



Informatica® Data Quality
(版本 10.1.1)

入门指南

Informatica Data Quality 入门指南

版本 10.1.1
2016 年 12 月

© 版权所有 Informatica LLC 1998, 2017

本软件和文档包含 Informatica LLC 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 LLC 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、Informatica Master Data Management 和 Live Data Map 是 Informatica LLC 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 (C) Genivia, Inc。保留所有权利。版权所有 (C) 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) Unicon, Inc。保留所有权利。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。版权所有 (C) jQWidgets。保留所有权利。版权所有 (C) Tableau Software, Inc。保留所有权利。版权所有 (C) MaxMind, Inc。保留所有权利。版权所有 (C) Tmate Software s.r.o。保留所有权利。版权所有 (C) MapR Technologies Inc。保留所有权利。版权所有 (C) Amazon Corporate LLC。保留所有权利。版权所有 (C) Highsoft。保留所有权利。版权所有 (C) Python Software Foundation。保留所有权利。版权所有 (C) BeOpen.com。保留所有权利。版权所有 (C) CNRI。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>。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 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/hsqldbLicense.html>、<http://httpunit.sourceforge.net/doc/license.html>、<http://jung.sourceforge.net/license.txt>、http://www.gzip.org/zlib/zlib_license.html、<http://www.openldap.org/software/release/license.html>、<http://www.libssh2.org>、<http://sif4j.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://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://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>、<https://github.com/hjiang/jsonxx/blob/master/LICENSE>、<https://code.google.com/p/lz4/>、<https://github.com/jedisctl/libsodium/blob/master/LICENSE>、<http://one-jar.sourceforge.net/index.php?page=documents&file=license>、<https://github.com/EsotericSoftware/kryo/blob/master/license.txt>、<http://www.scala-lang.org/license.html>、<https://github.com/tinkerpop/blueprints/blob/master/LICENSE.txt>、<http://gee.cs.oswego.edu/dl/classes/EDU/oswego/cs/dl/util/concurrent/intro.html>、<https://aws.amazon.com/asl/>、<https://github.com/twbs/bootstrap/blob/master/LICENSE> 和 <https://sourceforge.net/p/xmlunit/code/HEAD/tree/trunk/LICENSE.txt> 下许可的软件。

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。

请参阅位于以下位置的专利：<https://www.informatica.com/legal/patents.html>。

免责声明：Informatica LLC 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 LLC 不保证本软件和文档中没有错误。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。

发布日期：2017-01-17

目录

前言	8
Informatica 资源.....	8
Informatica Network.....	8
Informatica 知识库.....	8
Informatica 文档.....	8
Informatica 产品可用性矩阵.....	9
Informatica Velocity.....	9
Informatica Marketplace.....	9
Informatica 全球客户支持部门.....	9
第 1 章： 入门概览	10
Informatica 域概览.....	10
功能的可用性.....	12
介绍 Informatica Analyst.....	12
Informatica Developer 概览.....	13
Informatica Developer 欢迎页.....	14
备忘单.....	14
数据质量和剖析.....	14
教程讲解.....	15
教程结构.....	15
Informatica Analyst 教程.....	15
Informatica Developer Tool.....	16
教程先决条件.....	17
第 I 部分： Informatica Analyst 入门	18
第 2 章： 课程 1。设置 Informatica Analyst	19
设置 Informatica Analyst 概览.....	19
任务 1. 登录 Informatica Analyst.....	20
任务 2. 创建项目.....	20
任务 3. 创建文件夹.....	20
设置 Informatica Analyst 摘要.....	21
第 3 章： 课程 2。创建数据对象	22
创建数据对象概览.....	22
任务 1. 创建平面文件数据对象.....	23
任务 2. 查看数据对象属性.....	23
创建数据对象摘要.....	24

第 4 章： 课程 3. 创建默认配置文件.....	25
创建默认配置文件概览.....	25
任务 1. 创建并运行默认配置文件.....	26
任务 2. 在摘要视图中查看配置文件结果.....	26
创建默认配置文件摘要.....	27
第 5 章： 课程 4. 创建自定义配置文件.....	28
创建自定义配置文件概览.....	28
任务 1. 创建自定义配置文件.....	29
任务 2. 运行配置文件.....	29
任务 3. 向下钻取配置文件结果.....	30
创建自定义配置文件课程总结.....	30
第 6 章： 课程 5. 创建表达式规则.....	31
创建表达式规则概览.....	31
任务 1. 创建表达式规则并运行配置文件.....	32
任务 2. 查看表达式规则的输出.....	32
任务 3. 编辑表达式规则.....	33
创建表达式规则课程总结.....	33
第 7 章： 课程 6. 创建并运行结果卡.....	34
创建并运行结果卡概览.....	34
任务 1. 使用配置文件结果创建结果卡.....	35
任务 2. 运行结果卡.....	36
任务 3. 查看结果卡.....	36
任务 4. 编辑结果卡.....	37
任务 5. 配置阈值.....	37
任务 6. 查看得分趋势图表.....	37
创建并运行结果卡课程总结.....	38
第 8 章： 课程 7. 使用配置文件列创建引用表.....	39
使用配置文件列创建引用表概览.....	39
任务 1. 使用配置文件列创建引用表.....	40
任务 2. 编辑引用表.....	40
使用配置文件列创建引用表课程总结.....	41
第 9 章： 课程 8. 创建引用表.....	42
创建引用表概览.....	42
任务 1. 创建引用表.....	43
创建引用表课程总结.....	43

第 II 部分： Informatica Developer 入门.....	44
第 10 章： 课程 1。设置 Informatica Developer.....	45
设置 Informatica Developer 概览.....	45
任务 1。启动 Informatica Developer.....	46
任务 2。添加域.....	46
任务 3。添加模型存储库.....	47
任务 4。创建项目.....	47
任务 5。创建文件夹.....	47
任务 6。选择默认数据集成服务.....	48
设置 Informatica Developer 课程总结.....	48
第 11 章： 课程 2： 导入物理数据对象.....	49
导入物理数据对象概览.....	49
任务 1。导入 Boston_Customers 平面文件数据对象.....	50
任务 2。导入 LA_Customers 平面文件数据对象.....	56
任务 3。导入 All_Customers 平面文件数据对象.....	56
导入物理数据对象课程总结.....	57
第 12 章： 课程 3。对源数据运行配置文件.....	58
剖析数据概览.....	58
任务 1。对两个数据源执行联接分析.....	59
任务 2。查看联接分析结果.....	60
任务 3。对数据源运行配置文件.....	60
任务 4。查看列剖析结果.....	60
剖析数据课程总结.....	61
第 13 章： 课程 4。解析数据.....	62
解析数据概览.....	62
任务 1。创建目标数据对象.....	63
步骤 1。创建 LA_Customers_tgt 数据对象.....	63
步骤 2。配置读取和写入选项.....	63
步骤 3。将列添加到数据对象.....	64
任务 2。创建映射以解析数据.....	64
步骤 1。创建映射.....	65
步骤 2。将数据对象添加到映射.....	65
步骤 3。将解析器转换添加到映射.....	65
步骤 4。配置解析器转换.....	66
任务 3。对解析器转换运行配置文件.....	66
任务 4。运行映射.....	66
任务 5。查看映射输出.....	67
解析数据课程总结.....	67

第 14 章： 课程 5。标准化数据	68
数据标准化概览.....	68
任务 1。创建目标数据对象.....	69
步骤 1。创建 All_Customers_Stdz_tgt 数据对象.....	69
步骤 2。配置读取和写入选项.....	69
任务 2。创建映射以将数据标准化.....	70
步骤 1。创建映射.....	70
步骤 2。将数据对象添加到映射.....	70
步骤 3。将标准创建器转换添加到映射.....	71
步骤 4。配置标准创建器转换.....	71
任务 3。运行映射.....	72
任务 4。查看映射输出.....	72
数据标准化课程总结.....	73
第 15 章： 课程 6。验证地址数据.....	74
验证地址数据概览.....	74
任务 1。创建目标数据对象	75
步骤 1。创建 All_Customers_av_tgt 数据对象.....	75
步骤 2。配置读取和写入选项.....	76
步骤 3。将端口添加到数据对象.....	76
任务 2。创建映射以验证地址.....	77
步骤 1。创建映射.....	77
步骤 2。将数据对象添加到映射.....	77
步骤 3。将地址验证器转换添加到映射.....	77
任务 3。配置地址验证器转换.....	78
步骤 1。设置用于地址验证“默认国家/地区”.....	78
步骤 2。配置地址验证器转换输入端口.....	78
步骤 3。配置地址验证器转换的输出端口.....	79
步骤 4。将不使用的数据源端口连接到数据目标.....	80
任务 4。运行映射.....	80
任务 5。查看映射输出.....	81
验证地址数据课程总结.....	83
附录 A： 常见问题.....	84
Informatica Analyst 常见问题解答.....	84
Informatica Developer 常见问题.....	84
索引.....	86

前言

《Data Quality 入门指南》的目标读者是数据质量开发人员和分析师。该指南提供教程来帮助初次使用软件的用户学习如何使用 Informatica Developer 和 Informatica Analyst。本指南假定您已了解数据质量概念、平面文件和关系数据库概念以及您环境中的数据库引擎。

Informatica 资源

Informatica Network

Informatica Network 囊括了 Informatica 全球客户支持部门、Informatica 知识库和其他产品资源。要访问 Informatica Network，请访问 <https://network.informatica.com>。

成员可以执行以下操作：

- 在一个位置访问您的所有 Informatica 资源。
- 在知识库中搜索文档、常见问题和最佳实践等产品资源。
- 查看产品可用性信息。
- 查看支持案例。
- 查找当地的 Informatica 用户组网络并与您的伙伴进行协作。

Informatica 知识库

使用 Informatica 知识库可在 Informatica Network 中搜索文档、入门知识文章、最佳实践和 PAM 等产品资源。

要访问知识库，请访问 <https://kb.informatica.com>。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。

Informatica 文档

要获取有关产品的最新文档，请浏览 Informatica 知识库，网址为 https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx。

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。

Informatica 产品可用性矩阵

产品可用性矩阵 (PAM) 指明了产品版本支持的操作系统版本、数据库以及其他类型的数据源和目标。如果您是 Informatica Network 成员, 您可以访问 PAM, 网址为

<https://network.informatica.com/community/informatica-network/product-availability-matrices>。

Informatica Velocity

Informatica Velocity 收集了 Informatica 专业服务开发的一系列提示和最佳实践。Informatica Velocity 基于数百计的数据管理项目的实际经验而开发, 汇集了我们曾在世界各地组织就职的顾问在成功规划、开发、部署和维护数据管理解决方案方面的知识。

如果您是 Informatica Network 成员, 您可以访问 Informatica Velocity 资源, 网址为

<http://velocity.informatica.com>。

如果您对 Informatica Velocity 有任何疑问、意见或建议, 请通过 ips@informatica.com 与 Informatica 专业服务联系。

Informatica Marketplace

Informatica Marketplace 是一个论坛, 该论坛中提供的解决方案可补充、扩展或增强您的 Informatica 实现。您可以利用 Informatica 开发人员和合作伙伴提供的数百计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 <https://marketplace.informatica.com>。

Informatica 全球客户支持部门

您可以通过电话或 Informatica Network 上的联机支持与全球支持中心联系。

要查找您当地的 Informatica 全球客户支持部门电话号码, 请访问 Informatica 网站, 链接为:

<http://www.informatica.com/us/services-and-training/support-services/global-support-centers>。

如果您是 Informatica Network 成员, 您可以使用联机支持, 网址为 <http://network.informatica.com>。

第 1 章

入门概览

本章包括以下主题：

- [Informatica 域概览, 10](#)
- [介绍 Informatica Analyst, 12](#)
- [Informatica Developer 概览, 13](#)
- [教程讲解, 15](#)
- [教程结构, 15](#)

Informatica 域概览

Informatica 采用服务型体系结构，该结构具有扩展服务和在多个计算机之间共享资源的功能。Informatica 域是进行服务管理的主单元。

安装 Informatica 之后，您可以登录 Informatica Administrator。首先使用 Administrator 工具管理域和配置必要应用程序服务，然后才能访问剩下的应用程序客户端。

Informatica 域包含以下组件：

- 应用程序客户端。用来访问基本 Informatica 功能的一组客户端。应用程序客户端创建服务管理器请求或应用程序服务请求。
- 应用程序服务。代表基于服务器的功能的一组服务。Informatica 域可以包含应用程序服务的子集。您要创建和配置应用程序客户端需要的应用程序服务。
应用程序服务包括可在域中包含单个实例的系统服务。创建域时，将为您创建系统服务。您可以配置一个系统服务，然后启用该服务，以便使用该服务所提供的功能。
- 配置文件仓库。数据集成服务用于存储配置文件结果的关系数据库。
- 引用数据仓库。用来存储模型存储库中引用表对象的引用数据值的关系数据库。
- 存储库。存储处理来自应用程序客户端的用户请求时所需对象和进程的相关元数据的一组关系数据库。
- 服务管理器。内置在域中用来管理所有域操作的服务。服务管理器运行应用程序服务并执行域功能，包括身份验证、授权和日志记录。
- workflow 数据库。用来存储工作流的运行时元数据的关系数据库。

下表列出了应用程序客户端（不含 Administrator 工具）以及客户端所需的应用程序服务和存储库：

应用程序客户端	应用程序服务	存储库
Informatica Analyst	<ul style="list-style-type: none"> - 分析服务 - 内容管理服务 - 数据集成服务 - 模型存储库服务 - 搜索服务 	模型存储库
Informatica Developer	<ul style="list-style-type: none"> - 分析服务 - 内容管理服务 - 数据集成服务 - 模型存储库服务 	模型存储库
Metadata Manager	<ul style="list-style-type: none"> - Metadata Manager 服务 - PowerCenter 集成服务 - PowerCenter 存储库服务 	<ul style="list-style-type: none"> - Metadata Manager 存储库 - PowerCenter 存储库
PowerCenter 客户端	<ul style="list-style-type: none"> - PowerCenter 集成服务 - PowerCenter 存储库服务 	PowerCenter 存储库
Web 服务中心控制台	<ul style="list-style-type: none"> - PowerCenter 集成服务 - PowerCenter 存储库服务 - Web 服务中心 	PowerCenter 存储库

以下应用程序服务不属于 Informatica 应用程序客户端的访问范围：

- PowerExchange 侦听器服务。管理 PowerExchange 侦听器以移动海量数据并捕获变更数据。PowerCenter 集成服务可通过许可证服务连接到 PowerExchange 侦听器。
- PowerExchange 日志记录器服务。管理适用于 Linux、UNIX 以及 Windows 的 PowerExchange 日志记录器，以捕获变更数据并将其写入到 PowerExchange 日志记录器日志文件中。变更数据可能源自 DB2 恢复日志、Oracle 恢复日志、Microsoft SQL Server 分布数据库或 i5/OS 或 z/OS 系统上的数据源。
- SAP BW 服务。侦听来自 SAP BI 的 RFC 请求和 PowerCenter 集成服务运行工作流以从 SAP BI 提取或加载至 SAP BI 的请求。

功能的可用性

Informatica 产品使用一组通用的应用程序。您可以使用的产品功能取决于您的产品许可证。

下表说明了各种许可选项以及每个选项可用的应用程序功能：

许可选项	Informatica Developer 功能	Informatica Analyst 功能
Data Quality	<ul style="list-style-type: none">- 创建和运行映射- 创建和运行规则- 执行剖析，包括使用企业发现配置文件和发现主键、外键和功能相关性。- 管理推导得出的配置文件结果- 使用结果卡- 将对象导出到 PowerCenter	<ul style="list-style-type: none">- 执行剖析，包括企业发现- 使用发现搜索功能查找数据和元数据在剖析存储库中的位置- 使用结果卡- 管理推导得出的配置文件结果- 创建和运行剖析规则- 引用表管理- 创建规则规范并将规则规范编译为 Mapplet- 离散记录和重复记录管理
Data Services	<ul style="list-style-type: none">- 创建逻辑数据对象模型- 创建并运行具有 Data Services 转换的映射- 创建 SQL 数据服务- 创建 Web 服务- 将对象导出到 PowerCenter	<ul style="list-style-type: none">- 引用表管理
Data Services 和剖析选项	<ul style="list-style-type: none">- 创建逻辑数据对象模型- 创建并运行具有 Data Services 转换的映射- 创建 SQL 数据服务- 创建 Web 服务- 将对象导出到 PowerCenter- 创建和运行具有 Data Services 转换的规则- 剖析	<ul style="list-style-type: none">- 引用表管理

介绍 Informatica Analyst

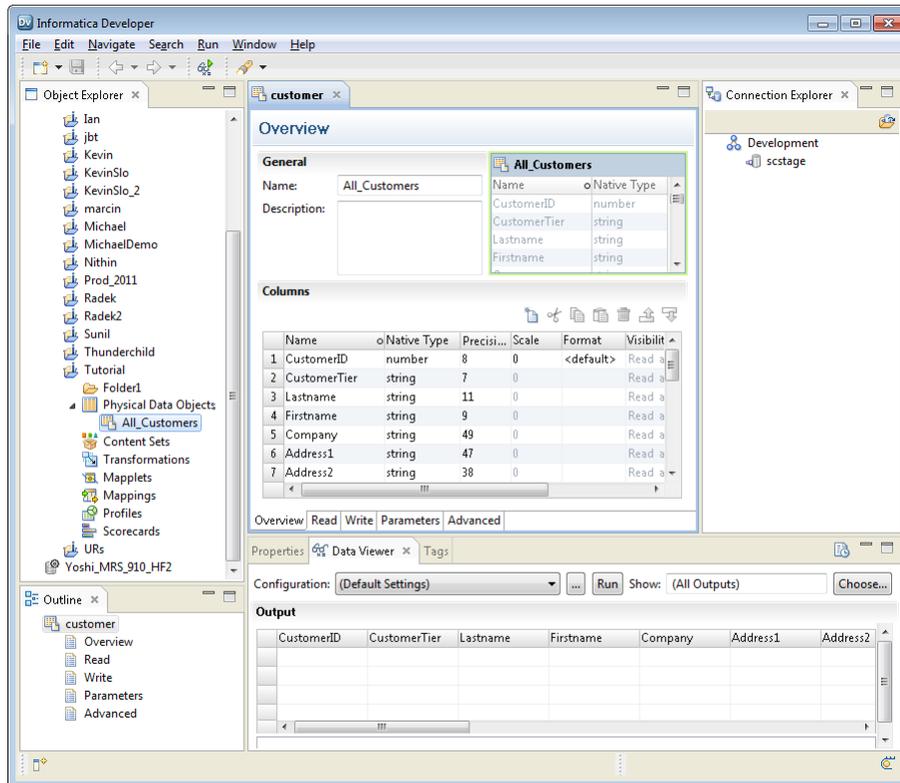
Informatica Analyst 是一个基于 Web 的应用程序客户端，分析人员可使用该客户端分析、清理、标准化、剖析企业内的数据，并为数据评分。

根据拥有的许可证，业务分析人员和开发人员可以使用 Analyst 工具进行数据驱动协作。您可以执行列和规则剖析、结果卡、离散记录和重复记录管理。还可以在数据质量解决方案中管理引用数据并将数据提供给开发人员。

Informatica Developer 概览

Informatica Developer 是一个供开发人员设计和实施数据解决方案使用的应用程序客户端。

下图显示了 Developer 工具：



Developer 工具可以显示以下视图：

“对象浏览器”视图

显示项目、文件夹以及其中的对象。

“连接浏览器”视图

显示与关系数据库的连接。

“大纲”视图

显示与对象浏览器视图中选定对象相关的对象。

“备忘单”视图

显示打开的备忘单。要打开备忘单，请单击**帮助** > **备忘单**，然后选择一个备忘单。

“数据查看器”视图

显示源数据、配置文件结果，以及预览转换的输出。

对象相关性视图

查看、修改或删除对象时显示对象相关性。

“警告”视图

显示连接状态警告。

“验证日志”视图

显示对象验证错误。

Developer 工具还可以显示其他视图。您可以隐藏视图和将视图移动到 Developer 工具工作台中的其他位置。单击 **窗口** > **显示视图** 选择要显示的视图。

Informatica Developer 欢迎页

首次打开 Developer tool 时，会显示欢迎页。使用欢迎页了解 Developer tool 的详细信息、设置 Developer tool 并开始使用 Developer tool。

欢迎页会显示以下选项：

- 概览。单击“概览”按钮可获取数据质量和数据服务解决方案概览。
- 前几步。单击“前几步”按钮可了解有关设置 Developer tool 以及访问 Informatica Data Quality 和 Informatica Data Services 课程的详细信息。
- 教程。单击“教程”按钮查看针对数据质量和数据服务解决方案的教程课程。
- Web 资源。单击“Web 资源”按钮获得指向 mysupport.informatica.com 的链接，从中可以访问 Informatica 入门知识库。Informatica 入门知识库包含有关 Informatica Data Quality、Informatica Data Services 及其他 Informatica 产品的文章。
- 新增内容。单击“新增内容”按钮可查看 Developer tool 的最新功能。

备忘单

Developer 工具将备忘单包含在联机帮助中。备忘单是可帮助您在 Developer 工具中完成一个或多个任务的分步指南。

完成备忘单后，即完成任务并可查看结果。例如，完成备忘单以导入并预览关系数据对象后，就已导入关系数据库表并在 Developer 工具中预览数据。

要访问备忘单，请单击 **帮助** > **备忘单**。

数据质量和剖析

使用 Developer 工具中的数据质量功能可分析数据的内容和结构。您可以采用符合业务需求的方式来改善数据。

使用 Developer 工具可设计并运行实现以下目标的过程：

- 剖析数据。剖析可揭示数据的内容和结构。剖析是任何数据项目中的关键一步，因为它可以识别数据中的优势和劣势，并帮助您制定项目计划。
- 创建结果卡以查看数据质量。结果卡是配置文件中对质量度量的一种图形表示形式。
- 标准化数据值。对数据进行标准化，以消除您在运行配置文件时发现的错误和不一致。您可以对标点、格式和拼写方面的变体进行标准化。例如，您可以确保城市、省/自治区/直辖市和邮政编码值一致。
- 解析记录。对数据记录进行解析，以改进记录结构并从数据中推导其他信息。您可以将一个自由格式数据字段拆分为多个包含不同信息类型的字段。您也可以将信息添加到记录中。例如，您可以将客户记录标记为个人客户或企业客户。
- 验证通信地址。地址验证可评估并提高通信地址数据的准确性及可送达性。地址验证会通过比较地址记录与国家邮政运营商提供的引用数据，更正地址中的错误并补充不完整的地址。地址验证还可以添加邮政信息，从而加快邮件送达并降低邮件成本。

- 查找重复记录。重复记录分析可对一组记录进行相互比较，找出选定的数据列中相似或一致的值。您可以设置一个表示字段值之间存在较好匹配的相似性级别。您还可以设置匹配计算中给予每个列的相对权重。例如，您可以设置姓氏信息的优先级高于名字信息。
- 创建和运行数据质量规则。Informatica 提供预设的规则，您可以运行或编辑这些规则来符合项目目标要求。您可以在 Developer 工具中创建这些规则。
- 与 Informatica 用户协作。您添加到模型存储库的规则和引用数据表可供用户在 Developer 工具和 Analyst 工具中使用。用户可以协作完成项目，并且不同用户可以在项目的不同阶段获得对象的所有权。
- 将映射导出到 PowerCenter。您可以将映射导出到 PowerCenter，以重用元数据来进行物理数据集成或创建 Web 服务。

教程讲解

HypoStores 公司是全国零售机构，总部设在波士顿，在多个州拥有分店。它定期将全国范围的分店业务数据与总部数据存储进行整合。最近在洛杉矶开设了一家分店。

总部包括一个中心 ICC 小组，由管理人员、开发人员和架构师构成，负责为所有复合和 BI 应用程序提供一个通用数据服务层。BI 应用程序包括一个 CRM 系统，包含用于结算和营销的主客户数据文件。

HypoStores Corporation 必须执行以下任务将 Los Angeles 分公司的数据与 Boston 总部的数据整合起来。

- 检查 Boston 和 Los Angeles 的数据是否存在数据质量问题。
- 从 Los Angeles 的数据中解析信息。
- 将 Boston 和 Los Angeles 的数据中的地址信息标准化。
- 验证数据中的通信地址信息的准确性，以用于客户关系管理 (CRM)。

教程结构

*入门指南*包含包括课程和任务的教程。

课程

每个课程都介绍了将帮助您了解要在该课程中完成的任务的概念。本课程从总体上介绍了业务要求。本课程的目标概述了您将完成以满足业务要求的任务。每个课程都提供了估计的完成时间。完成本课程中的任务后，可以查看课程摘要。

如果未配置工具内的环境，则每个教程中的第一课将帮助您进行配置。

任务

这些任务提供逐步说明。按完成本课程所列的顺序完成所有任务。

Informatica Analyst 教程

在本教程中，分析人员将登录到 Analyst 工具、创建项目和文件夹、创建配置文件和规则、对数据进行评分并创建引用表。

您可以进行的课程取决于您拥有 Informatica Data Quality 还是 Informatica Data Services 产品。

下表说明了您可以进行的课程（具体取决于您的产品）：

课程	说明	产品
课程 1. 设置 Informatica Analyst	登录到 Analyst 工具并为教程课程创建一个项目和文件夹。	Data Quality Data Services
课程 2. 创建数据对象	导入平面文件作为数据对象并预览数据。	Data Quality
课程 3. 创建快速配置文件	创建快速配置文件以快速了解数据质量。	Data Quality
课程 4. 创建自定义配置文件	创建自定义配置文件以配置列、采样和向下钻取选项。	Data Quality
课程 5. 创建表达式规则	创建表达式规则以修改和剖析列值。	Data Quality
课程 6. 创建并运行结果卡	创建并运行结果卡以衡量一段时间的数据质量进度。	Data Quality
课程 7. 使用配置文件结果创建引用表	创建可用于将源数据标准化的引用表。	Data Quality Data Services
课程 8. 创建引用表	创建引用表在源数据与有效的标准值之间建立关系。	Data Quality Data Services

Informatica Developer Tool

在此教程中，您将使用 Developer tool 执行几项数据质量操作。

Informatica Data Quality 用户使用 Developer tool 设计和运行用于改善数据质量的过程。Informatica Data Quality 用户也使用 Developer tool 来创建和运行分析数据内容和结构的配置文件。

您将完成数据质量教程中的以下课程：

课程 1. 设置 Informatica Developer

创建与某个域的模型存储库服务所管理的模型存储库的连接。创建项目和文件夹来存储教程课程的作业内容。选择默认的数据集成服务。

课程 2. 导入物理数据对象

您将为与这些对象相关联的客户数据文件定义数据质量过程。

课程 3. 剖析数据

剖析可揭示数据的内容和结构。

剖析包括联接分析，这种分析可判断两个数据列之间是否可能建立有效联接。

课程 4. 解析数据

解析可以丰富数据记录并改进记录结构。它可以查找数据中的有用信息，还可以从当前数据中推导出新的信息。

课程 5. 标准化数据

标准化可去除剖析过程中发现的数据错误和不一致问题。

课程 6. 验证地址数据

地址验证可评估通信地址的准确性和可送达性，并修复地址中的错误和遗漏。

教程先决条件

在您可以开始教程课程之前，Informatica 域必须运行并至少设置一个节点。

安装程序包含完成课程要使用的教程文件。您在客户端和服务端安装中均可找到所有文件：

- 您可在 Developer 工具安装路径中的以下位置找到教程文件：

<Informatica 安装目录>\clients\DeveloperClient\Tutorials

- 您可在服务安装路径中的以下位置找到教程文件：

<Informatica 安装目录>\server\Tutorials

您需要以下文件来完成教程课程：

- All_Customers.csv
- Boston_Customers.csv
- LA_customers.csv

第 I 部分： Informatica Analyst 入门

本部分包含以下章节：

- [课程 1。设置 Informatica Analyst, 19](#)
- [课程 2。创建数据对象, 22](#)
- [课程 3。创建默认配置文件, 25](#)
- [课程 4。创建自定义配置文件, 28](#)
- [课程 5。创建表达式规则, 31](#)
- [课程 6。创建并运行结果卡, 34](#)
- [课程 7。使用配置文件列创建引用表, 39](#)
- [课程 8。创建引用表, 42](#)

第 2 章

课程 1。设置 Informatica Analyst

本章包括以下主题：

- [设置 Informatica Analyst 概览, 19](#)
- [任务 1. 登录 Informatica Analyst, 20](#)
- [任务 2. 创建项目, 20](#)
- [任务 3. 创建文件夹, 20](#)
- [设置 Informatica Analyst 摘要, 21](#)

设置 Informatica Analyst 概览

在开始本教程中的课程之前，您必须先设置 Analyst 工具。要设置 Analyst 工具，请登录到 Analyst 工具并创建一个项目和文件夹来存储工作。

Informatica 域是可定义 Informatica 环境的节点和服务的集合。域中的服务包括分析服务和模型存储库服务。分析服务运行 Analyst 工具，模型存储库服务管理模型存储库。在 Analyst 工具中工作时，Analyst 工具会存储您在模型存储库中创建的资产。

您必须先创建一个项目，然后才能在 Analyst 工具中创建资产。项目包含 Analyst 工具中的资产。项目还可以包含存储相关资产的文件夹，例如属于同一业务要求的数据对象。

目标

在本课程中，您将完成以下任务：

- 登录到 Analyst 工具。
- 创建一个项目，用于存储在 Analyst 工具中创建的资产。
- 在项目中创建一个可存储相关资产的文件夹。

先决条件

开始本课程之前，请验证以下先决条件：

- 管理员已在 Administrator 工具中配置模型存储库服务和分析服务。
- 您有 Analyst 工具的主机名和端口号。
- 拥有用户名和密码才能访问分析服务。可以通过管理员获取此信息。

计时

完成本课程需要 5-10 分钟。

任务 1. 登录 Informatica Analyst

登录到 Analyst 工具，开始学习本教程。

1. 启动 Microsoft Internet Explorer 或 Google Chrome 浏览器。
2. 在“地址”字段中，输入 Informatica Analyst 的 URL：
`http[s]://<fully qualified host name>:<port number>/analyst`
3. 如果该域使用 LDAP 或本地身份验证，请在登录页面上输入您的用户名和密码。
4. 选择**本地**或特定安全域的名称。

如果 Informatica 域使用 LDAP 或 Kerberos 身份验证，将显示“安全域”字段。如果您不知道用户帐户所属的安全域，请联系 Informatica 域管理员。

5. 单击**登录**。

Analyst 工具会在**启动**工作区上打开。

任务 2. 创建项目

在本任务中，创建项目以包含在 Analyst 工具中创建的资产。创建教程项目以包含项目的文件夹。

1. 在**管理**表头上，单击**项目**。
将显示**项目**工作区。
2. 从**操作**菜单中，单击**新建** > **项目**。
此时将显示**新建项目**窗口。
3. 输入加前缀“Tutorial_”的名称作为项目的名称。
4. 单击**确定**。

任务 3. 创建文件夹

在本任务中，创建存储相关资产的文件夹。您可以在项目或在其他文件夹中创建文件夹。创建名为 Customers 的文件夹，以存储与数据质量项目相关的资产。

1. 在**项目**面板中，选择教程项目。
2. 从**操作**菜单中，单击**新建** > **文件夹**。
此时将显示**新建文件夹**窗口。
3. 输入 Customers 作为文件夹名称。
4. 单击**确定**。

该文件夹会显示在教程项目下。

设置 Informatica Analyst 摘要

在本课程中，您学习到 Analyst 工具在项目和文件夹中存储资产。模型存储库包含项目和文件夹。分析服务运行 Analyst 工具。模型存储库服务用于管理模型存储库。分析服务和模型存储库服务是 Informatica 域中的应用程序服务。

已登录到 Analyst 工具并创建了项目和文件夹。

现在，可以使用 Analyst 工具完成本教程中的其他课程。

第 3 章

课程 2。创建数据对象

本章包括以下主题：

- [创建数据对象概览, 22](#)
- [任务 1。创建平面文件数据对象, 23](#)
- [任务 2. 查看数据对象属性, 23](#)
- [创建数据对象摘要, 24](#)

创建数据对象概览

在 Analyst 工具中，数据对象是基于平面文件或关系数据库表的数据表示。创建平面文件或表对象，然后对照平面文件或关系数据库表中的数据运行配置文件。当您在 Analyst 工具中创建平面文件数据对象时，可以将文件上载到运行 Analyst 工具的计算机上的平面文件缓存中，或者可以指定存储平面文件的网络位置。

故事

HypoStores 会将洛杉矶客户数据保留在平面文件中。HypoStores 需要剖析并分析数据以及执行数据质量任务。

目标

在本课程中，您将完成以下任务：

1. 将平面文件上传到平面文件缓存位置，并创建数据对象。
2. 预览平面文件数据对象的数据。

先决条件

开始本课程之前，请验证以下先决条件：

- 已完成本课程中的课程 1。
- 您有 LA_Customers.csv 平面文件。可以在 <安装根目录>\<发行版本>\clients\DeveloperClient\Tutorials 文件夹中查找该文件。

计时

完成本任务需要 5-10 分钟。

任务 1。创建平面文件数据对象

在此任务中，您将通过 LA_Customers 文件创建平面文件数据对象。

1. 在 Analyst 工具中，单击**新建** > **平面文件数据对象**。
此时将显示**添加平面文件**向导。
2. 选择**浏览并上载**，然后单击**浏览**。
3. 浏览到 LA_Customers.csv 所在的位置，然后单击**打开**。
4. 单击**下一步**。
选择导入类型面板将显示**带分隔符**和**固定宽度**选项。选择**带分隔符**选项。默认选项为**带分隔符**。
5. 单击**下一步**。
6. 在**指定在数据中使用的分隔符和文本限定符**下，选择**双引号**作为文本限定符。
7. 在**指定要导入的行**下，选择**从第一行导入**以便从第一个非空行导入列名。
预览面板将更新并显示第一行的列标题。
8. 单击**下一步**。
列属性面板显示每一列的数据类型、精度、小数位数和格式。
9. 单击**下一步**。
名称字段将显示 LA_Customers。
10. 选择 **Tutorial_** 项目和 **Customers** 文件夹。
11. 单击**完成**。
该数据对象将显示在 Customers 文件夹的文件夹内容中。

任务 2。查看数据对象属性

在本任务中，您可以查看 LA_Customers 数据对象的属性。

1. 单击**打开**以打开**库**工作区。
2. 在**库**工作区中，单击**资产**面板中的**数据对象**。
数据对象的列表将显示在**数据对象**面板中。
3. 单击 **LA_Customers** 平面文件。
此时将显示**数据预览**面板，其中包含从 LA_Customers 数据对象检索到的数据。
4. 在**数据预览**面板中，查看 LA_Customers 数据对象的结构和内容。
Analyst 工具显示平面文件数据对象的前 100 行。
5. 单击**属性**。
属性面板显示数据对象的名称、类型、说明和位置。您还可以查看数据对象的列名称和列属性。

创建数据对象摘要

在本课程中，您学习到数据对象是基于平面文件或关系数据库源的数据的表示形式。您学习到，可以创建平面文件数据对象并预览其中的数据。

您上载了一个平面文件并创建了一个平面文件数据对象，预览了该数据对象的数据，并查看了该数据对象的属性。

创建数据对象后，在课程 3 中为该数据对象创建默认配置文件，并在课程 4 中为该数据对象创建自定义配置文件。

第 4 章

课程 3. 创建默认配置文件

本章包括以下主题：

- [创建默认配置文件概览, 25](#)
- [任务 1. 创建并运行默认配置文件, 26](#)
- [任务 2. 在摘要视图中查看配置文件结果, 26](#)
- [创建默认配置文件摘要, 27](#)

创建默认配置文件概览

配置文件是基于数据内容和结构的数据质量分析。默认配置文件是使用默认选项创建的配置文件。默认配置文件用于获取配置文件结果，而不必为配置文件配置所有列和选项。

在启动数据质量项目时，创建并运行默认配置文件以分析数据的质量。创建默认配置文件对象时，选择您想要分析的数据对象和数据对象列。默认配置文件会跳过配置文件列和选项配置。Analyst 工具会剖析平面文件数据对象的实时平面文件。

故事

HypoStores 想要将来自新收购的洛杉矶办公室的数据合并到其数据仓库中。必须先对数据仓库进行清理，数据才能合并到其中。您是分析人员，负责评估数据质量并将信息传递给负责清理数据的开发人员。您想要快速查看配置文件结果并大致了解数据质量。

目标

在本课程中，您将完成以下任务：

1. 为 LA_Customers 平面文件数据对象创建并运行默认配置文件。
2. 查看配置文件结果。

先决条件

开始本课程之前，请确认满足以下先决条件：

- 已完成本课程中的课程 1 和 2。

计时

完成本课程需要 5-10 分钟。

任务 1. 创建并运行默认配置文件

在本任务中，为数据对象中的所有列创建默认配置文件，并使用默认采样和向下钻取选项。

1. 在**库**工作区中，在**资产**面板中选择一个数据对象。
2. 右键单击数据对象并选择**创建配置文件**。
此时将显示**新建配置文件**向导。
3. 默认情况下将选择**单源**选项。单击**下一步**。
4. 在**指定常规属性**屏幕中，默认情况下会填充名称、说明和位置。
5. 单击**下一步**。
6. 在**选择源**屏幕中，默认情况下会显示数据对象。您可以在**选择列**窗格中查看列。
7. 单击**下一步**。
8. 在**指定设置**屏幕中，默认情况下选择以下选项：
 - **运行列配置文件**
 - **运行配置文件**窗格中的**所有行**
 - **向下钻取**窗格中的**实时**
 - **在后续运行配置文件时，从数据类型和数据域推理中排除已批准的数据类型和数据域。**
 - **运行时环境**窗格中的**本地连接**。
9. 单击**下一步**。
10. 在**指定规则和筛选器**屏幕中，单击**保存并运行**以创建并运行配置文件。
Analyst 工具会创建并运行配置文件。配置文件结果将显示在摘要视图中。

任务 2. 在摘要视图中查看配置文件结果

在本任务中，使用 LA_Customers 配置文件的摘要视图可快速了解配置文件结果。

1. 在**库 > 资产 > 配置文件**窗格中，单击 **LA_Customers** 配置文件。
配置文件结果将显示在摘要视图中。
2. 在摘要视图中，单击**筛选依据**窗格中的**列**，以查看列的配置文件结果。
3. 将指针移到水平条形图上方，可查看百分比形式的值。
4. 在**空值、相异值和非相异值百分比**部分，可查看列中空值、相异值和非相异值的百分比。
5. 在**模式**部分，可通过水平条形图查看列中的多个模式。将指针移到条形图上方时，可查看列中的模式字符和相似模式数的百分比。
6. 在**长度**部分，您可以查看列中值的最小和最大长度。
7. 在**值**部分，您可以查看列中的最小值和最大值。
8. 在**数据类型**部分，将指针移到值上方时，可以查看列的所有推理的数据类型和已记录的数据类型。
9. 在**数据域**部分，将指针移到值上方时，可以查看列的所有推理的数据域。
10. 要查看离群值数据，请单击**操作 > 检测离群值**来检测配置文件结果中的离群值。
11. 单击**模式离群值**或**值频率离群值**筛选器，以查看配置文件结果中的离群值。
12. 单击列名称，在详细视图中查看此列的配置文件结果。

创建默认配置文件摘要

在本课程中，您学习了以下内容：默认配置文件会显示配置文件结果，而不必为配置文件配置所有列和行采样选项；在启动数据质量项目时，创建并运行默认配置文件以分析数据的质量；Analyst 工具会剖析平面文件数据对象的实时平面文件。

您已创建默认配置文件并已分析配置文件结果，并且进一步了解了配置文件中列的相关信息，包括空值和数据类型。还使用了列值和模式来识别数据质量问题。

在您分析快速配置文件的結果后，可以执行以下任务：

- 创建自定义配置文件，以将列从配置文件中排除并仅包含需要的列。
- 创建表达式规则以创建虚拟列并对其进行剖析。
- 创建引用表以包括列的有效值。

第 5 章

课程 4。创建自定义配置文件

本章包括以下主题：

- [创建自定义配置文件概览, 28](#)
- [任务 1. 创建自定义配置文件, 29](#)
- [任务 2. 运行配置文件, 29](#)
- [任务 3. 向下钻取配置文件结果, 30](#)
- [创建自定义配置文件课程总结, 30](#)

创建自定义配置文件概览

配置文件是基于数据内容和结构的数据质量分析。自定义配置文件是您想要配置列、采样选项和向下钻取选项以更快进行剖析时创建的配置文件。配置采样选项可在源中选择采样行。配置向下钻取选项可向下钻取源数据或暂存数据中的数据行。您可以选择在 Hive 或本机环境中运行配置文件。

您可以创建并运行配置文件以在启动数据质量项目时分析数据的质量。在创建配置文件对象时，您首先要选择希望对其运行配置文件的数据对象和数据对象列。

故事

HypoStores 需要将来自新收购的洛杉矶办事处的数据合并到其数据仓库中。HypoStores 想要访问 LA 客户数据文件中客户层数据的质量。您是分析人员，负责评估数据质量并将信息传递给负责清理数据的开发人员。

目标

在本课程中，您将完成以下任务：

1. 为平面文件数据对象创建自定义配置文件，并排除值为空的列。
2. 运行配置文件以分析 CustomerTier 列的内容和结构。
3. 向下钻取配置文件结果的行。

先决条件

开始本课程之前，请确认满足以下先决条件：

- 已完成本教程中的课程 1、2 和 3。

计时

完成本课程需要 5-10 分钟。

任务 1. 创建自定义配置文件

在本任务中，您将创建一个自定义配置文件。创建自定义配置文件时，选择要对其运行配置文件的数据对象和列。您还可以配置采样和向下钻取选项。

1. 单击**新建** > **配置文件**。
此时将显示**新建配置文件**向导。
2. 默认情况下将选择**单源**选项。单击**下一步**。
3. 在**指定常规属性**屏幕上，设置以下选项：
 - 在“名称”字段中，输入 **Profile_LA_Customers**。
 - 在“位置”字段中，选择 **Customers** 文件夹。
4. 单击**下一步**。
5. 在**选择源**屏幕中，单击**选择**。
此时将显示**选择数据对象**对话框。
6. 在**选择数据对象**对话框中，选择 **LA_Customers**。单击**确定**。
7. 在**选择源**屏幕中，清除“地址 2”、“地址 3”和“城市 2”列。
8. 单击**下一步**。
9. 在**指定设置**屏幕中，设置以下选项：
 - **运行列配置文件**。
 - 在**运行配置文件**窗格中选择**随机采样**选项。
 - 选择在**后续运行配置文件时**，从**数据类型和数据域推理**中排除**已批准的数据类型和数据域**选项。
 - 在**向下钻取**窗格中选择**暂存**选项。
10. 单击**下一步**。
11. 在**指定规则和筛选器**屏幕中，单击**保存并完成**以创建并运行配置文件。
Analyst 工具会创建配置文件，并在**发现**工作区中显示该配置文件。您需要运行配置文件以查看结果。

任务 2. 运行配置文件

在本任务中，运行配置文件以剖析数据对象并显示配置文件结果。Analyst 工具会剖析平面文件数据对象的暂存平面文件。

1. 验证您是否处于**发现**工作区中。
可以在工作区中查看配置文件 **Profile_LA_Customers_Custom**。
2. 在工作区中单击 **Profile_LA_Customers_Custom**。
3. 此时将显示配置文件屏幕，您可以在此屏幕中选择编辑配置文件或运行配置文件。单击**运行**。
4. 配置文件结果将显示在摘要视图中。

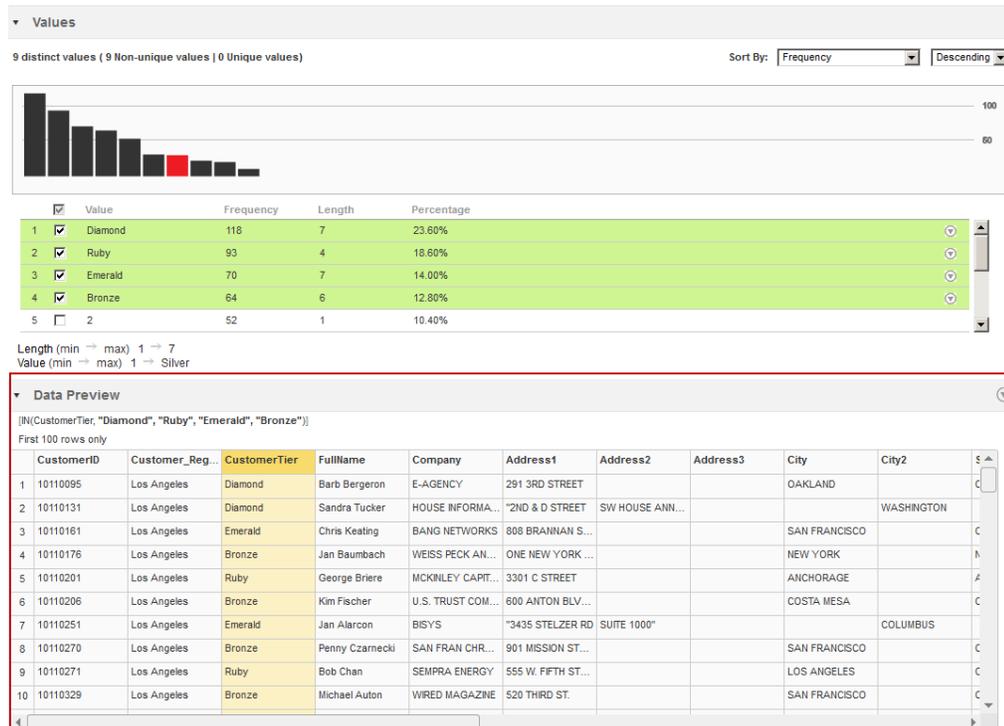
任务 3. 向下钻取配置文件结果

在本任务中，向下钻取 CustomerTier 列值，以查看配置文件的数据对象中的源行。

1. 确认已进入 **Profile_LA_Customers** 配置文件结果的摘要视图。
2. 单击 **CustomerTier** 列。
此列的配置文件结果将显示在详细视图中。
3. 在详细视图中，选择 Diamond、Ruby、Emerald 和 Bronze 值。右键单击值窗格中的值，然后选择**向下钻取**。

列表中值为 Diamond、Ruby、Emerald 或 Bronze 的行将显示在**数据预览**窗格中。

下图显示了向下钻取 Diamond、Ruby、Emerald 或 Bronze 值时，**数据预览**窗格中显示的向下钻取结果：



数据预览窗格显示选定列的前 100 行。**数据预览**窗格的标题显示源列所用的逻辑。

创建自定义配置文件课程总结

在本课程中您学习到，可以配置剖析的列，并可以配置采样和向下钻取选项。您学习到，可以向下钻取以查看列值的基本行，并可以在查看列值时配置包括的列。

您创建了包括 CustomerTier 列的自定义配置文件，运行了该配置文件，并向下钻取到结果中 CustomerTier 列的基本行。

在课程 5 中使用自定义配置文件对象创建表达式规则。

第 6 章

课程 5。创建表达式规则

本章包括以下主题：

- [创建表达式规则概览, 31](#)
- [任务 1. 创建表达式规则并运行配置文件, 32](#)
- [任务 2. 查看表达式规则的输出, 32](#)
- [任务 3. 编辑表达式规则, 33](#)
- [创建表达式规则课程总结, 33](#)

创建表达式规则概览

表达式规则使用表达式函数和源列定义规则逻辑。您可以创建表达式规则，并在 Analyst 工具中将其添加到配置文件中。一个表达式规则可以与一个或多个配置文件关联。

表达式规则的输出是配置文件中的虚拟列。Analyst 工具会在您运行配置文件时剖析虚拟列。

您可以使用表达式规则来验证源列，或根据这些源列的值创建其他源列。

故事

HypoStores 想要将来自新收购的洛杉矶办公室的数据合并到其数据仓库中。HypoStores 想要分析客户名称，并从客户名称中分离出名和姓。HypoStores 想要使用表达式规则将包含名和姓的列解析为单独的虚拟列，然后剖析这些列。HypoStores 还想要将这些规则设置为可供需要分析这些规则的其他分析人员使用。

目标

在本课程中，您将完成以下任务：

1. 创建表达式规则以将 FullName 列分离出“名”和“姓”列。创建可从全名中分离出名的规则。创建可从全名中分离出姓的另一规则。为 Profile_LA_Customers 配置文件创建这些规则。
2. 运行配置文件并查看配置文件中规则的输出。
3. 编辑这些规则以将其设置为可供其他 Analyst 工具用户使用。

先决条件

开始本课程之前，请确认满足以下先决条件：

- 已完成课程 1 到 4。

计时

完成本课程需要 10-15 分钟。

任务 1. 创建表达式规则并运行配置文件

在本任务中，创建两个表达式规则，以将 FullName 列解析为名为 FirstName 和 LastName 的两个虚拟列。规则名称为 FirstName 和 LastName。

1. 在库工作区中，单击 **Profile_LA_Customers** 配置文件。
配置文件结果将显示在摘要视图中。
2. 单击 **编辑**。
此时将显示 **配置文件** 向导。
3. 单击 **指定规则和筛选器** 屏幕。
4. 在 **规则** 窗格中，单击 **操作** > **创建规则**。
5. 在 **名称** 字段中，输入 **FirstName**。
6. 在 **表达式** 部分中，输入以下表达式以从 **名称** 列中分离出名字：

```
SUBSTR (FullName, 1, INSTR (FullName, ' ', -1, 1) - 1)
```
7. 单击 **验证**。
8. 单击 **确定**。
9. 重复第 4 步到第 8 步，以创建名为 **LastName** 的规则。输入以下表达式，将姓氏从 **名称** 列分隔出来：

```
SUBSTR (FullName, INSTR (FullName, ' ', -1, 1), LENGTH (FullName))
```
10. 单击 **保存并运行** 以创建并运行配置文件。

任务 2. 查看表达式规则的输出

在本任务中，查看在您运行配置文件后分离出名和姓的表达式规则的输出。

1. 在摘要视图中，单击 **编辑**。
此时将显示配置文件向导
2. 在配置文件向导中，单击 **选择源**。
3. 在 **选择源** 屏幕中，选中工具栏上 **名称** 旁边的复选框以清除所有列。
默认情况下会选择一列，因为您需要在 **列** 部分中至少选择一列。
4. 选择 **FullName** 列以及 **FirstName** 和 **LastName** 规则。
5. 清除选中的其他所有列。
6. 单击 **保存并运行**。
配置文件结果将显示在摘要视图中。
7. 单击 **FirstName** 规则，规则的配置文件结果将显示在详细视图中。
8. 在 **值** 窗格中选择一个值。右键单击此值，然后单击 **向下钻取**。
FullName 列以及 **FirstName** 和 **LastName** 规则的值以及其他列值显示在 **数据预览** 窗格中。请注意，Analyst 工具将 **FullName** 列分离为名字和姓氏。

任务 3. 编辑表达式规则

在本任务中，将表达式规则设置为可重用并可供所有 Analyst 工具用户使用。

1. 在 **Profile_LA_Customers** 配置文件的摘要视图中，单击**编辑**。
此时将显示配置文件向导。
2. 单击**指定规则和筛选器**屏幕。
3. 在**指定规则和筛选器**屏幕中，选择 **FirstName** 规则，然后单击**操作 > 编辑规则**。
此时将显示**编辑规则**对话框。
4. 选择**是否要保存此规则作为可重用规则?** 选项，然后选择要保存规则的位置。
5. 单击**确定**。
6. 选择 **LastName** 规则，然后重复执行第 3 步到第 5 步。
7. 单击**保存并完成**以保存配置文件。

任何 Analyst 工具用户都可以使用 **FirstName** 和 **LastName** 规则将具有姓氏和名字的列分为单独的列。

创建表达式规则课程总结

在本课程中，您学习到表达式规则使用表达式函数和源列来定义规则逻辑。您还了解到表达式规则的输出是配置文件中的虚拟列。Analyst 工具会在您运行配置文件时包括虚拟列。

您创建了两个表达式规则、将其添加到配置文件中并运行了该配置文件。您查看了规则的输出并将其设置为可供所有 Analyst 工具用户使用。

第 7 章

课程 6。创建并运行结果卡

本章包括以下主题：

- [创建并运行结果卡概览, 34](#)
- [任务 1. 使用配置文件结果创建结果卡, 35](#)
- [任务 2. 运行结果卡, 36](#)
- [任务 3. 查看结果卡, 36](#)
- [任务 4. 编辑结果卡, 37](#)
- [任务 5. 配置阈值, 37](#)
- [任务 6. 查看得分趋势图表, 37](#)
- [创建并运行结果卡课程总结, 38](#)

创建并运行结果卡概览

结果卡是配置文件结果中列的有效值或规则输出的图形表示形式。结果卡用于衡量和监控数据质量随时间的进度。

要创建结果卡，请将配置文件中的列作为度量添加到结果卡，为度量分配权重，并配置得分阈值。您可以将基于源数据的筛选器添加到结果卡。要运行结果卡，请为度量选择有效值并运行结果卡以查看度量的得分。

结果卡会将配置文件中各列的值频率显示为得分。得分反映度量的有效值的百分比。

故事

HypoStores 想要将来自新收购的洛杉矶办公室的数据合并到其数据仓库中。组织合并数据之前，他们想要验证不同客户层中具有不同状态的数据是否执行了数据质量分析。您是负责监视数据质量分析执行进度的分析人员，想要通过客户层和状态配置文件列创建结果卡、为数据质量配置阈值以及查看得分趋势图表，以确定得分如何随着时间的变化而提高。

目标

在本课程中，您将完成以下任务：

1. 通过 Profile_LA_Customers_Custom 配置文件的结果创建结果卡，以查看 CustomerTier 和“状态”列的得分。
2. 运行结果卡，为 CustomerTier 和“状态”列生成得分。
3. 查看结果卡，以查看每列的得分。
4. 编辑结果卡，为得分指定不同的有效值。
5. 配置得分阈值并运行结果卡。

6. 查看得分趋势图表，以确定今后如何改进得分。

先决条件

开始本课程之前，请确认满足以下先决条件：

- 已完成本课程中的课程 1 到 5。

计时

完成本课程中的任务需要 15 分钟。

任务 1. 使用配置文件结果创建结果卡

在本任务中，从 Profile_LA_Customers_Custom 配置文件创建结果卡，以为 CustomerTier 和“状态”列值评分。

1. 在库工作区中，单击 **Profile_LA_Customers** 配置文件。
此时会显示配置文件结果的摘要视图。
2. 在摘要视图中，选择 **CustomerTier** 列，然后右键单击该列，并选择**添加到** > **结果卡**。
此时将显示**添加到结果卡**向导。
3. 在**添加到结果卡**向导中，默认情况下会选择**新建结果卡**选项。单击**下一步**。
4. 在**第 2 步(共 8 步)** 屏幕上，输入 **sc_LA_Customer** 作为结果卡名称，然后导航到 **Customers** 文件夹作为结果卡位置。
5. 单击**下一步**。
6. 在**第 3 步(共 8 步)** 屏幕中，选择 **CustomerTier** 和**州**列，以将它们添加到结果卡。
7. 单击**下一步**。
8. 在**第 4 步(共 8 步)** 屏幕上，可以为度量创建、编辑或删除筛选器。在本教程中，将不创建结果卡筛选器。单击**下一步**。
9. 在**第 5 步(共 8 步)** 屏幕上，在**度量**窗格中选择 **CustomerTier** 度量。
10. 在**正在使用的得分：值**窗格中，选择所有值并单击**全部添加**按钮，将值移动到**有效值**部分。
使用 **Shift** 键可选择多个值。
11. 在**度量**窗格中，选择**州**度量，并在**正在使用的得分：值**部分选择两个字母的州代码值。
12. 单击**添加**按钮以将值移动到**有效值**部分。
您可以在该部分顶部查看有效值总数和有效值百分比。
13. 对于**度量**部分中的每个度量，接受**度量阈值**部分中的得分阈值的默认设置。
14. 单击**下一步**。
15. 在**第 6 步(共 8 步)** 屏幕中，可以选择一个度量组来添加度量。默认情况下，Analyst 工具将度量添加到**默认度量组**。
16. 单击**下一步**。
17. 在**第 7 步(共 8 步)** 屏幕中，在**默认 - 度量**窗格中双击 **CustomerTier** 度量的**权重**列。
当运行结果卡时，Analyst 工具会根据度量得分和您分配给每个度量的权重计算每个度量组的加权平均值。
18. 输入 **CustomerTier** 和**状态**度量的权重。
19. 单击**下一步**。

20. 在**第 8 步(共 8 步)** 屏幕中，默认情况下会选择**本地**选项。单击**保存**以创建结果卡。结果卡显示在**结果卡**工作区中。

任务 2. 运行结果卡

在本任务中，运行 `sc_LA_Customer` 结果卡，为 `CustomerTier` 和“状态”列生成得分。

1. 在**库**工作区中，单击**资产** > **结果卡**。
2. 在**结果卡**窗格中选择一个结果卡。
3. 单击**操作** > **打开**。
结果卡将显示。
4. 单击**操作** > **运行结果卡**
此时将显示**运行结果卡**对话框。
5. 验证对话框中的设置，然后单击**运行**。

结果卡工作区显示 `CustomerTier` 和“状态”列的得分。

任务 3. 查看结果卡

在本任务中，查看 `sc_LA_Customer` 结果卡，以查看 `CustomerTier` 和“状态”列的得分。

1. 选择包含要查看的“状态”得分的**状态**行。
在 `sc_LA_Customer` - **度量**部分，可以查看结果卡的以下属性：
 - 结果卡名称。
 - 结果卡中的总行数。
 - 无效行的数量。
 - 得分以及水平条形图。
 - 得分趋势。您可以单击得分趋势，在**趋势图表详细信息**屏幕中查看图形表示形式。
 - 度量的权重。
 - 无效数据的成本。
 - 成本趋势。
 - 数据对象。单击数据对象可以在**发现**工作区中查看数据对象的数据预览。
 - 列或规则名称。
 - 源类型。
 - 向下钻取图标。
2. 单击“状态”行中的向下钻取图标。
“状态”列的无效得分显示在**向下钻取**窗格中的**无效行**部分。
3. 选择**有效行数**以查看对“状态”列有效的得分。
4. 单击 `CustomerTier` 行中的向下钻取图标。
`CustomerTier` 列的所有得分均有效。

任务 4. 编辑结果卡

在本任务中，编辑 sc_LA_Customer 结果卡以指定 Ruby 值为 CustomerTier 得分的无效值。

1. 确认已进入**结果卡**工作区，并已打开 sc_LA_Customer 结果卡。
2. 选择**操作 > 编辑 > 度量**。
此时将显示**编辑结果卡**对话框。
3. 在**度量**部分中，选择 **CustomerTier**。
4. 在**正在使用的得分：值**部分中，将 **Ruby** 从**有效值**部分移动到**可用值**部分。
接受**度量阈值**部分中的默认设置。
5. 单击**保存并运行**，以保存结果卡更改并运行结果卡。
6. 再次查看 CustomerTier 得分。
CustomerTier 得分更改为 81.4%。

任务 5. 配置阈值

在本任务中，配置 sc_LA_Customer 结果卡中“状态”得分的阈值，以确定“状态”列中数据的可接受范围。具有两个字母代码的值（例如 CA）都可接受，多于两个字母的代码（例如 Calif）不可接受。

1. 确认已进入**结果卡**工作区，并已打开 sc_LA_Customer 结果卡。
2. 选择**操作 > 编辑 > 度量**。
此时将显示**编辑结果卡**对话框。
3. 在**度量**部分中，选择**状态**。
4. 在**度量阈值**部分中，为“正常”和“不可接受”得分输入以下范围：90 到 100% 为“正常”；0 到 50% 为“不可接受”；51% 到 89% 为“可接受”。
这些阈值表示可接受和正常范围的下限。
5. 单击**保存并运行**，以保存结果卡更改并运行结果卡。
在**结果卡**面板中，查看得分百分比更改和在“状态”得分显示为条形图的得分。

任务 6. 查看得分趋势图表

在本任务中，查看“状态”得分的趋势图表。您可以查看趋势图表以监控今后的得分。

1. 确认已进入**结果卡**工作区，并已打开 sc_LA_Customer 结果卡。
2. 选择**状态**行。
3. 单击**操作 > 显示趋势图**，或单击**得分趋势**列下的箭头。
此时将显示**趋势图表详细信息**对话框。您可以查看得分的**正常、可接受和不可接受**阈值。编辑结果卡中的得分值后，每次运行结果卡都会更改阈值。
4. 指向图表中的任何圆，以查看图表底部的**有效值**部分中的有效值。
5. 单击**关闭**以返回到结果卡。

创建并运行结果卡课程总结

在本课程中，您学习到可以通过配置文件的结果创建结果卡。结果卡包含配置文件中的列。您还学习到，可以运行结果卡以为列生成得分。编辑结果卡以配置有效值并设置得分阈值。还学习到如何查看得分趋势图表。

通过配置文件中的 CustomerTier 和“状态”列创建结果卡，以分析客户层和状态列的数据质量。运行结果卡以为每列生成得分。编辑结果卡，为得分指定不同的有效值。配置得分阈值并查看得分趋势图表。

第 8 章

课程 7。使用配置文件列创建引用表

本章包括以下主题：

- [使用配置文件列创建引用表概览, 39](#)
- [任务 1. 使用配置文件列创建引用表, 40](#)
- [任务 2. 编辑引用表, 40](#)
- [使用配置文件列创建引用表课程总结, 41](#)

使用配置文件列创建引用表概览

引用表包含您可以用于标准化源数据的引用数据。引用数据可以包括有效的标准值。创建引用表以建立源数据值和有效标准值之间的关系。

可以通过配置文件的结果创建引用表。创建引用表后，您可以编辑引用表以添加列或行以及添加或编辑标准有效值。可以在审计跟踪中查看对引用表所做的更改。

故事

HypoStores 想要剖析数据以发现异常并使用有效值标准化数据。您是分析人员，负责标准化数据中的有效值。您想要根据配置文件列中的有效值创建引用表。

目标

在本课程中，您将完成以下任务：

1. 通过选择列的有效值，从 Profile_LA_Customers_Custom 配置文件中的 CustomerTier 列创建引用表。
2. 编辑引用表来为列配置不同的有效值。

先决条件

开始本课程之前，请确认满足以下先决条件：

- 已完成本课程中的课程 1 到 6。

计时

完成本课程中的任务需要 15 分钟。

任务 1. 使用配置文件列创建引用表

在本任务中，创建引用表并将 Profile_LA_Customers 配置文件中的 CustomerTier 列添加到引用表中。

1. 在**库**工作区中，单击**资产** > **配置文件**。
2. 单击 **Profile_LA_Customers** 配置文件以在摘要视图中打开配置文件结果。
3. 在摘要视图中，选择要添加到引用表的 CustomerTier 列。单击右键并选择**添加到引用表**。
此时将显示**添加到引用表**对话框。
4. 选择**创建引用表**。
5. 单击**下一步**。
6. 在**名称**字段中，输入 **Reftab_CustTier_HypoStores**。
7. 输入说明并将 0 设置为默认值。
Analyst 工具为任何不包含值的表记录使用该默认值。
8. 单击**下一步**。
9. 在**列属性**部分中，配置 CustomerTier 列的以下列属性：

属性	说明
名称	CustomerTier
数据类型	字符串
精度	10
说明	引用客户层值

10. （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。
11. 验证**预览**部分中的 CustomerTier 列值。
12. 单击**下一步**。
此时将显示 Reftab_CustomerTier_HypoStores 引用表名称。您可以输入可选的说明。
13. 在**保存位置**部分中，选择要创建引用表所在的教程项目。
引用表：面板将列出您选择的位置中的引用表。
14. 输入可选的审计说明。
15. 单击**完成**。

任务 2. 编辑引用表

在本任务中，编辑 Reftab_CustomerTier_HypoStores 表以为客户层添加备用值。

1. 在**库**工作区中，单击**资产** > **引用表**。
2. 单击 Reftab_CustomerTier_HypoStores 引用表。
该引用表将在**设计**工作区中打开。
3. 要编辑某行，请选择该行并单击**操作** > **编辑**或单击**编辑**图标。

此时将显示**编辑行**对话框。（可选）选择多行，以将同一备用值添加到每行。

4. 为“Diamond”、“Diamond”、“Gold”、“Silver”和“Bronze”行输入以下备用值：1、2、3、4、5。
输入可选的审计说明。
5. 单击**应用**以应用更改。
6. 单击**关闭**。
更改后的引用表值会显示在**设计**工作区中。

使用配置文件列创建引用表课程总结

在本课程中，您学习了如何通过配置文件的结果创建引用表来为源数据配置有效值。

通过选择列的有效值，从配置文件列创建引用表。编辑了引用表来为列配置不同的有效值。

第 9 章

课程 8。创建引用表

本章包括以下主题：

- [创建引用表概览，42](#)
- [任务 1。创建引用表，43](#)
- [创建引用表课程总结，43](#)

创建引用表概览

引用表包含您可以用于标准化源数据的引用数据。引用数据可以包括有效的标准值。创建引用表以建立源数据值和有效标准值之间的关系。

可以使用引用表编辑器手动创建引用表。使用引用表定义并标准化源数据。您可以与开发人员共享引用表，以在 Developer tool 中的标准创建器和查找转换中使用。

故事

HypoStores 想要使用有效值标准化数据。您是分析人员，负责标准化数据中的有效值。您想要创建引用表以定义引用 LA 客户数据的标准客户层代码。然后可以与开发人员共享该引用表。

目标

在本课程中，您将完成以下任务：

- 使用引用表编辑器创建引用表，以定义引用 LA 客户数据的标准客户层代码。

先决条件

开始本课程之前，请确认满足以下先决条件：

- 已完成本课程中的课程 1 和 2。

计时

完成本课程中的任务需要 10 分钟。

任务 1。创建引用表

在本任务中，您将创建 Reftab_CustomerTier_Codes 引用表，以标准化客户层数据的有效值。

1. 单击**新建** > **引用表**。
此时将显示**新建引用表**向导。
2. 选择**使用引用表编辑器**。
3. 单击**下一步**。
4. 对于想要纳入引用表中的每列，单击**添加新列**图标并为每列配置列属性。
添加以下列名称：CustomerID、CustomerTier 和“状态”。您可以将列重新排序或删除列。
5. 输入可选说明并将默认值设置为 **0**。
Analyst 工具为任何不包含值的表记录使用该默认值。
6. 单击**下一步**。
7. 在**名称**字段中，输入 **Reftab_CustomerTier_Codes**。
8. 在**文件夹**部分中，选择教程项目中的 **Customers** 文件夹。
9. 单击**完成**。
该引用表将显示在**设计**工作区中。
10. 从**操作**菜单中，选择**添加行**以使用以下四个值填充每个引用表列。
CustomerID = LA1、LA2、LA3、LA4
CustomerTier = 1、2、3、4。
状态 = 活动、非活动

创建引用表课程总结

在本课程中，您学习到如何使用引用表编辑器来创建引用表，以创建用于源数据的标准有效值。

您使用引用表编辑器创建了引用表，以标准化 LA 客户数据的客户层值。

第 II 部分： Informatica Developer 入门

本部分包含以下章节：

- [课程 1。设置 Informatica Developer, 45](#)
- [课程 2: 导入物理数据对象, 49](#)
- [课程 3. 对源数据运行配置文件, 58](#)
- [课程 4。解析数据, 62](#)
- [课程 5。标准化数据 , 68](#)
- [课程 6。验证地址数据, 74](#)

第 10 章

课程 1。设置 Informatica Developer

本章包括以下主题：

- [设置 Informatica Developer 概览, 45](#)
- [任务 1。启动 Informatica Developer, 46](#)
- [任务 2。添加域, 46](#)
- [任务 3。添加模型存储库, 47](#)
- [任务 4。创建项目, 47](#)
- [任务 5。创建文件夹, 47](#)
- [任务 6。选择默认数据集成服务, 48](#)
- [设置 Informatica Developer 课程总结, 48](#)

设置 Informatica Developer 概览

在开始本教程中的课程之前，您必须先启动并设置 Developer tool。要设置 Developer tool，需要添加域。添加位于域中的模型存储库，并创建一个项目和文件夹来存储工作。还要选择默认数据集成服务。

Informatica 域是可定义 Informatica 环境的节点和服务的集合。域中的服务包括模型存储库服务和数据集成服务。

模型存储库服务用于管理模型存储库。模型存储库是存储您在 Developer tool 中所创建项目的元数据的关系数据库。项目用于存储在 Developer tool 中创建的对象。项目还可以包含存储相关对象的文件夹，例如属于同一业务要求的对象。

数据集成服务在 Developer tool 中执行数据集成任务。

目标

在本课程中，您将完成以下任务：

- 启动 Developer tool，转到 Developer tool 工作台。
- 在 Developer tool 中添加域。
- 添加模型存储库，以便可以创建项目。
- 创建一个项目，用于存储在 Developer tool 中创建的对象。
- 在项目中创建一个可存储相关对象的文件夹。

- 选择默认数据集成服务，以执行数据集成任务。

先决条件

开始本课程之前，请验证以下先决条件：

- 您已安装 Developer tool。
- 您具有连接到域所需的域名、主机名和端口号。可以通过域管理员获取此信息。
- 域管理员已在 Administrator 工具中配置模型存储库服务。
- 拥有用户名和密码才能访问模型存储库服务。可以通过域管理员获取此信息。
- 域管理员已配置数据集成服务。
- 数据集成服务正在运行。

计时

完成本课程中的任务需要 5-10 分钟。

任务 1。启动 Informatica Developer

启动 Developer 工具以开始教程。

1. 启动 Developer 工具。
此时将显示 Developer 工具的**欢迎**页面。
2. 单击**工作台**按钮。
此时将显示 Developer 工具的工作台。

任务 2。添加域

在本任务中，在 Developer 工具中添加域，以访问模型存储库。

1. 单击 **Window > 首选项**。
此时将显示**首选项**对话框。
2. 单击 **Informatica > 域**。
3. 单击**添加**。
此时将显示**新建域**对话框。
4. 输入域名、主机名和端口号。
5. 单击**完成**。
6. 单击**确定**。

任务 3。添加模型存储库

在本任务中，添加要用于存储项目和文件夹的模型存储库。

1. 单击**文件** > **连接到存储库**。
此时将显示**连接到存储库**对话框。
2. 单击**浏览**以选择模型存储库服务。
3. 单击**确定**。
4. 单击**下一步**。
5. 输入用户名和密码。
6. 选择命名空间。
7. 单击**完成**。
模型存储库显示在**对象浏览器**视图中。

任务 4。创建项目

在本任务中，创建项目以存储在 Developer 工具中创建的对象。您可以创建一个项目用于本指南中的所有教程。

1. 在**对象浏览器**视图中，选择一个模型存储库服务。
2. 单击**文件** > **新建** > **项目**。
此时将显示**新建项目**对话框。
3. 输入加前缀“Tutorial_”的名称作为项目的名称。
4. 单击**完成**。
项目显示在**对象浏览器**视图中的“模型存储库服务”下。

任务 5。创建文件夹

在本任务中，创建存储相关对象的文件夹。您可以创建一个文件夹用于存储本指南中的所有教程。

1. 在**对象浏览器**视图中，选择要将文件夹添加到的项目。
2. 单击**文件** > **新建** > **文件夹**。
3. 输入文件夹的名称。
4. 单击**完成**。
Developer 工具将在**对象浏览器**视图中的项目下添加文件夹。展开项目以查看文件夹。

任务 6。选择默认数据集成服务

在本任务中，选择默认数据集成服务，以便可以运行映射和预览数据。

1. 单击 **Window** > **首选项**。
此时将显示**首选项**对话框。
2. 选择 **Informatica** > **数据集成服务**。
3. 展开域。
4. 选择数据集成服务。
5. 单击**设置为默认值**。
6. 单击**确定**。

设置 Informatica Developer 课程总结

在本课程中，您学习到 Informatica 域包括模型存储库服务和数据集成服务。模型存储库服务用于管理模型存储库。模型存储库包含项目和文件夹。数据集成服务执行数据集成任务。

启动并设置 Developer 工具。将域添加到 Developer 工具、添加模型存储库以及创建项目和文件夹。还选择了默认数据集成服务。

现在，可以使用 Developer 工具完成本教程中的其他课程。

第 11 章

课程 2：导入物理数据对象

本章包括以下主题：

- [导入物理数据对象概览, 49](#)
- [任务 1. 导入 Boston Customers 平面文件数据对象, 50](#)
- [任务 2. 导入 LA Customers 平面文件数据对象, 56](#)
- [任务 3. 导入 All Customers 平面文件数据对象, 56](#)
- [导入物理数据对象课程总结, 57](#)

导入物理数据对象概览

物理数据对象是基于平面文件或关系数据库表的数据表示。您可以导入平面文件或关系数据库表作为物理数据对象，以用作映射中的源或目标。

故事

HypoStores Corporation 在平面文件中存储来自洛杉矶和波士顿办公室的客户数据。您想要在 Developer tool 中处理该客户数据。要执行该操作，您需要导入每个平面文件作为物理数据对象。

目标

在本课程中，您导入平面文件作为物理数据对象。还要设置源文件目录，以便数据集成服务可以从正确的目录中读取源数据。

先决条件

开始本课程之前，请确认满足以下先决条件：

- 已完成本课程中的课程 1。

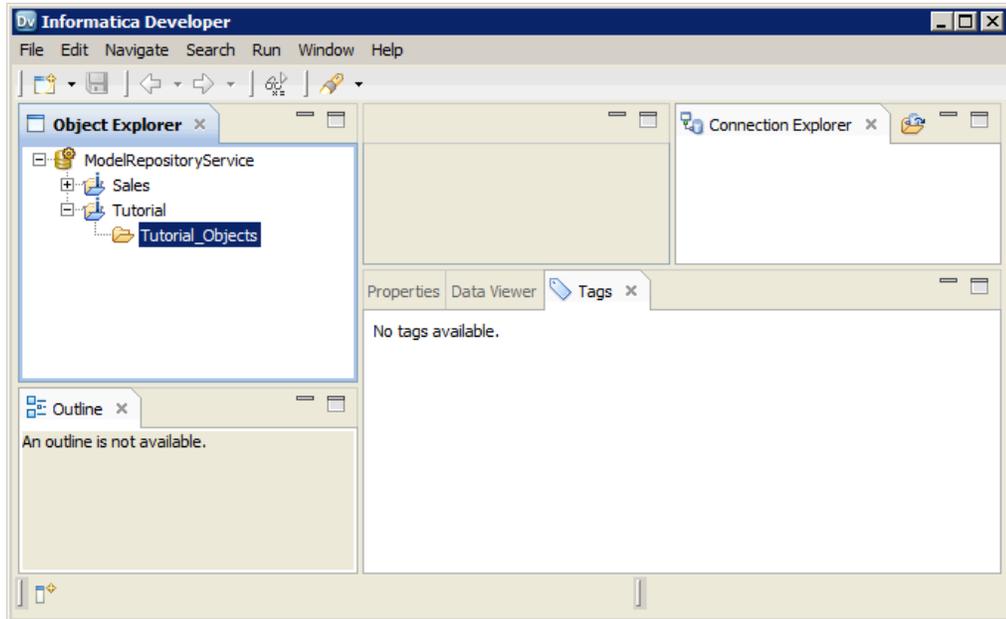
计时

完成本课程中的任务需要 10-15 分钟。

任务 1. 导入 Boston_Customers 平面文件数据对象

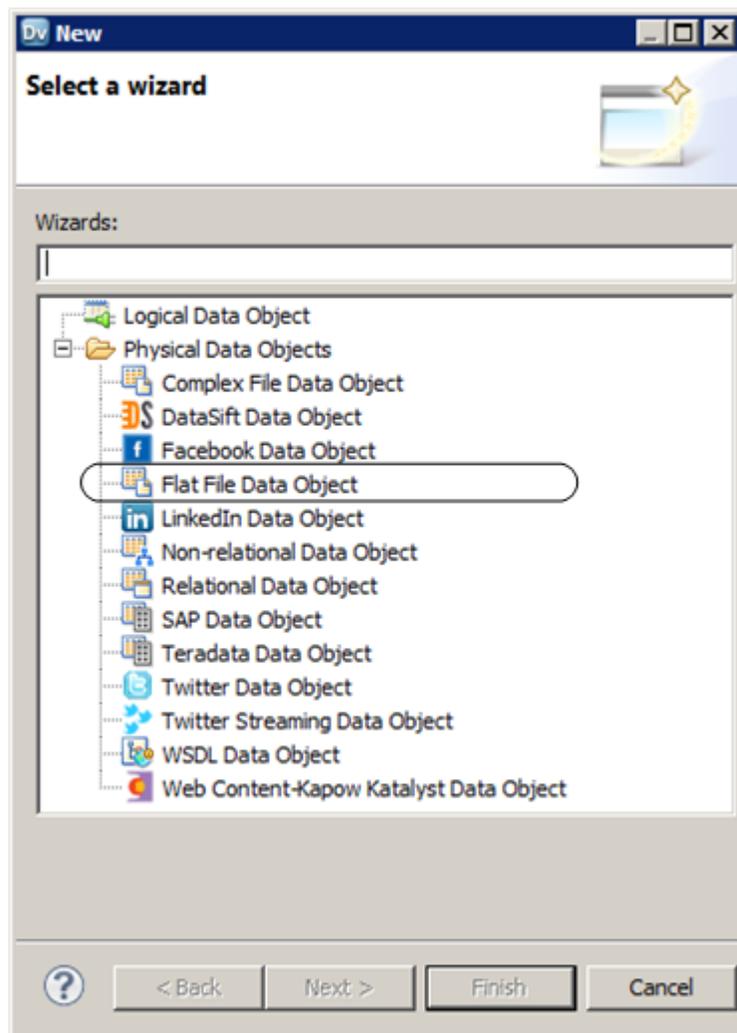
在本任务中，从包含波士顿办公室的客户数据的文件导入物理数据对象。

1. 在**对象浏览器**视图中，选择 Tutorial_Objects 文件夹。



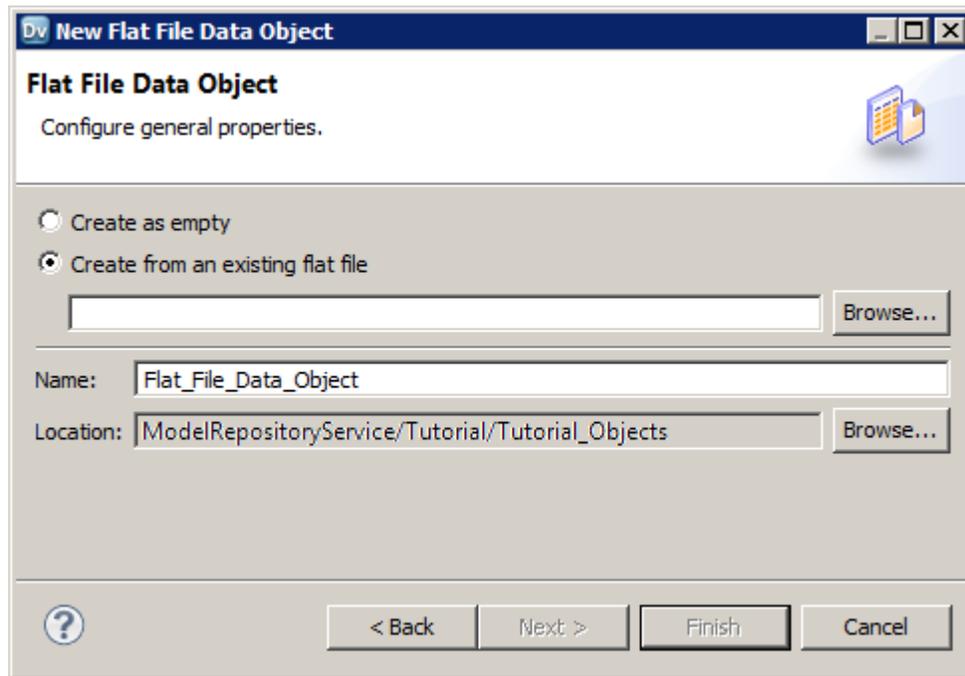
2. 右键单击 Tutorial_Objects 文件夹并选择**新建** > **数据对象**。

此时将显示新建对话框。



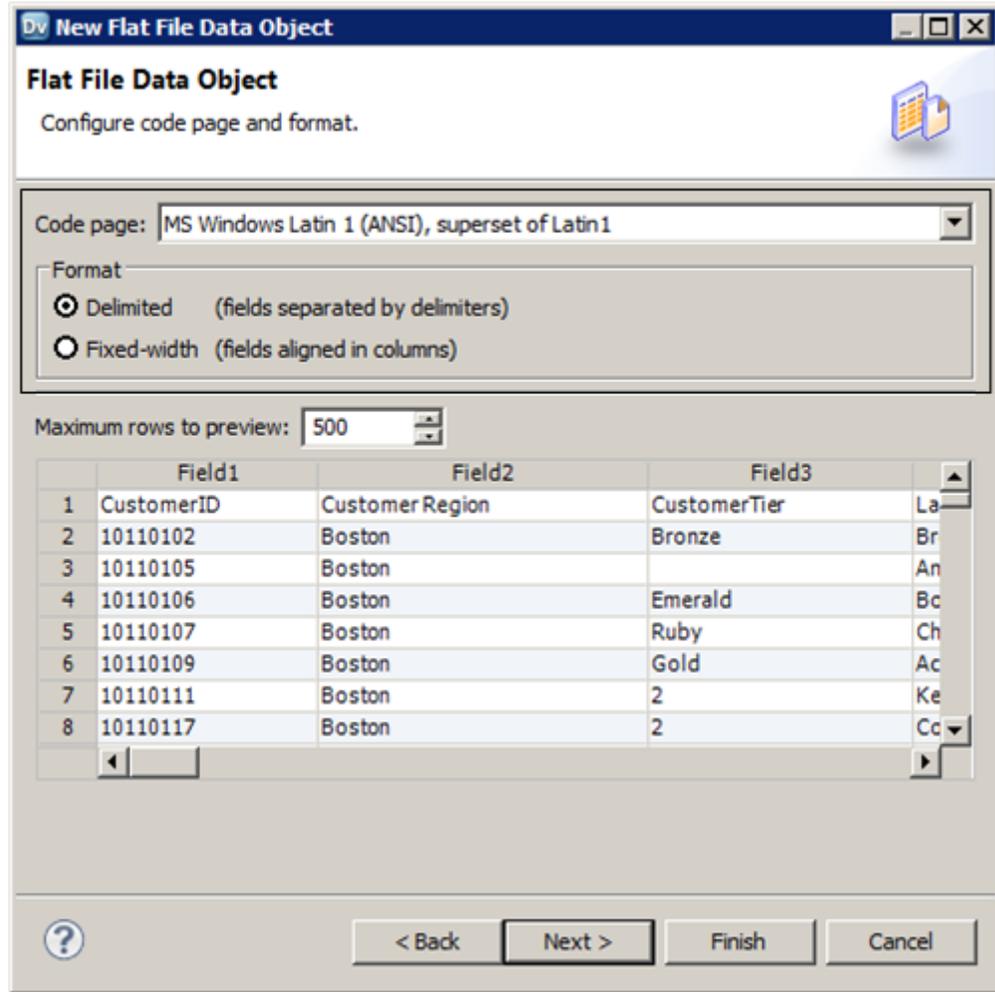
3. 选择物理数据对象 > 平面文件数据对象，然后单击下一步。

此时将显示新建平面文件数据对象对话框。



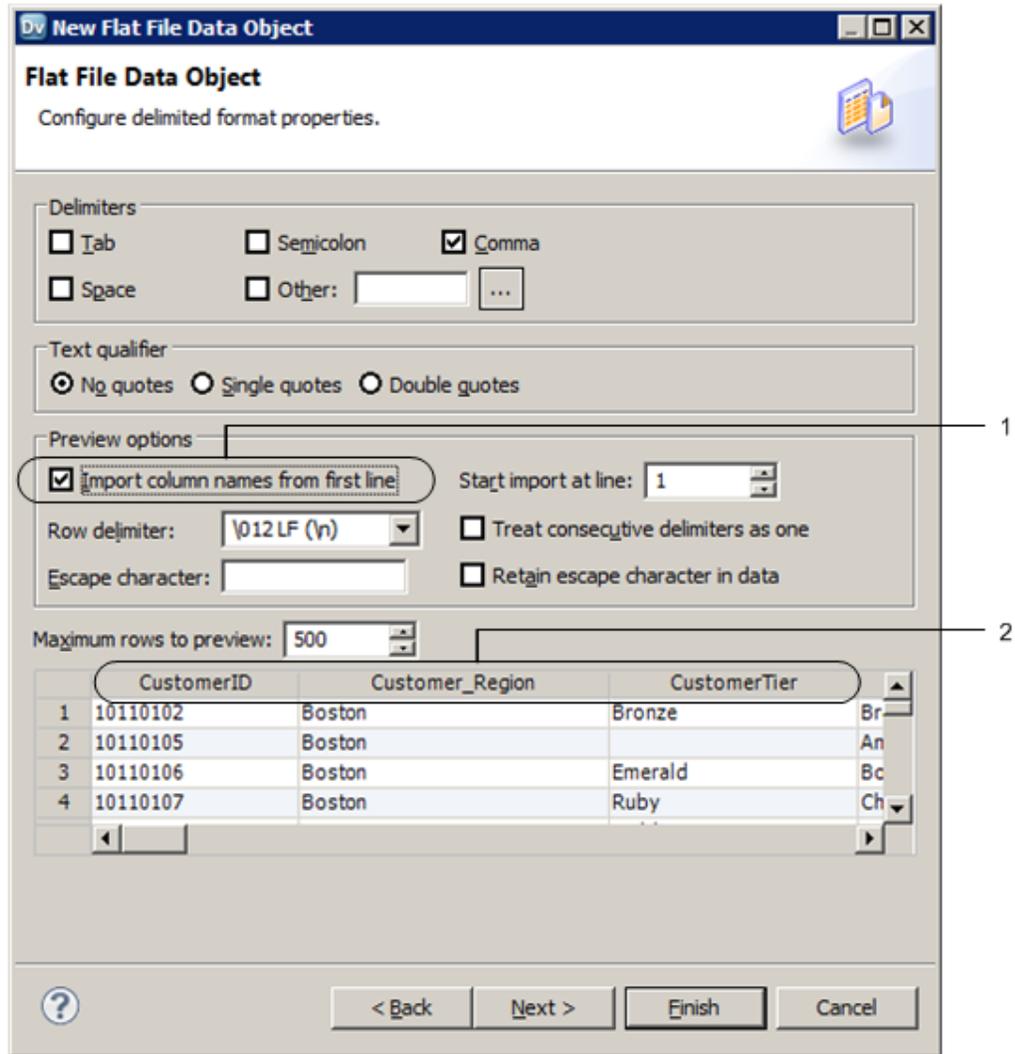
4. 选择**从现有平面文件创建**。
5. 单击**浏览**并导航到 Developer tool 计算机的以下目录中的 Boston_Customers.csv: <Informatica 安装目录>\clients\DeveloperClient\Tutorials
6. 单击**打开**。
向导会将该数据对象命名为“Boston_Customers”。
7. 单击**下一步**。
8. 验证代码页已设置为 **MS Windows Latin 1 (ANSI)**, **Latin 1 的超集**, 并且格式设置为**带分隔符**。

新建平面文件数据对象对话框将显示默认的代码页、格式和平面文件数据的预览。



- 单击下一步。
- 选择导入第一行中的列名称。

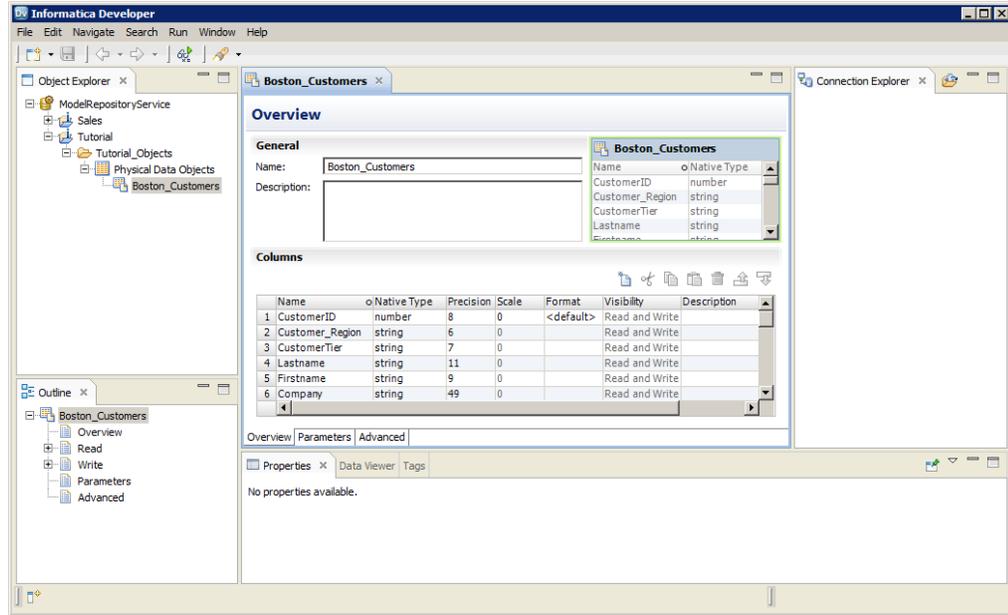
新建平面文件数据对象对话框将在平面文件数据的预览中显示列名称。



1. 导入第一行中的列名称选项
2. 列名称

11. 单击完成。

Boston_Customers 物理数据对象将显示在 Tutorial_Objects 文件夹中的 Physical Data Objects 文件夹下。**概览**视图将显示文件内容，并且文件将在编辑器中打开。



- 单击**高级**视图。

高级视图会显示物理数据对象的属性。

- 在**高级**视图中，滚动至**运行时：读取**部分。

- 在**运行时：读取**部分中，将**源文件目录**设置为数据集成服务计算机上的以下目录：<Informatica 安装目录>\server\Tutorials

数据集成服务会在运行数据集成服务的计算机的服务器目录中搜索源文件。服务器安装中含有教程文件的副本。数据集成服务无法从客户端安装目录读取文件，除非更改了源文件和目录的访问权限。

下图显示了一个示例源文件目录：

Advanced	
Name	Value
Runtime : Read	
Input type	File
Source type	Direct
Source file name	Boston_Customers.csv
Source file directory	\\MyMachine\Informatica\10.0\server\Tutorials

注意：Developer tool 计算机必须对运行数据集成服务的计算机上的源文件目录拥有访问权限。如果 Developer tool 无法访问源文件目录，它将无法预览源文件中的数据，也无法运行对源文件中数据进行访问的映射。如果您运行多个数据集成服务，每个数据集成服务都会有一个单独的源文件目录。

- 单击**数据查看器**视图。

- 在**数据查看器**视图中，单击**运行**。

数据集成服务将从 Boston_Customers 文件中读取数据，并在“输出”窗口中显示结果。

- 单击**文件** > **保存**以保存 Boston_Customers 物理数据对象。

任务 2。导入 LA_Customers 平面文件数据对象

在本任务中，从平面文件中导入物理数据对象，该平面文件包含来自洛杉矶办公室的客户数据。

1. 在**对象浏览器**视图中，选择教程项目。
2. 单击**文件 > 新建 > 数据对象**。
此时将显示**新建**对话框。
3. 选择**物理数据对象 > 平面文件数据对象**，然后单击**下一步**。
此时将显示**新建平面文件数据对象**对话框。
4. 选择**从现有平面文件创建**。
5. 单击**浏览**，导航到以下目录中的 LA_Customers.csv：<Informatica 安装目录>\clients\DeveloperClient\Tutorials
6. 单击**打开**。
该向导会将数据对象命名为 LA_Customers。
7. 单击**下一步**。
8. 验证代码页面是否为 MS Windows Latin 1 (ANSI)，Latin 1 的超集。
9. 验证格式是否带分隔符。
10. 单击**下一步**。
11. 验证分隔符是否设置为逗号。
12. 选择**导入第一行中的列名称**。
13. 单击**完成**。
LA_Customers 物理数据对象显示在教程项目中的**物理数据对象**下。
14. 单击**读取**视图，然后选择输出转换。
15. 单击**属性**视图中的**运行时**选项卡。
16. 将源文件目录设置为数据集成服务计算机上的以下目录：<Informatica 安装目录>\server\Tutorials
17. 单击**文件 > 保存**。

任务 3。导入 All_Customers 平面文件数据对象

在本任务中，从平面文件中导入物理数据对象，该平面文件合并了来自洛杉矶和波士顿办公室的客户订单数据。

1. 在**对象浏览器**视图中，选择教程项目。
2. 单击**文件 > 新建 > 数据对象**。
此时将显示**新建**对话框。
3. 选择**物理数据对象 > 平面文件数据对象**，然后单击**下一步**。
此时将显示**新建平面文件数据源**对话框。
4. 选择**从现有平面文件创建**。
5. 单击**浏览**，导航到以下目录中的 All_Customers.csv：<Informatica 安装目录>\clients\DeveloperClient\Tutorials。
6. 单击**打开**。
该向导会将数据对象命名为 All_Customers。

7. 单击**下一步**。
8. 验证代码页面是否为 MS Windows Latin 1 (ANSI), Latin 1 的超集。
9. 验证格式是否带分隔符。
10. 单击**下一步**。
11. 验证分隔符是否设置为逗号。
12. 选择**导入第一行中的列名称**。
13. 单击**完成**。
All_Customers 物理数据对象显示在教程项目中的**物理数据对象**下。
14. 单击**读取**视图, 然后选择输出转换。
15. 单击**属性**视图中的**运行时**选项卡。
16. 将源文件目录设置为数据集成服务计算机上的以下目录: <Informatica 安装目录>\server\Tutorials
17. 单击**文件** > **保存**。

导入物理数据对象课程总结

在本课程中, 您学习到物理数据对象是基于平面文件或关系数据库表的数据的表示形式。

您通过平面文件创建了物理数据对象。还要设置源文件目录, 以便数据集成服务可以从正确的目录中读取源数据。

使用数据对象作为数据质量课程中的映射源。

第 12 章

课程 3. 对源数据运行配置文件

本章包括以下主题：

- [剖析数据概览, 58](#)
- [任务 1。对两个数据源执行联接分析, 59](#)
- [任务 2。查看联接分析结果, 60](#)
- [任务 3。对数据源运行配置文件, 60](#)
- [任务 4。查看列剖析结果, 60](#)
- [剖析数据课程总结, 61](#)

剖析数据概览

配置文件是一组描述数据集内容和结构的元数据。

剖析和数据发现通常是要在项目中执行的第一个步骤。可以运行配置文件来评估数据的结构，并验证数据列是否使用所需的信息类型填充。如果配置文件显示数据中出现问题，则可以在项目中定义步骤来修复这些问题。例如，如果配置文件显示某列包含的值的长度大于所需长度，则可以设计数据质量进程以删除或修复问题值。

分析选定列的数据质量的配置文件称为列配置文件。

注意：还可以使用 Developer tool 来发现主键、外键和功能相关性关系，以及分析数据列的联接条件。

列配置文件提供有关数据的以下事实：

- 每个列中的相异值和空值的数量，用数字和百分比表示。
- 每个列中数据的模式以及这些值出现的频率。
- 有关列值的统计信息，例如，每个列中值的最大和最小长度，以及第一个值和最后一个值。
- 对于联接分析配置文件，两个数据列之间的重叠度以韦恩图和百分比值的形式显示。使用联接分析配置文件识别列联接条件可能发生的问题。

您可以在项目的任何阶段运行列配置文件，以衡量数据质量并验证数据更改是否符合项目的目标。您可以对映射中的转换运行列配置文件，以指示转换对数据的影响。

故事

HypoStores 想要验证客户数据是否没有错误、不一致情况和重复信息。在 HypoStores 设计达到数据质量目标的流程之前，需要衡量其源数据文件的质量并确认数据已准备好可供处理。

目标

在本课程中，您将完成以下任务：

- 对 Boston_Customers 数据源和 LA_Customers 数据源执行联接分析。
- 查看联接分析的结果，以确定是否可以成功合并两个办公室的数据。
- 对 All_Customers 数据源运行列配置文件。
- 查看列剖析结果，以观察数据中包含的值和模式。

先决条件

开始本课程之前，请确认满足以下先决条件：

- 已完成本课程中的课程 1 和 2。

所需时间

- 完成本课程需要 20 分钟。

任务 1。对两个数据源执行联接分析

在本任务中，对 Boston_Customers 和 LA_Customers 数据源执行联接分析，以查看联接条件。

1. 选择教程文件夹并单击**文件 > 新建 > 配置文件**。
2. 选择**企业发现配置文件**。
3. 单击**下一步**。
4. 在**名称**字段中，输入 **Tutorial_Profile**。
5. 单击**完成**。
Tutorial_Profile 配置文件显示在对象浏览器中。
6. 将 **Boston_Customers** 和 **LA_Customers** 数据源拖到右侧编辑器中。
提示：按住 Shift 键以选择多个数据对象。
7. 右键单击数据对象名称并选择**联接配置文件**。
此时将显示**新建联接配置文件**向导。
8. 在**名称**字段中，输入 **JoinAnalysis**。
9. 验证 Boston_Customers 和 LA_Customers 是否显示为数据对象，然后单击**下一步**。
10. 验证是否已在两个数据源中选定 **CustomerID** 列。
向下滚动向导窗格以查看两个数据集中的列。
单击**下一步**。
11. 单击**添加**以添加联接条件。
此时将显示**联接条件**对话框。
12. 在**列**部分中，单击**添加行**。
13. 双击左侧列中的第一行并选择 **CustomerID**。
14. 双击右侧列中的第一行并选择 **CustomerID**。
15. 单击**确定**，然后单击**完成**。
16. 如果 Developer 工具提示您保存更改，请单击**是**。
Developer 工具会运行配置文件。

注意：请不要关闭配置文件。查看下一任务中的配置文件结果。

任务 2。查看联接分析结果

在本任务中，在 JoinAnalysis 配置文件的“联接结果”视图中查看联接分析结果。

1. 单击编辑器中的 **JoinAnalysis** 选项卡。

2. 在**联接结果**部分中，单击第一行。

详细信息部分会显示详细列明联接分析结果的韦恩图和颜色键。

3. 验证**联接行**列是否显示零作为包含联接的行数。

该值指示没有重复项的 CustomerID 字段。您可以成功合并这两个数据源。

4. 要查看 LA_Customers 数据对象的 CustomerID 值，请双击韦恩图中名为 **LA_Customers** 的圆。

提示：双击韦恩图中的圆以查看数据行。如果韦恩图中的圆相交，则双击相交区域可查看两个数据集共用的数据值。

数据查看器显示 LA_Customers 数据对象中的 CustomerID 值。

任务 3。对数据源运行配置文件

在本任务中，对 All_Customers 数据源运行配置文件，以查看数据的内容和结构。

1. 在**对象浏览器**视图中，浏览到教程项目中的数据对象。

2. 选择 **All_Customers** 数据源。

3. 单击**文件 > 新建 > 配置文件**。

此时将显示**新建**对话框。

4. 选择**配置文件**。

5. 单击**下一步**。

6. 在**名称**字段中，输入 **All_Customers**。

7. 单击**完成**。

All_Customers 配置文件会在编辑器中打开，并且该配置文件将会运行。

任务 4。查看列剖析结果

在本任务中，查看 All_Customers 数据对象的列剖析结果并检查数据中包含的值和模式。

1. 单击**窗口 > 显示视图 > 进度**以查看 All_Customers 配置文件的进度。

此时将打开“进度”视图。

2. 当“进度”视图报告 All_Customers 配置文件运行完成时，单击编辑器中的**结果**视图。

3. 在**列剖析**部分中，单击 **CustomerTier** 列。

详细信息部分会显示 CustomerTier 列中包含的所有值，并显示数据集中的值发生频率的相关信息。

4. 在**详细信息**部分中，双击 **Ruby**。
数据查看器会运行并显示 CustomerTier 列包含值 **Ruby** 的记录。
5. 在**列剖析**部分中，单击 **OrderAmount** 列。
6. 在**详细信息**部分中，单击**显示**列表并选择**模式**。
详细信息部分显示在 OrderAmount 列中找到的模式。“模式”列中的字符串 **9(5)** 是指包含五位数订单金额的记录。字符串 **9(4)** 是指包含四位数订单金额的记录。
7. 在**模式**列中，双击字符串 **9(4)**。
数据查看器会运行并显示 OrderAmount 列包含四位数订单金额的记录。
8. 在**详细信息**部分中，单击**显示**列表并选择**统计信息**。
详细信息部分显示 OrderAmount 列的统计信息，包括平均值、标准偏差、最大和最小长度、五个最常用的值以及五个最不常用的值。

剖析数据课程总结

在本课程中，您学习到配置文件提供有关数据内容和结构的信息，

并学习到可以对两个数据对象执行联接分析，以及查看两个数据对象之间的重叠度。您还学习到，可以对数据对象运行列配置文件并查看与数据对象中的每个列相关的值、模式和统计信息。

您创建了 JoinAnalysis 配置文件来确定 Boston_Customers 数据对象中的数据是否可以与 LA_Customers 数据对象中的数据合并。您查看了该配置文件的结果并确定 CustomerID 列中的所有值都是唯一的，您可以成功合并这些数据对象。

您创建了 All_Customers 配置文件并对 All_Customers 数据对象运行列配置文件。您查看了该配置文件的结果，以发现 All_Customers 数据对象中列的值、模式和统计信息。最后，您运行了数据查看器来查看包含您选定的值和模式的行，这可以让您验证数据质量。

第 13 章

课程 4。解析数据

本章包括以下主题：

- [解析数据概览, 62](#)
- [任务 1。创建目标数据对象, 63](#)
- [任务 2。创建映射以解析数据, 64](#)
- [任务 3。对解析器转换运行配置文件, 66](#)
- [任务 4。运行映射, 66](#)
- [任务 5。查看映射输出, 67](#)
- [解析数据课程总结, 67](#)

解析数据概览

您通过解析数据来确定输入字段中的一个或多个数据元素，并将每个元素写入不同的输出字段。

解析数据使您能够对每列中的信息有更大的控制权限。例如，假设一个数据字段包含某人的全名 Bob Smith。您可以使用解析器转换将该全名拆分为单独的名字和姓氏数据列。将数据解析为新列之后，您可以为每个列创建自定义的数据质量操作。

您可以配置解析器转换，以使用标志集将数据列解析为成分字符串。标志集可识别词语、邮政编码、电话号码和社会保障号等数据元素。

您还可以使用解析器转换解析与您输入的引用表条目或自定义正则表达式匹配的数据。

故事

HypoStores 希望 Los Angeles 办事处的客户数据文件格式与 Boston 办事处的数据文件格式一致。Los Angeles 办事处的客户数据将客户姓名存储在 FullName 列中，而 Boston 办事处的客户数据将客户姓名存储在单独的 FirstName 和 LastName 列中。HypoStores 需要将 Los Angeles 的 FullName 列的数据解析为名字和姓氏，使 Los Angeles 的数据格式与 Boston 的数据格式一致。

目标

在本课程中，您将完成以下任务：

- 创建并配置 LA_Customers_tgt 数据对象用以包含解析的数据。
- 创建映射将 FullName 列解析为单独的 FirstName 和 LastName 列。
- 将 LA_Customers 数据对象添加到映射，以连接到源数据。
- 将 LA_Customers_tgt 数据对象添加到映射，以创建目标数据对象。

- 将解析器转换添加到映射，并配置它使用标志集将全名解析为名字和姓氏。
- 对解析器转换运行配置文件，以便在生成目标数据源之前检查数据。
- 运行映射生成解析的名称。
- 运行数据查看器查看映射输出。

先决条件

开始本课程之前，请确认满足以下先决条件：

- 已完成本教程的第 1 课和第 2 课。

计时

完成本课程中的任务需要 20 分钟。

任务 1。创建目标数据对象

在此任务中，您将创建一个可向其中写入解析后的姓名的 LA_Customers_tgt 数据对象。

要创建目标数据对象，请完成以下步骤：

1. 基于 LA_Customers.csv 文件创建 LA_Customers_tgt 数据对象。
2. 配置该数据对象的读取和写入选项，包括文件位置和文件名。
3. 将 Firstname 和 Lastname 列添加到 LA_Customers_tgt 数据对象。

步骤 1。创建 LA_Customers_tgt 数据对象

在此步骤，您基于 LA_Customers.csv 文件创建 LA_Customers_tgt 数据对象。

1. 单击**文件** > **新建** > **数据对象**。
此时将打开**新建**窗口。
2. 选择**平面文件数据对象**，然后单击**下一步**。
3. 确认已选择**从现有平面文件创建**。
4. 单击**浏览**并导航到以下目录中的 LA_Customers.csv：<Informatica 安装目录>\clients\DeveloperClient\Tutorials
5. 单击**打开**。
6. 在名称字段中，输入 LA_Customers_tgt。
7. 单击**下一步**。
8. 单击**下一步**。
9. 在**预览选项**部分，选择**导入第一行中的列名称**，然后单击**下一步**。
10. 单击**完成**。

LA_Customers_tgt 数据对象将显示在编辑器中。

步骤 2。配置读取和写入选项

在此步骤，您配置 LA_Customers_tgt 数据对象的读取和写入选项，包括文件位置和文件名。

1. 确认 LA_Customers_tgt 数据对象已在编辑器中打开。

2. 在编辑器中，选择**读取**视图。
3. 单击**窗口 > 显示视图 > 属性**。
4. 在**属性**视图中，选择**运行时**视图。
5. 在**值**列中，双击源文件名并键入 LA_Customers_tgt.csv。
6. 在**值**列中，双击鼠标突出显示源文件目录。
7. 右键单击突出显示的名称并选择**复制**。
8. 在编辑器中，选择**写入**视图。
9. 在**属性**视图中，选择**运行时**视图。
10. 在**值**列中，双击**输出文件目录**条目。
11. 右键单击并选择**粘贴**以粘贴您从**读取**视图复制的目录位置。
12. 在**值**列中，双击**表头选项**条目并选择输出字段名称。
13. 在**值**列中，双击**输出文件名**条目并键入 LA_Customers_tgt.csv。
14. 单击**文件 > 保存**以保存该数据对象。

步骤 3。将列添加到数据对象

在此步骤，您将 Firstname 和 Lastname 列添加到 LA_Customers_tgt 数据对象。

1. 在“对象浏览器”视图中，浏览到教程项目中的数据对象。
2. 双击 LA_Customers_tgt 数据对象。
LA_Customers_tgt 数据对象将在编辑器中打开。
3. 确认已选中**概览**视图。
4. 选择 **FullName** 列并单击“新建”按钮添加列。
此时将显示一个名为 FullName1 的列。
5. 将该列重命名为 Firstname。单击**精度**字段并输入“30”。
6. 选择 Firstname 列并单击“新建”按钮添加列。
此时将显示一个名为 FirstName1 的列。
7. 将该列重命名为 Lastname。单击**精度**字段并输入“30”。
8. 单击**文件 > 保存**以保存该数据对象。

任务 2。创建映射以解析数据

在此任务中，您将创建一个映射并配置该映射，以使用数据对象和解析器转换。

要创建映射以解析数据，请完成以下步骤：

1. 创建映射。
2. 将源数据对象和目标数据对象添加到映射。
3. 将解析器转换添加到映射。
4. 配置解析器转换，以将包含客户全名的源列解析为包含名字和姓氏的单独列。

步骤 1。创建映射

在此步骤，您创建映射并为其命名。

1. 在“对象浏览器”视图中，选择教程项目。
2. 单击**文件** > **新建** > **映射**。
此时将打开**新建映射**窗口。
3. 在**名称**字段中，输入 ParserMapping。
4. 单击**完成**。
映射将在编辑器中打开。

步骤 2。将数据对象添加到映射

在此步骤，您将 LA_Customers 数据对象和 LA_Customers_tgt 数据对象添加到映射。

1. 在“对象浏览器”视图中，浏览到教程项目中的数据对象。
2. 选择 LA_Customers 数据对象并将其拖动到编辑器中。
此时将打开**将物理数据对象添加到映射**窗口。
3. 确认已选中**读取**，然后单击**确定**。
数据对象将显示在编辑器中。
4. 在“对象浏览器”视图中，浏览到教程项目中的数据对象。
5. 选择 LA_Customers_tgt 数据对象并将其拖动到编辑器中。
此时将打开**将物理数据对象添加到映射**窗口。
6. 选择**写入**，然后单击**确定**。
数据对象将显示在编辑器中。
7. 选择 LA_Customers 数据对象中的 CustomerID、CustomerTier 和 FullName 端口。将这些端口拖动到 LA_Customers_tgt 数据对象中的 CustomerID 端口。
提示：按住 **Ctrl** 键可选择多个端口。
LA_Customers 数据对象的端口将连接到 LA_Customers_tgt 数据对象中的对应端口。

步骤 3。将解析器转换添加到映射

在此步骤，您将解析器转换添加到 ParserMapping 映射。

1. 选择包含 ParserMapping 映射的编辑器。
2. 在转换选项板中，选择解析器转换。
3. 单击编辑器。
此时将打开**新建解析器转换**窗口。
4. 确认已选中**标志解析器**，然后单击**完成**。
解析器转换将显示在编辑器中。
5. 选择 LA_Customers 数据对象中的 FullName 端口，然后将其拖动到解析器转换的 Input 组。
该 FullName 端口将显示在解析器转换中，并连接到数据对象中的 FullName 端口。

步骤 4。配置解析器转换

在此步骤，您配置解析器转换来将包含客户全名的列解析为包含名字和姓氏的单独列。

1. 选择包含 ParserMapping 映射的编辑器。
2. 单击解析器转换。
3. 单击窗口 > 显示视图 > 属性。
4. 在“属性”视图中，选择策略视图。
5. 单击新建。此时将显示“新建策略”向导。
6. 单击“输入”列中的选择箭头，然后选择 FullName 端口。
7. 选择字符空格分隔符 [\s]。
8. 单击下一步。
9. 选择使用标志集解析操作，单后单击下一步。
10. 选中固定字符集(仅限单个输出)，然后选择未定义标志集。
11. 单击输出字段并选择新建。
12. 在操作输出对话框中，将输出名称更改为 Undefined_Output。
13. 单击完成。
14. 在解析器转换中，单击 Undefined_Output 端口并将其拖至 LA_customers_tgt 数据对象中的 FirstName 端口。端口之间将显示一个连接。
15. 在解析器转换中，单击 OverflowField 端口并将其拖至 LA_customers_tgt 数据对象中的 LastName 端口。端口之间将显示一个连接。
16. 单击文件 > 保存以保存该映射。

任务 3。对解析器转换运行配置文件

在此任务中，您将对解析器转换运行配置文件，以验证您是否已正确配置了解析器转换来解析全名。

1. 选择包含 ParserMapping 映射的编辑器。
2. 右键单击解析器转换并选择立即剖析。
配置文件将运行并在编辑器中打开。
3. 在编辑器中，单击结果视图显示剖析操作的结果。
4. 选择 Undefined_output 列以便在详细信息部分显示有关该列的信息。
Undefined_output 列中包含的值，连同每个值的频率和百分比统计信息，将显示在详细信息部分。
5. 查看这些数据并验证 Undefined_output 列中仅显示名字。

任务 4。运行映射

在此任务中，您将运行映射来创建映射输出。

1. 选择包含 ParserMapping 映射的编辑器。

2. 单击**运行** > **运行映射**。
映射将运行，并将输出写入 LA_Customers_tgt.csv 文件。

任务 5。查看映射输出

在此任务中，您将运行数据查看器来查看映射输出。

1. 在“对象浏览器”视图中，找到教程项目中的 LA_Customers_tgt 数据对象并双击该数据对象。
数据对象将在编辑器中打开。
2. 单击**窗口** > **显示视图** > **数据查看器**。
此时将打开“数据查看器”视图。
3. 在“数据查看器”视图中，单击**运行**。
数据查看器将运行并显示数据。
4. 验证 FirstName 和 LastName 列是否可正确显示解析的数据。

解析数据课程总结

在本课程中，您学习到解析数据会识别输入字段中的数据元素，并将每个元素写入新列。

您学习到可以使用解析器转换来解析数据。另外您还学习到，可以在映射中为转换创建配置文件，以分析该转换的输出。最后，您学习到可以使用数据查看器查看映射输出。

您创建并配置了 LA_Customers_tgt 数据对象来包含解析后的输出。您创建了一个映射用于解析数据。在此映射中，您配置了一个具有标志集的解析器转换，以从 Los Angeles 客户文件的 FullName 列中解析出名字和姓氏。您配置该映射将解析的数据写入 LA_Customers_tgt 数据对象中的 Firstname 和 Lastname 列。您还运行了配置文件，在运行映射之前查看转换的输出。最后，您运行了该映射，并使用数据查看器查看了 LA_Customers_tgt 数据对象中的新数据列。

第 14 章

课程 5。标准化数据

本章包括以下主题：

- [数据标准化概览, 68](#)
- [任务 1。创建目标数据对象, 69](#)
- [任务 2。创建映射以将数据标准化, 70](#)
- [任务 3。运行映射, 72](#)
- [任务 4。查看映射输出, 72](#)
- [数据标准化课程总结, 73](#)

数据标准化概览

数据标准化可以通过去除数据中的错误和不一致问题来改善数据质量。

要改善数据质量，请将包含以下类型值的数据标准化：

- 不正确的值
- 信息正确但格式不正确的值
- 要用于派生新信息的值

请使用标准创建器转换在数据中搜索这些值。您可以选择以下搜索操作类型之一：

- **文本。** 搜索您输入的自定义字符串。删除这些字符串或将它们替换为自定义文本。
- **引用表。** 搜索您选择的引用表中包含的字符串。删除这些字符串，或将它们替换为引用表条目或自定义文本。

例如，您可以配置标准创建器转换，以使用替换字符串 ST. 将包含自定义字符串 Street 和 St. 的地址数据标准化。标准创建器转换会将搜索词语替换为词语 ST.，并将结果写入新的数据列。

故事

HypoStores 需要将其客户地址数据标准化，以便所有地址使用统一的词语。All_Customers 数据对象中的地址数据对 Street、Boulevard、Avenue、Drive 和 Park 等通用词语的条目采用的格式不一致。

目标

在本课程中，您将完成以下任务：

- 创建并配置 All_Customers_Stdz_tgt 数据对象以用于包含标准化的数据。
- 创建映射将地址词语 Street、Boulevard、Avenue、Drive 和 Park 标准化为统一格式。
- 将 All_Customers 数据对象添加到映射，以连接源数据。

- 将 All_Customers_Stdz_tgt 数据对象添加到映射，以创建目标数据对象。
- 将标准创建器转换添加到映射，并配置该转换以便将地址词语标准化。
- 运行映射生成标准化的地址数据。
- 运行数据查看器查看映射输出。

先决条件

开始本课程之前，请确认满足以下先决条件：

- 已完成本教程的第 1 课和第 2 课。

计时

完成本课程需要 15 分钟。

任务 1。创建目标数据对象

在此任务中，您将创建一个可以向其中写入标准化数据的 All_Customers_Stdz_tgt 数据对象。

要创建目标数据对象，请完成以下步骤：

1. 基于 All_Customers.csv 文件创建 All_Customers_Stdz_tgt 数据对象。
2. 配置该数据对象的读取和写入选项，包括文件位置和文件名。

步骤 1。创建 All_Customers_Stdz_tgt 数据对象

在此步骤，您基于 All_Customers.csv 文件创建 All_Customers_Stdz_tgt 数据对象。

1. 单击**文件** > **新建** > **数据对象**。
此时将打开**新建**窗口。
2. 选择**平面文件数据对象**，然后单击**下一步**。
3. 确认已选择**从现有平面文件创建**。
4. 单击**浏览**并导航到以下目录中的 All_Customers.csv：<Informatica 安装目录>\clients\DeveloperClient\Tutorials
5. 单击**打开**。
6. 在名称字段中，输入 All_Customers_Stdz_tgt。
7. 单击**下一步**。
8. 单击**下一步**。
9. 在**预览选项**部分，选择**导入第一行中的列名称**，然后单击**下一步**。
10. 单击**完成**。

All_Customers_Stdz_tgt 数据对象将显示在编辑器中。

步骤 2。配置读取和写入选项

在此步骤，您配置 All_Customers_Stdz_tgt 数据对象的读取和写入选项，包括文件位置和文件名。

1. 确认 All_Customers_Stdz_tgt 数据对象已在编辑器中打开。
2. 在编辑器中，选择**读取**视图。

3. 单击**窗口** > **显示视图** > **属性**。
4. 在**属性**视图中，选择**运行时**视图。
5. 在**值**列中，双击源文件名并键入 All_Customers_Stdz_tgt.csv。
6. 在**值**列中，双击**源文件目录**条目。
7. 右键单击突出显示的名称并选择**复制**。
8. 在编辑器中，选择**写入**视图。
9. 在**属性**视图中，选择**运行时**视图。
10. 在**值**列中，双击**输出文件目录**条目。
11. 右键单击并选择**粘贴**以粘贴您从**读取**视图复制的目录位置。
12. 在**值**列中，双击**表头选项**条目并选择输出字段名称。
13. 在**值**列中，双击**输出文件名**条目并键入 All_Customers_Stdz_tgt.csv。
14. 单击**文件** > **保存**以保存该数据对象。

任务 2。创建映射以将数据标准化

在此任务中，您将创建一个映射并配置该映射，以使用数据对象和标准创建器转换。

要创建映射以将数据标准化，请完成以下步骤：

1. 创建映射。
2. 将源数据对象和目标数据对象添加到映射。
3. 将标准创建器转换添加到映射。
4. 配置标准创建器转换将通用地址词语标准化为一致的格式。

步骤 1。创建映射

在此步骤，您创建映射并为其命名。

1. 在**对象浏览器**视图中，选择您的教程项目。
2. 单击**文件** > **新建** > **映射**。
此时将打开**新建映射**窗口。
3. 在**名称**字段中，输入 StandardizerMapping。
4. 单击**完成**。
映射将在编辑器中打开。

步骤 2。将数据对象添加到映射

在此步骤，您将 All_Customers 数据对象和 All_Customers_Stdz_tgt 数据对象添加到映射。

1. 在**对象浏览器**视图中，浏览到教程项目中的数据对象。
2. 选择 All_Customers 数据对象并将其拖动到编辑器。
此时将打开**将物理数据对象添加到映射**窗口。
3. 确认已选中**读取**，然后单击**确定**。

数据对象将显示在编辑器中。

4. 在**对象浏览器**视图中，浏览到教程项目中的数据对象。
5. 选择 All_Customers_Stdz_tgt 数据对象并将其拖动到编辑器。
此时将打开**将物理数据对象添加到映射**窗口。
6. 选择**写入**，然后单击**确定**。
数据对象将显示在编辑器中。
7. 选择 All_Customers 数据对象中的所有端口。将这些端口拖动到 All_Customers_Stdz_tgt 数据对象中的 CustomerID 端口。

提示：按住 Shift 键可选择多个端口。您可能需要向下滚动端口列表以选择所有端口。

All_Customers 数据对象的端口将连接到 All_Customers_Stdz_tgt 数据对象中的对应端口。

步骤 3。将标准创建器转换添加到映射

在此步骤，您添加标准创建器转换以便将地址数据中的字符串标准化。

1. 选择包含 StandardizerMapping 映射的编辑器。
2. 在转换选项板中，选择标准创建器转换。
3. 单击编辑器。
一个名为 NewStandardizer 的标准创建器转换将显示在映射中。
4. 要重命名该标准创建器转换，双击转换的标题栏并键入 AddressStandardizer。
5. 选择 All_Customers 数据对象中的 Address1 端口，然后将其拖动到标准创建器转换的 **Input** 组。
一个名为 Address1 的端口将显示在输入组中。该端口连接到 All_Customers 数据对象中的 Address1 端口。

注意：您在配置标准化策略时将输出端口添加到该转换。

步骤 4。配置标准创建器转换

在此步骤，您配置标准创建器转换将源数据中的地址词语标准化。

注意：在此任务中，您将定义五个标准化操作。每个操作会将输入列中的一个字符串替换为一个新字符串。

1. 选择包含 StandardizerMapping 映射的编辑器。
2. 单击标准创建器转换。
3. 单击**窗口** > **显示视图** > **属性**。
4. 在“属性”视图中，选择**策略**。
5. 单击**新建**。此时将显示“新建策略”向导。
6. 单击“输入”列中的选择箭头，并选择 Address1 输入端口。
“输出”字段将显示 Address1 作为输出端口。
7. 选择字符空格和逗号分隔符 [\s] 和 [,]。您也可以选择性地选择选项删除尾随空格。
8. 单击**下一步**。
9. 选择**替换自定义字符串**操作，然后单击**下一步**。
10. 在**属性**下，单击**新建**。

11. 编辑**自定义字符串**和**替换为**字段，使其包含以下表中的第一对字符串：

自定义字符串	替换为
STREET	ST.
BOULEVARD	BLVD.
AVENUE	AVE.
DRIVE	DR.
PARK	PK.

12. 重复步骤 9 至 12 为表中所有字符串定义标准化操作。
13. 将 Address1 输出端口拖动到 All_Customers_Stdz_tgt 数据对象中的 Address1 端口。
14. 单击**文件** > **保存**以保存该映射。

任务 3。运行映射

在此任务中，您将运行映射以将标准化地址写入输出数据对象。

1. 选择包含 StandardizerMapping 映射的编辑器。
2. 单击**运行** > **运行映射**。
映射将运行，并将输出写入 All_Customers_Stdz_tgt.csv 文件。

任务 4。查看映射输出

在此任务中，您将运行数据查看器来查看映射输出，并验证地址数据是否已正确标准化。

1. 在**对象浏览器**视图中，找到教程项目中的 All_Customers_Stdz_tgt 数据对象并双击该数据对象。
数据对象将在编辑器中打开。
2. 单击**窗口** > **显示视图** > **数据查看器**。
此时将打开“数据查看器”视图。
3. 在“数据查看器”视图中，单击**运行**。
“数据查看器”视图将显示映射输出。
4. 验证 Address1 列是否显示正确标准化的数据。例如，字符串 STREET 的所有实例均应替换为字符串 ST。

数据标准化课程总结

在本课程中，您学习了可以将数据标准化，以去除数据中的错误和不一致问题。

您学习了可以使用标准创建器转换将输出列中的字符串标准化。您还学习了可以使用数据查看器查看映射输出。

您创建并配置了 `All_Customers_Stdz_tgt` 数据对象来包含标准化的输出。您创建了一个映射来将数据标准化。在此映射中，您配置了一个标准创建器转换来将 `All_Customers` 数据对象中的 `Address1` 列标准化。您配置了该映射将标准化的输出写入 `All_Customers_Stdz_tgt` 数据对象。最后，您运行了该映射，并使用数据查看器查看了 `All_Customers_Stdz_tgt` 数据对象中的标准化数据。

第 15 章

课程 6。验证地址数据

本章包括以下主题：

- [验证地址数据概览, 74](#)
- [任务 1。创建目标数据对象, 75](#)
- [任务 2。创建映射以验证地址, 77](#)
- [任务 3。配置地址验证器转换, 78](#)
- [任务 4。运行映射, 80](#)
- [任务 5。查看映射输出, 81](#)
- [验证地址数据课程总结, 83](#)

验证地址数据概览

地址验证是一个评估并改进通信地址质量的过程。它通过比较输入地址与有效地址的引用数据集来评估地址质量。它可以发现不正确的地址值，并使用引用数据集创建包含正确值的地址，从而提高地址质量。

有效地址是指可送达的地址。地址可能格式正确并包含真实的街道、城市和邮政编码信息，但是如果这些数据无法生成可送达的地址，则该地址是无效的。Developer 工具可使用地址引用数据集检查输入地址的可送达性。Informatica 提供地址引用数据集。

地址引用数据集包含描述了一个国家/地区所有可送达地址的数据。地址验证过程会在引用数据集中搜索与输入地址数据最接近的地址。如果该过程在引用数据集中发现近似匹配，将为任何不正确或不完整的数据值写入新值。该过程将创建一组字母数字代码，以描述在输入地址和引用地址之间找到的匹配类型。它还可以重新构造地址，并可添加输入地址中没有的信息，例如美国地址的四位数邮政编码后缀。

使用地址验证器转换可在 Developer 工具中构建地址验证过程。这种多组转换包含一系列与输入地址中所有可能的字段相对应的预定义输入端口和输出端口。配置地址验证器转换时，您需要选择默认的引用数据集，并使用转换端口创建输入和输出地址结构。在本课程中，您将配置该转换来验证美国地址数据。

故事

HypoStores 需要更正和完善地址数据，以确保直邮推广活动和其他消费者邮寄品可以送达至其客户。更正和完善地址数据还可降低该公司的邮寄成本。此外，HypoStores 还需要在客户资料中以一种足以灵活包含各种长度地址的可打印格式来包含地址信息。

为满足这些业务要求，HypoStores 的 ICC 团队在 Developer 工具中创建了一个地址验证映射。

目标

在本课程中，您将完成以下任务：

- 创建一个将用于包含已验证的地址字段和匹配代码的目标数据对象。
- 创建一个具有源数据对象、目标数据对象和地址验证器转换的映射。
- 配置地址验证器转换以验证客户的地址数据。
- 运行映射以验证地址数据，并查看匹配代码输出确认地址数据的有效性。

先决条件

开始本课程之前，请验证以下先决条件：

- 已完成本教程的第 1 课和第 2 课。
- 已在域中安装了美国地址引用数据并向 Administrator 工具注册了该数据。联系 Informatica 管理员确认您的系统上已安装了美国地址数据。引用数据通过 Data Quality 内容安装程序进行安装。

计时

完成本课程需要 25 分钟。

任务 1。创建目标数据对象

在此任务中，您将创建目标数据对象、配置写入选项并添加端口。

要创建并配置目标数据对象，请完成以下步骤：

1. 基于 All_Customers.csv 文件创建 All_Customers_av_tgt 数据对象。
2. 配置该数据对象的读取和写入选项，包括文件位置和文件名。
3. 将端口添加到该数据对象，以接收地址验证器转换生成的匹配代码值。

步骤 1。创建 All_Customers_av_tgt 数据对象

在此步骤，您基于 All_Customers.csv 文件创建 All_Customers_av_tgt 数据对象。

1. 单击**文件** > **新建** > **数据对象**。
此时将打开**新建**窗口。
2. 选择**平面文件数据对象**，然后单击**下一步**。
3. 确认已选择**从现有平面文件创建**。单击此选择旁边的**浏览**，找到 All_Customers.csv 文件并单击**打开**。
4. 在名称字段中，输入 All_Customers_av_tgt。
5. 单击**下一步**。
6. 单击**下一步**。
7. 在**预览选项**部分，选择**导入第一行中的列名称**，然后单击**下一步**。
8. 单击**完成**。

All_Customers_av_tgt 数据对象将显示在编辑器中。

步骤 2。配置读取和写入选项

在此步骤，您配置 All_Customers_av_tgt 数据对象的读取和写入选项，包括目标文件名和位置。

1. 确认 All_Customers_av_tgt 数据对象已在编辑器中打开。
2. 在编辑器中，选择**读取**视图。
3. 选择**窗口 > 显示视图 > 属性**。
4. 在**属性**视图中，选择**运行时**视图。
5. 在**值**列中，双击源文件名并键入 All_Customers_av_tgt.csv。
6. 在**值**列中，双击鼠标突出显示源文件目录路径。
7. 右键单击突出显示的路径和名称，然后选择**复制**。
8. 在编辑器中，选择**写入**视图。
9. 在**属性**视图中，选择**运行时**视图。
10. 在**值**列中，双击**输出文件目录**条目。
11. 右键单击此条目并选择**粘贴**，添加您从**读取**视图中复制的路径。
12. 在**值**列中，双击**表头选项**条目并选择输出字段名称。
13. 在**值**列中，双击**输出文件名**条目并键入 All_Customers_av_tgt.csv。
14. 选择**文件 > 保存**以保存该数据对象。

步骤 3。将端口添加到数据对象

在此步骤，您将两个端口添加到 All_Customers_av_tgt 数据对象，以便地址验证器转换可以将匹配代码值写入目标文件。请将端口命名为 MailabilityScore 和 MatchCode。

MailabilityScore 的值描述输入地址的可送达性。MatchCode 的值描述转换在输入地址与引用数据地址之间进行的匹配类型。

1. 在**对象浏览器**视图中，浏览到教程项目中的数据对象。
2. 双击 All_Customers_av_tgt 数据对象。
All_Customers_av_tgt 数据对象将在编辑器中打开。
3. 确认已选中**概览**。
4. 选择端口列表中的最后一个端口。此端口命名为 MiscDate。
5. 单击**新建**。
此时将显示一个命名为 MiscDate1 的端口。
6. 将 MiscDate1 端口重命名为 MailabilityScore。
7. 选择 MailabilityScore 端口。
8. 单击**新建**。
此时将显示一个命名为 MailabilityScore1 的端口。
9. 将 MailabilityScore1 端口重命名为 MatchCode。
10. 单击**文件 > 保存**以保存该数据对象。

任务 2。创建映射以验证地址

在此任务中，您将创建一个映射并添加数据对象和地址验证器转换。

要创建映射并添加需要的对象，请完成以下步骤：

1. 创建映射对象。
2. 将源数据对象和目标数据对象添加到映射。
3. 将地址验证器转换添加到映射。

步骤 1。创建映射

在此步骤，您创建映射并为其命名。

1. 在“对象浏览器”视图中，选择教程项目。
2. 选择**文件** > **新建** > **映射**。
此时将打开**新建映射**窗口。
3. 在**名称**字段中，输入 ValidateAddresses。
4. 单击**完成**。
映射将在编辑器中打开。

步骤 2。将数据对象添加到映射

在此步骤，您将源数据对象和目标数据对象添加到映射。

All_Customers 是映射的源数据对象。地址验证器转换会从此对象中读取数据。All_Customers_av_tgt 是映射的数据目标对象。此对象从地址验证器转换读取数据。

1. 在**对象浏览器**视图中，浏览到教程项目中的数据对象。
2. 选择 All_Customers 数据对象并将其拖动到编辑器。
此时将打开**将物理数据对象添加到映射**窗口。
3. 确认已选中**读取**，然后单击**确定**。
数据对象将显示在编辑器中。
4. 在**对象浏览器**视图中，浏览到教程项目中的数据对象。
5. 选择 All_Customers_av_tgt 数据对象并将其拖动到编辑器。
此时将打开**将物理数据对象添加到映射**窗口。
6. 选择**写入**，然后单击**确定**。
数据对象将显示在编辑器中。
7. 单击**保存**。

步骤 3。将地址验证器转换添加到映射

在此步骤，您将地址验证器转换添加到包含源和数据对象的映射中。

完成此步骤后，您可以配置该转换并将其端口连接到数据对象。

1. 选择包含 ValidateAddresses 映射的编辑器。
2. 在转换选项板中，选择地址验证器转换。

3. 单击编辑器。
地址验证器转换将显示在编辑器中。

任务 3。配置地址验证器转换

在此任务中，您将配置地址验证器转换，以便从 All_Customers 数据源读取并验证地址。

注意：地址验证器转换包含一系列预定义的输入和输出端口。请选择您需要的端口并将它们连接到映射中的对象。

要配置转换，请完成以下步骤：

1. 选择用于地址验证的默认国家/地区。
2. 配置转换输入端口。
3. 配置转换输出端口。
4. 将不使用的源端口连接到数据目标。

步骤 1。设置用于地址验证“默认国家/地区”

在此步骤中，选择用于地址验证的默认国家/地区。地址验证器转换使用的地址引用数据文件按国家/地区组织。选择默认国家/地区时，您需要标识转换应用到任何不含国家/地区信息输入地址的国家/地区数据集。

1. 在编辑器中选择地址验证器转换。
2. 在**属性**下，单击**常规设置**。
3. 在**默认国家/地区**菜单中，选择美国。

步骤 2。配置地址验证器转换输入端口

在此步骤，您选择转换的输入端口，并将这些端口连接到 All_Customers_av 数据对象。

地址验证器转换包含几组预定义的输入端口。请选择与您的输入地址中的字段相对应的输入端口，然后将这些端口添加到转换中。

在以下步骤中选择端口时按照 Ctrl 键可一次选择多个端口。

1. 在编辑器中选择地址验证器转换。
2. 在**属性**下，单击**模板**。
3. 展开**基本模型**端口组。
4. 展开**混合**输入端口组并选择以下端口：

端口名称	说明
收件人地址行 1	街道地址数据，如街道名称和建筑物编号。
区域填写 1	城市或乡镇名称。
邮政编码 1	邮政编码或邮政区码。
省/市/自治区 1	省/市/自治区或州的名称

端口名称	说明
国家/地区名称	国家/地区的名称或缩写。

注意：按住 Ctrl 键可一次选择多个端口。

- 在端口名称列表上方的工具栏上，单击**将端口添加到转换**。

此工具栏将在您选择**模板**时显示。

选定的端口将显示在映射编辑器中的转换中。

- 按以下方式将源端口连接到地址验证器转换端口：

源端口	地址验证器转换端口
地址 1	收件人地址行 1
城市	区域填写 1
邮政编码	邮政编码 1
省/自治区/直辖市	省/市/自治区 1
国家/地区	国家/地区名称

步骤 3。配置地址验证器转换的输出端口

在此步骤，您选择转换器的输出端口，并将这些端口连接至 All_Customers_av_tgt 数据对象。

地址验证器转换包含几组预定义的输出端口。请选择可以定义您所需的地址结构的端口，然后将这些端口添加到转换中。

您也可以选择包含要为每个地址执行的验证类型信息的端口。

- 在映射编辑器中选择地址验证器转换。
- 在**属性**下，单击**模板**。
- 展开**基本模型**端口组。
- 展开**地址元素**输出端口组并选择以下端口：

端口名称	说明
街道填写 1	街道地址数据，如街道名称和建筑物编号。

- 展开**末行元素**输出端口组并选择以下端口：

端口名称	说明
区域填写 1	城市或乡镇名称。
邮政编码 1	邮政编码或邮政区码。
省/市/自治区缩写 1	省/市/自治区或州的标识符。

注意：按住 Ctrl 键可一次选择多个端口。

6. 展开**国家/地区**输出端口组并选择以下端口：

端口名称	说明
国家/地区名称 1	国家/地区名称。

7. 展开**状态信息**输出端口组并选择以下端口：

端口名称	说明
可邮寄得分	表示邮政递送成功几率的得分。
匹配代码	表示输入地址与引用数据之间相似性程度的代码。

8. 在端口名称列表上方的工具栏上，单击**将端口添加到转换**。

此工具栏将在您选择**模板**时显示。

9. 按以下方式将地址验证器转换端口连接到 All_Customers_av_tgt 端口：

地址验证器转换端口	目标端口
街道填写 1	地址 1
区域填写 1	城市
邮政编码 1	邮政编码
省/市/自治区缩写 1	省/自治区/直辖市
国家/地区名称 1	国家/地区
可邮寄得分	MailabilityScore
匹配代码	MatchCode

步骤 4。将不使用的数据源端口连接到数据目标

在此步骤，您将 All_Customers 数据源上不使用的端口连接到数据目标。

- ▶ 将数据源上不使用的端口连接到数据目标上的同名端口。

任务 4。运行映射

在此任务中，您将运行映射来创建映射输出。

1. 选择包含 ValidateAddresses 映射的编辑器。
2. 选择**运行** > **运行映射**。

映射将运行，并将输出写入到 All_Customers_av_tgt.csv 文件。

任务 5。查看映射输出

在此任务中，您将运行数据查看器来查看映射输出。请检查写入目标数据对象中的“可邮寄得分”和“匹配代码”列的值，查看您验证的地址的质量。

“匹配代码”值是一个字母数字代码，代表映射对地址执行的验证类型。

“可邮寄得分”值是一个概括地址可达性的个位数值。

1. 在“对象浏览器”视图中，找到教程项目中的 All_Customers_av_tgt 数据对象并双击该数据对象。
数据对象将在编辑器中打开。

2. 选择窗口 > 显示视图 > 数据查看器。

此时将打开数据查看器。

3. 在数据查看器中，单击运行。

“数据查看器”视图将显示映射输出。

4. 滚动浏览映射结果，使“可邮寄得分”和“匹配代码”列可见。

5. 查看“可邮寄得分”列中的值。

该得分的范围可以从 0 到 5。得分越高的地址越有可能成功送达。

6. 查看“匹配代码”列中的值。

匹配代码是一个字母数字代码。字母字符部分表示转换执行的验证类型，数字部分表示最终地址的质量。

下表说明了匹配代码的值：

代码	说明
A1	地址代码查找过程为输入代码发现了不完整地址或完整地址。
A0	地址代码查找过程没有为输入代码发现地址。
C4	已更正。已检查所有与邮政相关的元素。
C3	已更正。无法检查部分元素。
C2	已更正，但是由于缺少引用数据，送达状态未知。
C1	已更正，但是由于用户标准化产生了错误，送达状态未知。
I4	无法完全更正数据，但是存在一个与引用数据中的地址匹配的匹配项。
I3	无法完全更正数据，但是存在多个与引用数据中的地址匹配的匹配项。
I2	无法更正数据。批处理模式返回部分建议的地址。
I1	无法更正数据。批处理模式无法建议地址。
N7	验证错误。由于单行验证未解锁，因此未进行验证。
N6	验证错误。由于目标国家/地区不支持单行验证，因此未进行验证。
N5	验证错误。由于引用数据库已过期，因此未进行验证。

代码	说明
N4	验证错误。由于引用数据已损坏或格式化错误，因此未进行验证。
N3	验证错误。由于国家/地区数据无法解锁，因此未进行验证。
N2	验证错误。由于所需的引用数据库不可用，因此未进行验证。
N1	验证错误。由于国家/地区无法识别或不受支持，因此未进行验证。
Q3	建议列表模式。地址验证可以从地址引用数据中检索到一个或多个与输入地址对应的完整地址。
Q2	建议列表模式。地址验证可以合并输入地址的元素和地址引用数据的元素来生成完整地址。
Q1	建议列表模式。地址验证无法建议完整地址。要生成完整地址建议，请向输入地址中添加数据。
Q0	建议列表模式。用于生成建议的输入数据不足。
RB	通过缩写识别了国家/地区。识别 ISO 双字符和 ISO 三个字符的国家/地区代码。也可识别常见缩写，如“GER”代表德国。
RA	通过转换中的“强制国家/地区”设置识别了国家/地区。
R9	通过转换中的“默认国家/地区”设置识别了国家/地区。
R8	通过国家/地区名称识别了国家/地区。
R7	通过国家/地区名称识别了国家/地区，但是转换在国家/地区数据中发现了错误。
R6	通过领土数据识别了国家/地区。
R5	通过省/市/自治区数据识别了国家/地区。
R4	通过主要城镇数据识别了国家/地区。
R3	通过地址格式识别了国家/地区。
R2	通过语言字母识别了国家/地区。
R1	由于存在多个匹配，未识别国家/地区。
R0	未识别国家/地区。
S4	解析模式。地址已完全解析。
S3	解析模式。解析地址后有多个结果。
S1	解析模式。由于输入格式不匹配，出现解析错误。

代码	说明
V4	已验证。输入数据正确。地址验证已检查所有与邮政相关的元素，且输入完全匹配。
V3	已验证。输入数据正确，但部分或所有元素已标准化，或者输入中包含的名称或外来语已过期。
V2	已验证。输入数据正确，但是由于引用数据不完整，无法验证某些元素。
V1	已验证。输入数据正确，但是用户标准化对可送达性造成了负面影响。例如，邮政编码长度过短。

验证地址数据课程总结

在本课程中，您学习了地址验证会比较输入地址数据与引用数据，并返回尽可能最准确版本的地址。

您学习了地址验证过程还会返回有关每个地址质量的状态信息。

另外，您学习了 Administrator 工具用户可以运行 Data Quality 内容安装程序来安装地址引用数据。

还了解到地址验证器转换是一个多组转换，并且您要从端口组中为转换选择输入和输出端口。您选择的输入端口将决定验证的地址内容。输出端口则决定最终地址记录的内容。

附录 A

常见问题

本附录包括以下主题：

- [Informatica Analyst 常见问题解答, 84](#)
- [Informatica Developer 常见问题, 84](#)

Informatica Analyst 常见问题解答

请查看常见问题以便解答您对 Informatica Analyst 可能存在的疑问。

能否使用一个用户帐户访问 Administrator 工具、Developer 工具和 Analyst 工具？

能。您可以授予一个用户访问全部这三个工具的权限。不需要为每个客户端应用程序创建单独的用户帐户。

引用数据存储在哪里？

您可以使用 Developer 工具和 Analyst 工具创建和共享引用数据对象。模型存储库存储引用数据对象的元数据。引用数据数据库存储引用表的数据值。请在内容管理服务中配置引用数据数据库。

Informatica Developer 常见问题

请查看常见问题以便解答您对 Informatica Developer 可能存在的疑问。

PowerCenter 中的源和目标与 Developer 工具中的物理数据对象之间有何不同？

在 PowerCenter 中，您创建源定义作为映射的源，并创建目标定义作为映射的目标。而在 Developer 工具中，您创建物理数据对象来用作映射的源或目标。

Developer 工具中的映射与 PowerCenter 中的映射之间有何不同？

PowerCenter 映射指定如何在源与目标之间移动数据。Developer 工具映射则指定如何在映射输入与输出之间移动数据。

PowerCenter 映射必须包含一个或多个源定义、源限定符和目标定义。PowerCenter 映射还可以包含快捷方式、转换和 Mapplet。

Developer 工具映射必须包含映射输入和输出。Developer 工具映射还可以包含转换和 Mapplet。

Developer 工具具有以下类型的映射：

- 在源与目标之间移动数据的映射。此类型的映射与 PowerCenter 映射唯一的不同在于它不能使用快捷方式，并且不使用源限定符。

- 逻辑数据对象映射。逻辑数据对象模型中的一种映射。逻辑数据对象映射可以包含一个逻辑数据对象作为映射输入，以及一个数据对象作为映射输出。或者，可以包含一个或多个物理数据对象作为映射输入，以及一个逻辑数据对象作为映射输出。
- 虚拟表映射。SQL 数据服务中的一种映射。该映射包含一个数据对象作为映射输入，以及一个虚拟表作为映射输出。
- 虚拟存储过程映射。这种映射在 SQL 数据服务中定义一组业务逻辑。它包含一个输入参数转换或物理数据对象作为映射输入，以及一个输出参数转换或物理数据对象作为映射输出。

PowerCenter 中的 Mapplet 与 Developer 工具中的 Mapplet 之间有何不同？

PowerCenter 与 Developer 工具中的 Mapplet 是一种包含一组转换的可重用对象。您可以在多个映射中重用转换逻辑。

PowerCenter Mapplet 可以包含源定义或输入转换作为 Mapplet 输入。它必须包含输出转换作为 Mapplet 输出。

Developer 工具的 Mapplet 可以包含数据对象或输入转换作为 Mapplet 输入。并可以包含数据对象或输出转换作为 Mapplet 输出。Developer 工具中的映射还具有以下特征：

- 您可以将 Mapplet 验证为规则。您在配置文件中使用的规则。
- Mapplet 可以包含其他 Mapplet。

Mapplet 与规则之间有何不同？

您可以将 Mapplet 验证为规则。规则是一种业务逻辑，定义了运行配置文件时对源数据应用的条件。当 Mapplet 符合以下要求时，您可以将 Mapplet 验证为规则：

- 它包含一个输入和输出转换。
- Mapplet 不包含活动转换。
- 它不指定输入组之间的基数。

索引

C

创建表达式规则

概览 [31](#)

创建结果卡

概览 [34](#)

创建默认配置文件

概览 [25](#)

创建数据对象

概览 [22](#)

创建自定义配置文件

概览 [28](#)

从列创建引用表

概览 [39](#)

D

导入物理数据对象

概览 [49](#)

P

剖析数据

概览 [58](#)

S

设置 Analyst 工具

概览 [19](#)

设置 Developer tool

概览 [45](#)

Y

引用表

概览 [42](#)