入名称。
5. 单击**确定**。

## 为分类器模型数据分配标签

标签代表分类器模型可以在源数据中确定的信息类型。创建分类器模型时，确认模型中的每个数据行具有一个标签。

可以将一个标签分配给一个或多个行。

1. 打开包含该模型的内容集。
2. 选择模型名称并单击**编辑**。
3. 在模型中筛选数据行以显示不具有标签的行。  
要显示不具有标签的行，请在**标签**面板中清除所有标签名称。
4. 选择一个或多个数据行。可以使用“全选”选项选择显示的所有行。  
模型会将复选标记添加到您选择的行。
5. 在模型中浏览标记值，然后选择要应用至数据行的标签。  
模型将该标记分配给您选择的行。
6. 编译模型，以便将标记名称添加至分类器逻辑。

如果您分配已从标记名称显示中清除的标签，则模型隐藏这些行。在“**标签**”面板中选择该标签名称可查看这些行。

## 从分类器模型中删除标签

可以在分类器模型的默认视图或详细视图中删除标签。

1. 打开包含该模型的内容集。
2. 选择模型名称并单击**编辑**。
3. 单击**属性**。
4. 在**管理标签**对话框中，选择一个或多个要删除的标签。  
可以选择多个标签。
5. 单击**删除**。
6. 单击**确定**关闭对话框。

## 编译分类器模型

每次在分类器模型中编辑标签值或引用数据值时，必须编译模型。编译模型时，将更新模型中的编译数据。

- 要更新编译数据，请在 Developer 工具中打开模型并单击**编译**。

## 分类器得分

分类器转换将输入数据中的每一行与分类器模型中引用数据的每一行进行比较。转换计算每个比较的得分。得分代表输入行与引用数据行之间的相似度。

运行某个包含分类器转换的映射时，该映射返回使用最高得分标识引用数据行的标签。得分范围为从 0 到 1。高分指示输入数据与模型数据之间的强匹配。

查看分类器得分以确认标签输出准确描述输入数据的每一行。还可以查看得分以确认分类器模型适合输入数据。如果转换输出包含很大比例的低得分，则分类器模型可能不适合。要改进比较，请再次编译模型。如果已编译的模型不能提高得分，请替换转换中的模型。

## 分类器模型筛选器

可以在默认视图和详细视图中将筛选器应用至分类器模型数据。

可以使用筛选器执行以下任务：

- 查找不具有关联标签的数据值。使用标签选项筛选分类器模型中显示的数据行。如果某数据行不使用标签，则将标签添加到该行。
- 在引用数据行中查找数据值。在默认视图中使用筛选器，以便在引用数据中查找数据值。确认引用数据与映射中的源数据重叠。
- 在引用数据行中查找数据值。当您需要确认某个引用数据行包含某数据值时，在详细视图中使用筛选器。一个数据行中可以包含大量数据值。

## 在引用数据中查找值

在默认视图中使用筛选器确认引用数据行包含所需的数据值。

1. 打开包含该模型的内容集。
2. 选择模型名称并单击**编辑**。
3. 在筛选器字段中键入文本值。  
Developer 工具将显示包含筛选器文本的数据行。

## 查找不具有标签的数据行

清除所有标签选项以显示不具有标签的引用数据行。

打开筛选器模型时，默认情况下 Developer 工具显示所有行和标签。

1. 打开包含该模型的内容集。
2. 选择模型名称并单击**编辑**。
3. 在默认视图中，选择或清除某个标签。  
选择某个标签时，模型显示您将其与该标签关联的数据字符串。清除某个标签时，模型隐藏您将其与该标签关联的数据字符串。
4. 要确认所有数据字符串都使用标签，请清除所有标签值。模型将显示任何不使用标签的字符串。
5. 单击某个标签值将该标签添加到数据字符串。



## 在数据行中查找数据

在详细视图中使用过滤器，以便在单个数据行中搜索某个数据值。

1. 打开包含该模型的内容集。
2. 选择模型名称并单击**编辑**。
3. 选择详细视图。
4. 在详细视图中，在筛选器字段中输入某个值。  
模型将显示包含该值的数据行。
5. 选择要搜索的数据行。
6. 在该数据行下面的搜索字段中键入搜索值。  
模型在行中突出显示该值的第一个实例。
7. 单击向下箭头在行中查找该值的下一个实例。  
使用向上箭头和向下箭头在数据行中的该值之间移动。

## 通过数据对象创建分类器模型

将数据对象用作分类器模型数据的源。

将分类器转换的输入数据用作模型引用数据的源时，分类器模型的执行效果最佳。例如，可以对映射中的转换对象运行配置文件。从配置文件结果创建数据对象。

1. 在对象浏览器中，打开或创建一个内容集。
2. 选择**内容**视图。
3. 选择**分类器模型**，然后单击**添加**。  
此时将打开分类器模型向导。
4. 输入分类器模型的名称。  
或者，输入模型的文本说明。
5. 浏览模型存储库并选择包含引用数据的数据对象。  
单击**下一步**。
6. 查看数据对象上的列，并选择要作为模型的引用数据值或标签值进行添加的列。
  - 要添加某数据列作为引用数据，请选择该列名称并单击**数据**。
  - 要将某数据列用作标签值的源，请选择该列名称并单击**标签**。单击**下一步**。
7. 选择要从数据源复制的行数。  
选择所有行，或输入要复制的行数。如果输入数字，则模型会从数据集的开头对行进行计数。
8. 单击**完成**并保存模型。  
创建分类器模型后，编译该模型。

# 复制和粘贴操作

可以将模型存储库中一个内容集中的分类器模型复制到另一个内容集中。复制分类器模型，以便与其他 Developer 工具用户共享资源。

可以将模型复制到其他内容集，也可以将模型导入当前内容集。可以在单个操作中从存储库中的多个内容集导入多个模型。

复制模型时，内容管理服务在服务计算机上创建模型数据文件的副本。每个模型使用不同的数据文件。

## 将分类器模型复制到其他内容集

可以将模型存储库中一个内容集中的分类器模型复制到另一个内容集中。复制分类器模型时，应指定模型对象以及源内容集和目标内容集。

1. 打开包含该分类器模型的内容集。
2. 选择一个分类器模型并单击**复制到**。
3. 浏览模型存储库并选择一个内容集。

可以将分类器模型复制到当前项目或其他项目中的内容集。

4. 单击**确定**。

Developer 工具将分类器模型复制到所选的内容集中。

## 从其他内容集导入分类器模型

在模型存储库中，可以将一个内容集中的分类器模型导入另一个内容集。导入分类器模型时，应指定一个或多个模型对象以及源内容集和目标内容集。

1. 打开包含该分类器模型的内容集。
2. 选择一个分类器模型并单击**粘贴自**。
3. 浏览模型存储库并选择一个分类器模型。

可以从当前项目或其他项目中的内容集粘贴分类器模型。

4. 单击**确定**。

Developer 工具将分类器模型粘贴到当前内容集中。

## 第 5 章

# 概率模型

本章包括以下主题：

- [概率模型概览, 43](#)
- [概率模型结构, 44](#)
- [概率模型视图, 44](#)
- [标签创建器转换示例, 45](#)
- [解析器转换示例, 46](#)
- [概率模型引用数据, 47](#)
- [概率模型标签数据, 47](#)
- [概率模型属性, 49](#)
- [从数据对象创建概率模型, 50](#)
- [复制和粘贴操作, 50](#)

## 概率模型概览

概率模型是您在内容集中创建的引用数据对象。使用概率模型可以分析包含多个数据值的数据字符串。概率模型标识字符串中每个值中的信息类型。可以将概率模型添加到标签创建器转换和解析器转换中。

在标签创建器转换中使用概率模型可以为输入字符串中的每个值分配描述性标签。标签创建器转换将标签写入单个输出端口。在解析器转换中使用概率模型可以将输入字符串中的每个值写入代表值中的信息的端口。解析器转换为每种类型的信息创建一个输出端口。

您可以在 Developer 工具中设计和编译概率模型。定义概率模型时，将一系列数据行添加到模型中并为每个行中的每个值分配一个标签。编译概率模型时，Developer 工具在数据值和您添加的标签之间创建关联。标签创建器转换和解析器转换使用自然语言处理将概率模型数据与输入端口数据进行比较。

自然语言处理使用以下技术来确定数据值中信息的类型：

- 自然语言处理可以识别相似的数据值并为这些值分配同一个标签。
- 自然语言处理可以将某个数据值与字符串中的相邻值进行比较。自然语言处理分析值序列以了解每个字符串的用法并确认字符串代表的信息类型。

# 概率模型结构

概率模型包含一组数据行和一组对应的标签。数据行包含可能显示在转换输入数据中的不同值的示例。标签标识您希望输入值包含的信息类型。

定义概率模型时，需要将每个标签分配给数据行中的一个或多个值。编译模型时，Developer 工具生成表示标签值与数据值之间的关系的引用数据对象。

一个数据行可以包含单个值或多个值。每个数据行可能具有不同的结构。您可以为一个数据行中的多个值分配同一个标签。或者，可以为显示在一个行中不同位置的相同值分配不同的标签。在编译概率模型之前，将每个标签分配给至少一个数据值。

**注意：**要优化概率模型的功能，请确认每个数据行包含多个值。每个行中值的顺序必须尽可能接近地与值在源数据中出现的顺序相对应。如果数据行包含单个值，则标签创建器转换或解析器转换无法在输入数据分析中应用自然语言处理。

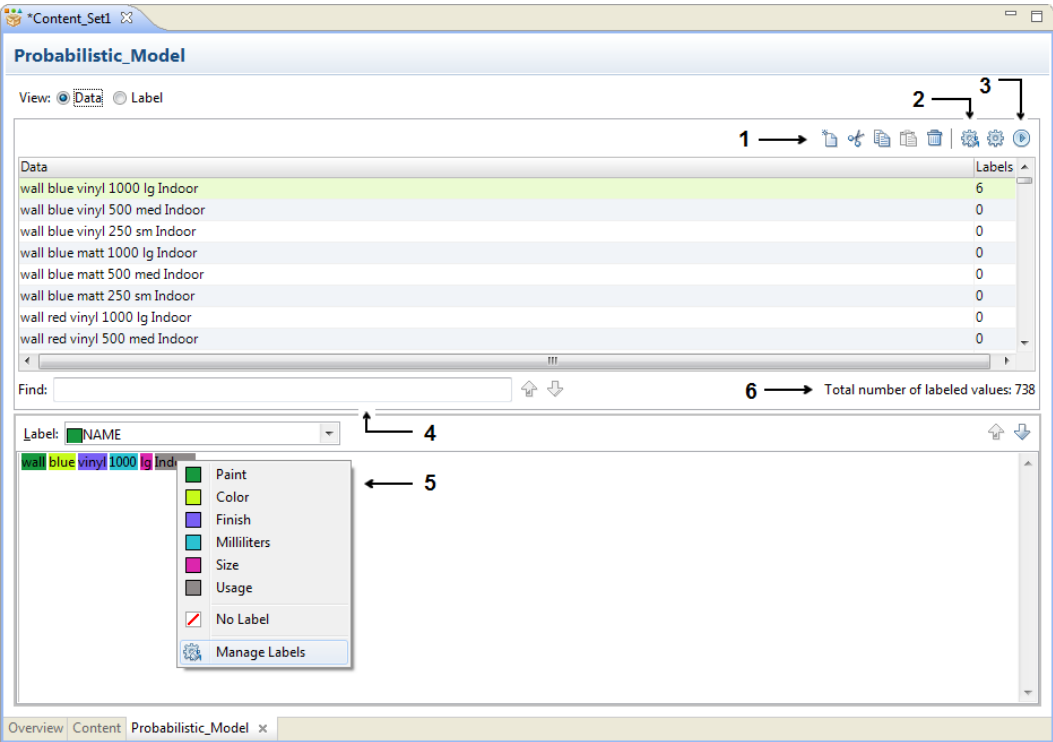
## 概率模型视图

编辑概率模型时，可以在“数据”视图或“标签”视图中工作。使用“数据”视图可以将数据行添加到模型中，并可以为每个行中的数据值分配标签。使用“标签”视图可以查找未使用的标签。

“数据”视图显示概率模型中的数据行并显示您为每个行分配的标签数。选择某个行时，该行将显示在编辑器中数据行列的下方。要为行中的数据值分配某个标签，请在编辑器中右键单击该值并从菜单中选择该标签。“数据”视图还显示您为当前概率模型中的值分配的标签总数。

“标签”视图列出您在概率模型中定义的标签。选择某个标签时，“标签”视图将显示包含该标签描述的值的行。“标签”视图显示您为每个行中的标签分配的数据值。

下图显示了概率模型的“数据”视图：



1. 插入一个数据行。
2. 管理标签。
3. 编译模型。
4. 在数据行中查找值。使用箭头可以查找包含该值的下一行或上一行。
5. 为数据值分配标签。
6. 显示具有标签的数据值的总数。

# 标签创建器转换示例

某保险组织中的客户数据库包含多个数据条目错误。您是该保险组织的数据管理者。您使用标签创建器转换配置一个映射，以确定每个列包含的不同数据类型。

下表介绍了客户数据库中的示例数据：

行 ID	字段 1	字段 2	字段 3
1	19132954	AIM SECURITIES	PETRIE TAYBRO
2	10110169	JASE TRAPANI	BANK OF NEW YORK
3	10111786	WANGER ASSET MANAGEMENT, LLP	JAN SEEDORF
4	10112299	FELIX LEVENER	HARVARD MAGAZINE
5	10112036	DESCHÊNES & FILS LTÉE (QUEBEC)	RICHARD TREMBLAY

行 ID	字段 1	字段 2	字段 3
6	BERGER ASSOCIATES	10111101	DAREEN HULSMAN
7	19131385	EAGLE FINANCIAL GROUP INC	PATRICK MCKINNIE
8	LAKENYA PASKETT	WHITEHALL FINANCIAL GROUP	15954710

当您运行该映射时，标签创建器转换将输入数据与概率模型引用数据进行比较。标签创建器转换为每个输入数据分配一个标签。该转换将这些标签写入一个输出端口。每个输出行包含一组用于定义对应输入行上的数据结构的标签。

下表介绍了标签创建器转换添加到输出端口的标签：

行 ID	输出标签
1	数字 组织 联系人
2	数字 联系人 组织
3	数字 组织 联系人
4	数字 联系人 组织
5	数字 组织 联系人
6	组织 数字 联系人
7	组织 数字 联系人
8	联系人 组织 数字

# 解析器转换示例

某超市在数据库表的单个列中存储产品说明。这些产品说明包含多个代表不同类型的信息的数据值。您是该超市的数据管理者。您需要为这些产品说明中不同类型的信息创建列。

您需要使用解析器转换配置一个映射，以便将数据值组织到正确的字段中。

以下数据片段包含橙汁的产品说明：

Sunnydream Orange Juice Unsweetened 12 oz

下表介绍了解析器转换从输入数据创建的输出数据：

产品名称	产品类型	产品详细信息	产品规格
Sunnydream	Orange Juice	Unsweetened	12 oz

# 概率模型引用数据

概率模型中的引用数据值代表标签创建器转换或解析器转换可能在映射中读取的输入数据的类型。

您可以在 Developer 工具中添加、编辑和删除引用数据行。您可以从剪贴板粘贴引用数据。您还可以将映射源数据用作概率模型中引用数据值的源。

添加数据值后，创建标签列并为模型的每个行中的每个数据值分配一个标签。

**注意：**模型中的标签值指示转换读取的输入行的可能结构。确认概率模型包含您希望转换在源数据中找到的标签结构。

## 将引用数据字符串添加到概率模型

使用“数据”视图可以将引用数据字符串添加到概率模型中的行。

1. 打开包含该模型的内容集。  
选择模型名称并单击**编辑**。
2. 选择“数据”视图。
3. 单击**新建**将空行添加到模型中。
4. 在该空行中输入数据字符串。
5. 为该字符串中的每个值分配一个标签。
6. 保存概率模型。

保存模型后，可以选择编译模型。

## 在概率模型中查找数据值

使用“数据”视图可以在模型中查找包含某个数据值的数据行。

1. 打开包含该模型的内容集。  
选择模型名称并单击**编辑**。
2. 选择“数据”视图。
3. 在**查找**字段中输入一个数据值。  
还可以输入以下通配符：  
\* ?  
“数据”视图显示模型中第一个包含您输入的值的行。
4. 使用向上箭头或向下箭头查看每一个包含该值的行。

# 概率模型标签数据

概率模型中的标签值代表引用数据值可能包含的信息类型。将数据行添加到模型时，为每个行中的每个数据值添加一个标签。您添加到模式的标签显示在“标签”视图以及“数据”视图的菜单选项中。

可以将模型中的任何标签分配给任何引用数据值。如果同一个值在引用数据的两行中具有不同的含义，则可以在每一行中为该值分配不同的标签。

可以为多个输入字符串定义同一个标签组合。一个标签的多个示例可以提高概率模型的准确性。

# 溢出标签

当转换无法为输入数据值分配您定义的标签时，转换将为该数据分配溢出标签。

标签创建器转换为其无法识别的任何数据值分配溢出标签。解析器转换为未分配的数据创建一个溢出列。

下表显示了解析器转换如何使用溢出端口解析概率模型无法识别的地址数据元素：

输入数据	Street_Name 端口	Street_Descriptor 端口	溢出端口
Park Place	Park	Place	
Park Avenue	Park	Avenue	
Madison Avenue	Madison	Avenue	
Central Park	Central	Park	
Washington Square Park	Washington	Square	Park
Madison Square Garden	Madison	Square	Garden

当输入数据值的数量大于概率模型中标签的数量时，该转换为溢出端口分配值。在转换中使用某个模型之前，预览映射源数据并确认该模型包含正确数量的标签值。

# 为概率模型数据分配标签

为概率模型中的所有数据值分配一个标签。

如果同一个数据值出现在数据行中的不同位置，则可以为该数据值分配不同的标签。如果数据值使用的标签不正确，则可以更新该标签。

1. 打开包含该模型的内容集。
2. 选择模型名称并单击**编辑**。
3. 选择“数据”视图。
4. 查找没有标签或标签不正确数据值。
5. 选择包含该数据值的数据行。  
此时将在编辑器中显示该行。
6. 在编辑器中右键单击某数据值并从上下文菜单中选择一个标签。  
Developer 工具将为该数据值分配此标签。
7. 保存概率模型。

保存模型后，可以选择编译模型。

# 将标签添加到概率模型

为模型数据值代表的每一个信息类型添加一个标签。如果在解析器转换中使用概率模型，则为您希望该转换创建的每个输出端口添加一个标签。

1. 打开包含该模型的内容集。



2. 选择模型名称并单击**编辑**。
3. 在“数据”视图或“标签”视图中，单击**管理标签**。  
此时将显示**管理标签**对话框。
4. 在**管理标签**对话框中，单击**新建**。  
此时将在该对话框的第一个空行中显示一个标签。
5. 编辑标签名称。（可选）更新标签的颜色。
6. 单击**确定**将标签添加到模型中。
7. 保存概率模型。

添加标签后，必须至少为一个数据值分配该标签。在转换中使用该模型之前，需要编译该模型。

## 从概率模型中删除标签

从模型中删除某个标签时，与该标签关联的任何数据值都会保留在模型中。为每个数据值分配其他标签。

1. 打开包含该模型的内容集。
2. 选择模型名称并单击**编辑**。
3. 在“数据”视图或“标签”视图中，单击**管理标签**。  
此时将显示**管理标签**对话框。
4. 在**管理标签**对话框中，单击**删除**。
5. 单击**确定**删除标签。
6. 保存概率模型。

**注意：**标签是概率模型中的结构元素。如果您在将模型添加到转换之后添加或删除标签，使用该模型的操作将变得无效。要使用您更新的模型，请删除并重新创建该转换操作。

## 编译概率模型

每次您更新某个概率模型时，可以编译该模型。编译该模型以使用当前数据值和标签值更新模型逻辑。

在编译模型之前，请确认所有标签值至少标识一个数据值。

- 要编译模型，请在 Developer 工具中打开模型并单击**编译**。

## 概率模型属性

可以查看和更新概率模型的常规属性和高级属性。

要打开属性编辑器，请在“数据”视图中选择**编辑属性**选项或在“标签”视图中选择**编辑高级属性**选项。

常规属性显示概率模型的名称、模型的任何说明以及模型数据文件的名称。高级属性显示 Developer 工具用于编译概率模型的计算属性。

概率模型编译中的基本元素是 *n-gram*。n-gram 是一系列位于其他字母之后或之前以完成某个单词的字母。当映射运行时，标签创建器转换或解析器转换将为概率模型引用数据列中的每个值创建多个 n-gram。转换将输入数据值与引用数据值和 n-gram 进行比较。概率模型的高级属性决定概率模型如何处理 n-gram 和其他模型特征。

**注意：**高级属性的默认值代表概率分析和概率模型编译的首选设置。如果您编译高级属性，则可能会对概率分析的准确性产生不利影响。除非您了解所进行的更改的作用，否则不要编辑高级属性。

## 从数据对象创建概率模型

可以将数据对象用作概率模型数据的源。例如，使用映射中读取概率模型的源数据对象。还可以剖析映射中的对象并从配置文件结果创建数据对象。

将标签创建器或解析器转换的输入数据用作模型引用数据的源时，概率模型的执行效果最佳。例如，可以对映射中的转换对象运行配置文件。从配置文件结果创建数据对象。

1. 打开或创建内容集。
2. 选择**内容**视图。
3. 选择**概率模型**，然后单击**添加**。

此时将打开概率模型向导。

4. 选择**来自数据对象的概率模型**选项。  
单击**下一步**。

5. 输入概率模型的名称。  
或者，输入模型的文本说明。

6. 浏览模型存储库并选择包含引用数据的数据对象。  
单击**下一步**。

7. 查看数据对象中的数据列，然后选择要添加到概率模型中的列。

- 要添加某数据列作为引用数据，请选择该列名称并单击**数据**。可以添加多个引用数据列。导入操作会联接这些列，以便同一行的每列中的值显示在模型中的同一数据行上。
- 要将某数据列用作标签值的源，请选择该列名称并单击**标签**。可以添加多个标签数据列。导入操作会连接这些列，以便每一列中的值以上下依次相邻的方式显示在模型中的单个列中。

单击**下一步**。

8. 选择要从数据源复制的行数。  
选择所有行，或指定要复制的行数。如果输入数字，则模型会从数据集的开头对行进行计数。
9. 设置要用于引用数据值的分隔符。指定用于标识代表单条信息的多个值的分隔符。  
默认分隔符是一个字符空格。

10. 单击**完成并保存模型**。  
此时将在 Developer 工具中打开概率模型。

创建概率模型后，为数据值分配标签并编译模型。

## 复制和粘贴操作

在模型存储库中，可以将一个内容集中的概率模型复制到另一个内容集中。复制概率模型，以便与其他 Developer 工具用户共享资源。

可以将模型复制到其他内容集，也可以将模型导入当前内容集。可以在单个操作中从存储库中的多个内容集导入多个模型。

复制模型时，内容管理服务在服务计算机上创建模型数据文件的副本。每个模型使用不同的数据文件。

## 将概率模型复制到其他内容集

在模型存储库中，可以将一个内容集中的概率模型复制到另一个内容集中。复制概率模型时，应指定模型对象以及源内容集和目标内容集。

1. 打开包含该概率模型的内容集。
2. 选择一个概率模型并单击**复制到**。
3. 浏览模型存储库并选择一个内容集。

可以将概率模型复制到当前项目或其他项目中的内容集。

4. 单击**确定**。

Developer 工具将概率模型复制到所选的内容集中。

## 从其他内容集导入概率模型

在模型存储库中，可以将一个内容集中的概率模型导入另一个内容集。导入概率模型时，应指定一个或多个模型对象以及源内容集和目标内容集。

1. 打开要包含该概率模型的内容集。
2. 选择一个概率模型并单击**粘贴自**。
3. 浏览模型存储库并选择一个概率模型。

可以从当前项目或其他项目中的内容集粘贴概率模型。

4. 单击**确定**。

Developer 工具将概率模型粘贴到当前内容集中。

# 索引

## A

Analyst 工具  
查找和替换引用数据值 [21](#)

## C

查看审计表事件  
引用表 [23](#)  
从列模式创建引用表  
引用表 [16](#)  
从列值创建引用表  
引用表 [16](#)  
从配置文件列创建引用表  
引用表 [15](#)

## D

导出引用表  
引用表 [21](#)  
导入引用表  
引用表 [18](#)

## G

管理列  
引用表 [20](#)  
管理行  
引用表 [20](#)

## L

列属性  
Analyst 工具中的引用表 [13](#)  
Developer 工具中的引用表 [26](#)

## N

内容管理服务  
引用表特权 [12](#)

## P

平面文件属性  
Analyst 工具中的引用表 [13](#)  
Developer 工具中的引用表 [26](#)

## S

手动创建引用表  
引用表 [14](#)

## Y

引用表  
Analyst 工具概览 [13](#)  
Analyst 工具中的列属性 [13](#)  
Analyst 工具中的平面文件属性 [13](#)  
查看审计跟踪表 [23](#)  
从列模式创建引用表 [16](#)  
从列值创建引用表 [16](#)  
从配置文件列创建引用表 [15](#)  
导出引用表 [21](#)  
导入引用表 [18](#)  
Developer 工具中的列属性 [26](#)  
Developer 工具中的平面文件属性 [26](#)  
管理列 [20](#)  
管理行 [20](#)  
手动创建引用表 [14](#)  
受管和非受管 [12](#)  
特权 [12](#)  
在 Analyst 工具中查找和替换值 [21](#)  
在 Analyst 工具中刷新 [22](#)