



Informatica®
10.4.0

剖析入门指南

Informatica 剖析入门指南

10.4.0

2019 年 12 月

© 版权所有 Informatica LLC 2010, 2020

本软件和文档仅根据包含使用与披露限制的单独许可协议提供。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）复制或传播本文档的任何部分。

Informatica 和 Informatica 标志是 Informatica LLC 在美国和世界其他许多司法管辖区的商标或注册商标。欲获得 Informatica 商标的最新列表，请访问 <https://www.informatica.com/trademarks.html>。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。

美国政府权利交付给美国政府客户的程序、软件、数据库及相关文档和技术数据是指适用的联邦采购条例和政府机构特定补充条例中定义的"商业计算机软件"或"商业技术数据"。因此，使用、复制、披露、修改和改编应遵循适用的政府合同中规定的限制和许可条款、政府合同条款的适用范围以及 FAR 52.227-19 商用计算机软件许可中规定的额外权利。

本软件和/或文档中的若干部分受第三方版权约束。所需的第三方声明随产品一起提供。

本文档中的信息如有更改，恕不另行通知。如发现本文档中有什么问题，请通过以下电子邮件地址向我们报告：infa_documentation@informatica.com。

Informatica 产品根据对应协议的条款和条件进行担保。INFORMATICA 按"原样"提供本文档中的信息，无任何明示或暗示的担保，包括但不限于任何适销性和特定用途适用性担保，也没有任何非侵权担保或条件。

发布日期: 2020-02-04

目录

前言	6
Informatica 资源	6
Informatica Network	6
Informatica 知识库	6
Informatica 文档	6
Informatica 产品可用性矩阵	7
Informatica Velocity	7
Informatica Marketplace	7
Informatica 全球客户支持部门	7
第 1 章：入门概览	8
剖析概览	8
介绍 Informatica Analyst	9
Informatica Analyst 用户界面	9
Informatica Developer 用户界面	11
Informatica Developer 欢迎页	12
备忘单	12
Informatica Developer 任务	13
教程讲解	13
教程结构	13
教程先决条件	14
Informatica Analyst 教程	14
Informatica Developer 教程	15
第 I 部分：Informatica Analyst 入门	16
第 2 章：课程 1。设置 Informatica Analyst	17
设置 Informatica Analyst 概览	17
任务 1. 登录 Informatica Analyst	18
任务 2. 创建项目	18
任务 3. 创建文件夹	18
设置 Informatica Analyst 摘要	19
第 3 章：课程 2。创建数据对象	20
创建数据对象概览	20
任务 1. 创建平面文件数据对象	21
任务 2. 查看数据对象属性	21
创建数据对象摘要	22

第 4 章： 课程 3. 创建默认配置文件.....	23
创建默认配置文件概览.....	23
任务 1. 创建并运行默认配置文件.....	24
任务 2. 在摘要视图中查看配置文件结果.....	24
任务 3. 在详细视图中查看配置文件结果.....	25
创建默认配置文件摘要.....	25
第 5 章： 课程 4。 创建自定义配置文件.....	26
创建自定义配置文件概览.....	26
任务 1. 创建自定义配置文件.....	27
任务 2. 运行配置文件.....	27
任务 3. 向下钻取配置文件结果.....	28
创建自定义配置文件课程总结.....	28
第 6 章： 课程 5。 创建表达式规则.....	29
创建表达式规则概览.....	29
任务 1. 创建表达式规则并运行配置文件.....	30
任务 2. 查看表达式规则的输出.....	30
任务 3. 编辑表达式规则.....	31
创建表达式规则课程总结.....	31
第 7 章： 课程 6。 创建并运行结果卡.....	32
创建并运行结果卡概览.....	32
任务 1. 使用配置文件结果创建结果卡.....	33
任务 2. 运行结果卡.....	34
任务 3. 查看结果卡.....	34
任务 4. 编辑结果卡.....	35
任务 5. 配置阈值.....	35
任务 6. 查看得分趋势图表.....	35
创建并运行结果卡课程总结.....	36
第 8 章： 课程 7。 使用配置文件列创建引用表.....	37
使用配置文件列创建引用表概览.....	37
任务 1. 使用配置文件列创建引用表.....	38
任务 2. 编辑引用表.....	38
使用配置文件列创建引用表课程总结.....	39
第 9 章： 课程 8。 创建引用表.....	40
创建引用表概览.....	40
任务 1. 创建引用表.....	41
创建引用表课程总结.....	41

第 II 部分：Informatica Developer 入门.....	42
第 10 章：课程 1。设置 Informatica Developer.....	43
设置 Informatica Developer 概览.....	43
任务 1。启动 Informatica Developer.....	44
任务 2。添加域.....	44
任务 3。添加模型存储库.....	45
任务 4。创建项目.....	45
任务 5。创建文件夹.....	45
任务 6。选择默认数据集成服务.....	46
设置 Informatica Developer 课程总结.....	46
第 11 章：课程 2：导入物理数据对象.....	47
导入物理数据对象概览.....	47
任务 1。导入 Boston_Customers 平面文件数据对象.....	48
任务 2。导入 LA_Customers 平面文件数据对象.....	54
任务 3。导入 All_Customers 平面文件数据对象.....	54
导入物理数据对象课程总结.....	55
第 12 章：课程 3。对源数据运行配置文件.....	56
剖析数据概览.....	56
任务 1。对两个数据源执行联接分析.....	57
任务 2。查看联接分析结果.....	58
任务 3。对数据源运行配置文件.....	58
任务 4。查看列剖析结果.....	58
剖析数据课程总结.....	59
索引.....	60

前言

要学习如何在 Developer tool 和 Analyst 工具中创建和运行配置文件，请按《剖析入门指南》中的说明操作。该指南作为一本教程，有助于初次使用此软件的用户学习如何使用 *Informatica Developer* 和 *Informatica Analyst*。

本指南面向数据分析师和开发人员。

Informatica 资源

Informatica 通过 Informatica Network 和其他在线门户为您提供一系列产品资源。使用这些资源，可以充分利用 Informatica 产品和解决方案，并向其他 Informatica 用户和主题专家学习。

Informatica Network

在 Informatica Network 中可以获得许多资源，包括 Informatica 知识库和 Informatica 全球客户支持。要进入 Informatica Network，请访问 <https://network.informatica.com>。

作为 Informatica Network 成员，您可以选择以下服务：

- 在知识库中搜索产品资源。
- 查看产品可用性信息。
- 创建并检查您的支持案例。
- 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 知识库可查找产品资源，例如操作方法文章、最佳实践、视频教程以及常见问题的答案。

要搜索知识库，请访问 <https://search.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

使用 Informatica 文档门户可浏览大量当前与最近产品版本的文档库。要浏览文档门户，请访问 <https://docs.informatica.com>。

如果您对产品文档有任何疑问、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、数据库以及数据源和目标的类型。您可以在以下网址中浏览 Informatica PAM:

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

Informatica Velocity

Informatica Velocity 是由 Informatica 专业服务根据数百个数据管理项目的实际经验所开发出来的，其中汇集了大量使用技巧和最佳实践。Informatica Velocity 代表了 Informatica 顾问的集体知识，这些顾问与世界各地的组织合作，共同计划、开发、部署和维护成功的数据管理解决方案。

您可以在以下网址中找到 Informatica Velocity 资源：<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扩展和增强您的 Informatica 实施。利用 Informatica 开发人员和合作伙伴在 Marketplace 中提供的数以百计的解决方案，可提高您的工作效率并加快项目实施时间。您可以在以下网址中找到 Informatica Marketplace：<https://marketplace.informatica.com>。

Informatica 全球客户支持部门

您可以通过电话或 Informatica Network 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：<https://www.informatica.com/services-and-training/customer-success-services/contact-us.html>。

要在 Informatica Network 上查找在线支持资源，请访问 <https://network.informatica.com>，然后选择 eSupport 选项。

第 1 章

入门概览

本章包括以下主题：

- [剖析概览, 8](#)
- [介绍 Informatica Analyst, 9](#)
- [Informatica Developer 用户界面, 11](#)
- [教程讲解, 13](#)
- [教程结构, 13](#)

剖析概览

使用剖析查找应用程序、架构或企业的数据源的内容、质量和结构。数据源内容包括值频率和数据类型。数据源结构包括键和功能相关性。

在发现过程中，可以创建并运行配置文件。配置文件是一个存储库对象，可查找和分析企业内数据源中的所有数据不规范问题和使数据项目处于危险中的隐藏数据问题。通过在企业内的任何数据源中运行配置文件，可以很好地了解企业数据和元数据的优势和劣势。

可以使用 Informatica Analyst 和 Informatica Developer 分析源数据和元数据。分析人员和开发人员可以使用这些工具进行协作、识别数据质量问题以及分析数据关系。您可以根据自己的职位选择使用 Analyst 工具或 Developer tool 的功能。您可以执行的剖析度因使用工具的不用而有所不同。

您可以在 Developer tool 和 Analyst 工具中执行以下任务：

- 执行列剖析。该过程包括发现列中唯一值、空值和数据模式的数量。
- 执行数据域发现。您可以发现企业内的关键数据特性。
- 管理配置文件结果，包括数据类型、数据域、主键和外键。
- 创建结果卡以监视数据质量。
- 选择一个操作系统配置文件，然后根据在操作系统配置文件中定义的操作系统用户的权限，创建并运行列配置文件、企业发现配置文件以及结果卡。
- 使用存储库资产锁定可防止其他用户覆盖所做的工作。
- 使用版本控制系统可保存配置文件的多个版本。
- 创建标记并将其分配给数据对象。
- 在 Business Glossary Desktop 中查找对象名称作为业务术语的含义。例如，可以查找列名或配置文件名称的含义，以了解其业务要求和当前的实现。

可以在 Developer tool 中执行以下任务：

- 发现数据源中两个数据列之间的潜在联接度。
- 确定一个或多个数据源内列中成对重叠数据的百分比。
- 比较列剖析的结果。
- 从配置文件中生成映射对象。
- 发现数据源中的主键。
- 发现一个或多个数据源中的外键。
- 发现数据源中各列之间的功能相关性。
- 对多个连接中的大量数据源运行数据发现任务。数据发现任务包括列配置文件、主键和外键关系推理、数据域发现以及生成数据关系的合并图形摘要。

可以在 Analyst 工具中执行以下任务：

- 对多个连接中的大量数据源执行企业发现。可以查看列元数据和数据域的合并的发现结果摘要。
- 执行发现搜索以查找数据和元数据在企业中存在的位置。您可以搜索特定资产，例如数据对象、规则和配置文件。发现搜索查找资产并识别与数据库和企业架构中的其他资产之间的关系。
- 查看最新配置文件运行的配置文件结果。
- 比较列配置文件中两次配置文件运行的配置文件结果。
- 查看每个结果卡度和度量组的结果卡沿袭。
- 查看结果卡仪表板。
- 向配置文件或配置文件中的列添加注释。
- 为配置文件或配置文件中的列分配标记。
- 为配置文件中的列分配业务术语。

介绍 Informatica Analyst

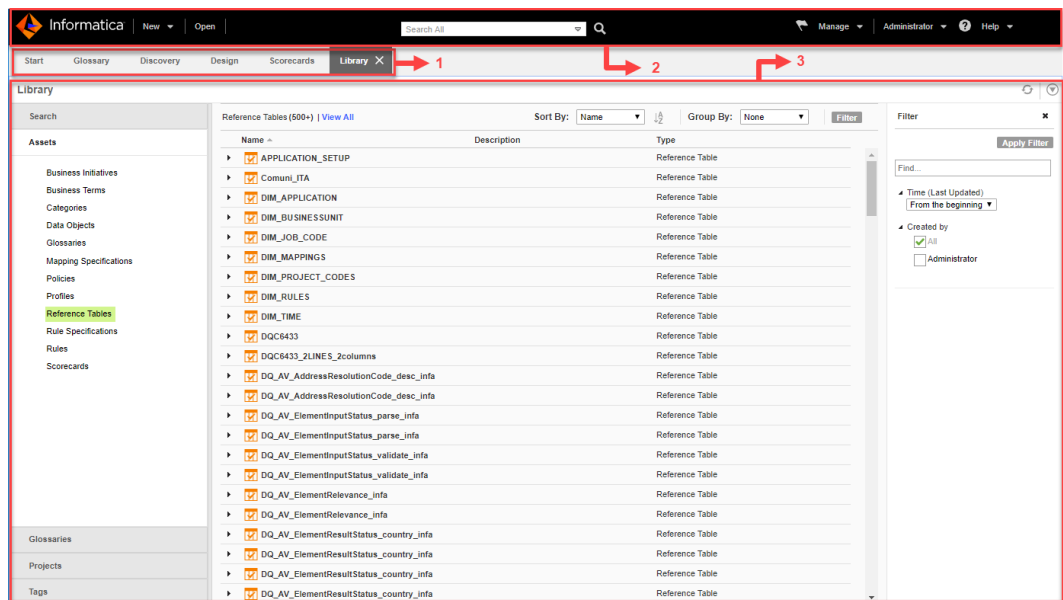
Informatica Analyst 是一个基于 Web 的应用程序客户端，分析人员可使用该客户端分析、清理、标准化、剖析企业内的数据，并为数据评分。

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 Analyst 用户界面

Analyst 工具用户界面包含多个工作区，可用于创建数据集成和数据质量数据联合解决方案。工作区是一个网页，可以根据许可证和特权进行访问。

下图显示了 Analyst 工具用户界面。



1. 工作区访问面板
2. 表头区域
3. 工作区

Analyst 工具会在**启动**工作区上打开。**开始**工作区会列出您具有相应许可证和特权可通过工作区访问面板使用的工作区。

Analyst 工具会显示以下表头菜单：

新建

在**词汇表**、**发现**和**设计**工作区中创建资产。

打开

打开**库**工作区。

通知警告

查看词汇表资产的通知。

管理

打开临时工作区和**通知**。您可以打开**连接**、**数据域**、**作业状态**、**项目**和 **Business Glossary** 安全工作区。

用户名

设置用户首选项以更改密码并注销 Analyst 工具。

帮助

访问当前工作区中的帮助。

Analyst 工具包含以下永久工作区：

启动

访问您有相应许可证可通过工作区访问面板访问的其他工作区。如果您有许可证来执行异常管理，则任务会显示在工作区的**我的任务**面板上。

词汇表

定义和描述对您的组织非常重要的业务概念。您可以创建和管理业务术语、类别、词汇表和策略。

发现

分析源系统中数据和元数据的质量。您可以创建和管理配置文件、平面文件数据对象和表数据对象。

设计

帮助分析师和开发人员进行协作的设计业务逻辑。您可以创建和管理映射规范、引用表和规则定义。

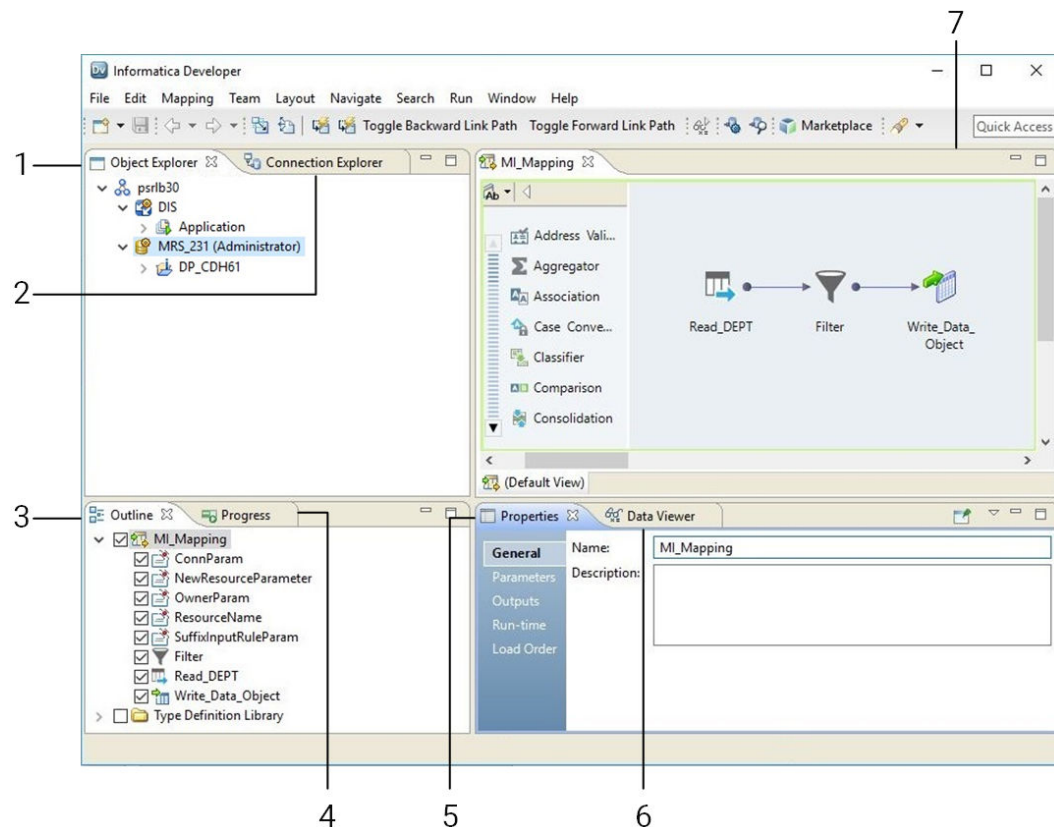
结果卡

打开、编辑和运行通过配置文件结果创建的结果卡。您可以添加度量，在列上向下钻取，添加结果卡筛选器以及查看结果卡的趋势图表。

Informatica Developer 用户界面

Developer 工具用户界面包含多视图工作台，可用于创建数据集成和数据质量 数据联合 解决方案。

下图显示了 Developer 工具用户界面：



1. “对象浏览器” 视图
2. “大纲” 视图
3. “属性” 视图
4. “数据查看器” 视图
5. “标记” 视图
6. 对象相关性视图
7. “警告” 视图
8. “连接浏览器” 视图
9. 编辑器

Developer 工具可以显示以下视图：

“对象浏览器”视图

显示域以及域中的设计时对象和运行时对象。设计时对象存储在模型存储库中的项目和文件夹中。运行时对象在数据集成服务上存储为运行时应用程序的一部分。

“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。还可以预览 Web 服务消息或者从“数据查看器”视图运行 SQL 查询。

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 工具工作台上的其他位置。单击**窗口 > 显示视图**选择要显示的视图。

Informatica Developer 欢迎页

第一次打开 Developer tool 时会显示欢迎页。使用欢迎页可以了解如何设置和开始使用 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”按钮可获取 Informatica 知识库链接。您可以访问 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。
- 工作台。单击“工作台”按钮可开始使用 Developer tool。

关闭欢迎页后，可以单击**帮助 > 欢迎**访问该页。

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](#)。

Informatica Developer 任务

使用 Developer tool 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 tool 可设计并运行实现以下目标的过程：

- 对源数据运行配置文件。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，因为它可以标识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质量度量的图形表示形式。
- 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 tool 中创建这些规则。
- 与 Informatica 用户协作。您添加到模型存储库中的规则可供 Developer tool 和 Analyst 工具中的用户使用。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。

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 想要先剖析波士顿和洛杉矶数据，然后再集成数据集。剖析操作可识别 HypoStores 可在集成之前修复的数据质量问题。

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。

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。您必须验证分析服务、模型存储库服务和数据集成服务是否正在运行。

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
- 您可在服务端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
- Customer_Order.xsd
- LA_customers.csv
- orders.csv

Informatica Analyst 教程

在本教程中，分析人员登录到 Analyst 工具，创建项目和文件夹，创建配置文件和规则，创建数据相关结果卡，以及创建引用表。

下表介绍了可执行的课程：

课程	说明
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。
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。
课程 3.创建默认配置文件	创建默认配置文件以快速了解数据质量。
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、采样选项和向下钻取选项。
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。
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。
课程 7.使用配置文件列创建引用表	创建可用于将源数据标准化的引用表。
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。

Informatica Developer 教程

在此教程中，您将使用 Developer tool 执行几项数据质量操作。您会创建与模型存储库的连接、创建项目和文件夹、将平面文件导入为物理数据对象、创建企业发现配置文件、执行对两个数据源的联接分析，以及查看配置文件结果。

下表介绍了可执行的课程：

课程	说明
课程 1.设置 Informatica Developer	添加一个域，创建与域中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。选择默认数据集成服务，以运行映射和预览数据。
课程 2.导入物理数据对象	导入平面文件以创建物理数据对象。
课程 3.对源数据运行配置文件	创建一个企业发现配置文件，执行对两个数据源的联接分析，并查看配置文件结果。 您可以在配置文件中执行联接分析，以确定是否能在两个数据列间建立有效联接。

第 I 部分： Informatica Analyst 入门

本部分包含以下章节：

- [课程 1。设置 Informatica Analyst, 17](#)
- [课程 2。创建数据对象, 20](#)
- [课程 3. 创建默认配置文件, 23](#)
- [课程 4。创建自定义配置文件, 26](#)
- [课程 5。创建表达式规则, 29](#)
- [课程 6。创建并运行结果卡, 32](#)
- [课程 7。使用配置文件列创建引用表, 37](#)
- [课程 8。创建引用表, 40](#)

第 2 章

课程 1。设置 Informatica Analyst

本章包括以下主题：

- [设置 Informatica Analyst 概览, 17](#)
- [任务 1. 登录 Informatica Analyst, 18](#)
- [任务 2. 创建项目, 18](#)
- [任务 3. 创建文件夹, 18](#)
- [设置 Informatica Analyst 摘要, 19](#)

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, 20](#)
- [任务 1. 创建平面文件数据对象, 21](#)
- [任务 2. 查看数据对象属性, 21](#)
- [创建数据对象摘要, 22](#)

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：

- 已完成本教程中的课程 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

在本任务中，使用**添加平面文件**向导从 LA_Customers 数据文件创建平面文件数据对象。

1. 在**项目**导航器中，选择教程项目中的 Customers 文件夹。
注意：必须选择要创建平面文件数据对象所在的项目或文件夹，然后才能进行创建。
2. 单击**新建 > 平面文件数据对象**。
平面文件数据向导会显示在**发现工作区**中。
3. 选择**浏览并上载**，然后单击**浏览**。
4. 浏览到 LA_Customers.csv 所在的位置，然后单击**打开**。
5. 单击**下一步**。
选择导入类型页面会显示**带分隔符**和**固定宽度**选项。默认选项为**带分隔符**。
6. 单击**下一步**。
7. 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。
8. 在**指定要导入的行**下，选择**从第一行导入**，以从不为空的第一行导入列名。
9. 单击**下一步**。
列属性部分显示每列的名称、数据类型、精度、小数位数和格式。
10. 单击**下一步**。
名称字段显示 **LA_Customers**。
11. （可选）更改文件名并添加说明。
12. 在**文件夹**部分中，选择 **Customers** 文件夹。
13. 单击**完成**。
LA_Customers 数据对象会显示在**发现工作区**中，并显示源数据的数据预览。

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, 23](#)
- [任务 1. 创建并运行默认配置文件, 24](#)
- [任务 2. 在摘要视图中查看配置文件结果, 24](#)
- [任务 3. 在详细视图中查看配置文件结果, 25](#)
- [创建默认配置文件摘要, 25](#)

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：

- 已完成本教程中的课程 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. 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。
4. 单击**下一步**。
5. 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。
6. 单击**下一步**。
7. 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**选项。
8. 单击**下一步**。
9. 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。

任务 3. 在详细视图中查看配置文件结果

在本任务中，使用 LA_Customers 配置文件的详细视图快速了解配置文件结果中的列。

1. 单击摘要视图中的列名称，在详细视图中查看此列的配置文件结果。
此列的配置文件结果将显示在详细视图中。
2. 在**常规**窗格中，可以使用水平条形图和百分比的形式查看空值、相异值和非相异值。
3. 在**值**窗格中，可以在上面板中查看所有值。在上面板中的值上方滑动滑块时，滑块中的所有值将显示在下面板中。
4. 在**数据预览**窗格中，单击**操作 > 显示预览**可查看源数据。
5. 在**模式**窗格中，可以使用水平条形图和百分比的形式查看列中的推理模式。
6. 在**数据类型**窗格中，可以使用水平条形图和百分比的形式查看列中的推理数据类型。
7. 在**数据域**窗格中，可以查看推理数据域。

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结果后，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, 26](#)
- [任务 1. 创建自定义配置文件, 27](#)
- [任务 2. 运行配置文件, 27](#)
- [任务 3. 向下钻取配置文件结果, 28](#)
- [创建自定义配置文件课程总结, 28](#)

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 Hadoop 或本地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。

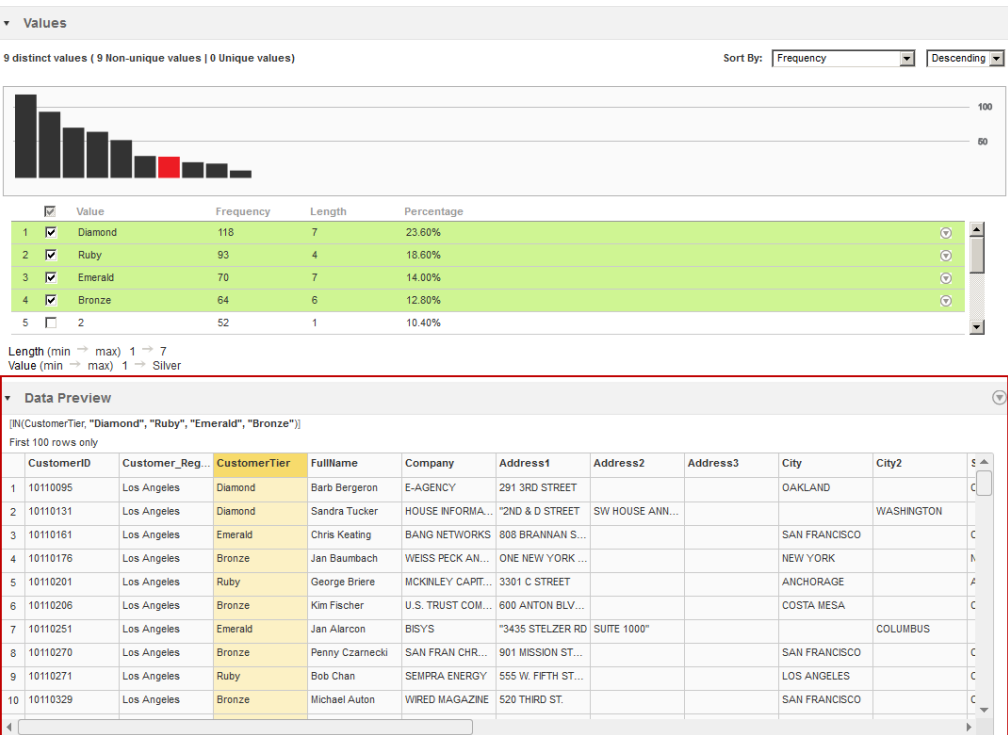
1. 验证您是否处于**发现工作区**中。
可以在工作区中查看配置文件 **Profile_LA_Customers_Custom**。
2. 在工作区中单击 **Profile_LA_Customers_Custom**。
3. 此时将显示配置文件屏幕，您可以在此屏幕中选择编辑配置文件或运行配置文件。单击**运行**。
4. 配置文件结果将显示在摘要视图中。

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, 29](#)
- [任务 1. 创建表达式规则并运行配置文件, 30](#)
- [任务 2. 查看表达式规则的输出, 30](#)
- [任务 3. 编辑表达式规则, 31](#)
- [创建表达式规则课程总结, 31](#)

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输出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, 32](#)
- [任务 1. 使用配置文件结果创建结果卡, 33](#)
- [任务 2. 运行结果卡, 34](#)
- [任务 3. 查看结果卡, 34](#)
- [任务 4. 编辑结果卡, 35](#)
- [任务 5. 配置阈值, 35](#)
- [任务 6. 查看得分趋势图表, 35](#)
- [创建并运行结果卡课程总结, 36](#)

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。

- 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。

- 确认已进入**结果卡**工作区，并已打开 sc_LA_Customer 结果卡。
- 选择**操作 > 编辑 > 度量**。
此时将显示**编辑结果卡**对话框。
- 在**度量**部分中，选择 **CustomerTier**。
- 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。
- 单击**保存并运行**，以保存结果卡更改并运行结果卡。
- 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。

- 确认已进入**结果卡**工作区，并已打开 sc_LA_Customer 结果卡。
- 选择**操作 > 编辑 > 度量**。
此时将显示**编辑结果卡**对话框。
- 在**度量**部分中，选择**状态**。
- 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。
- 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。

- 确认已进入**结果卡**工作区，并已打开 sc_LA_Customer 结果卡。
- 选择**状态**行。
- 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, 37](#)
- [任务 1. 使用配置文件列创建引用表, 38](#)
- [任务 2. 编辑引用表, 38](#)
- [使用配置文件列创建引用表课程总结, 39](#)

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, 40](#)
- [任务 1。创建引用表, 41](#)
- [创建引用表课程总结, 41](#)

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：

- 已完成本教程中的课程 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, 43](#)
- [课程 2： 导入物理数据对象, 47](#)
- [课程 3. 对源数据运行配置文件, 56](#)

第 10 章

课程 1。设置 Informatica Developer

本章包括以下主题：

- [设置 Informatica Developer 概览, 43](#)
- [任务 1。启动 Informatica Developer, 44](#)
- [任务 2。添加域, 44](#)
- [任务 3。添加模型存储库, 45](#)
- [任务 4。创建项目, 45](#)
- [任务 5。创建文件夹, 45](#)
- [任务 6。选择默认数据集成服务, 46](#)
- [设置 Informatica Developer 课程总结, 46](#)

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。如果域包含多个服务，则您还必须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。如果域中有多个数据集成服务，则需要执行此步骤。如果域仅包含一个数据集成服务，该服务将设置为默认服务。

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, 47](#)
- [任务 1. 导入 Boston_Customers 平面文件数据对象, 48](#)
- [任务 2。导入 LA_Customers 平面文件数据对象, 54](#)
- [任务 3. 导入 All_Customers 平面文件数据对象, 54](#)
- [导入物理数据对象课程总结, 55](#)

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：

- 已完成本教程中的课程 1。

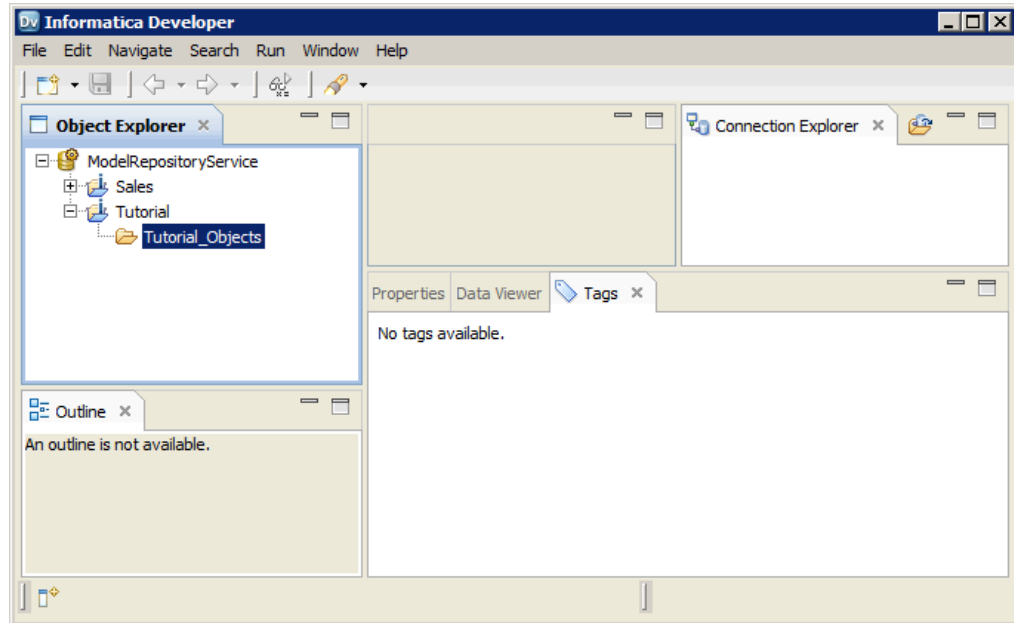
计时

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

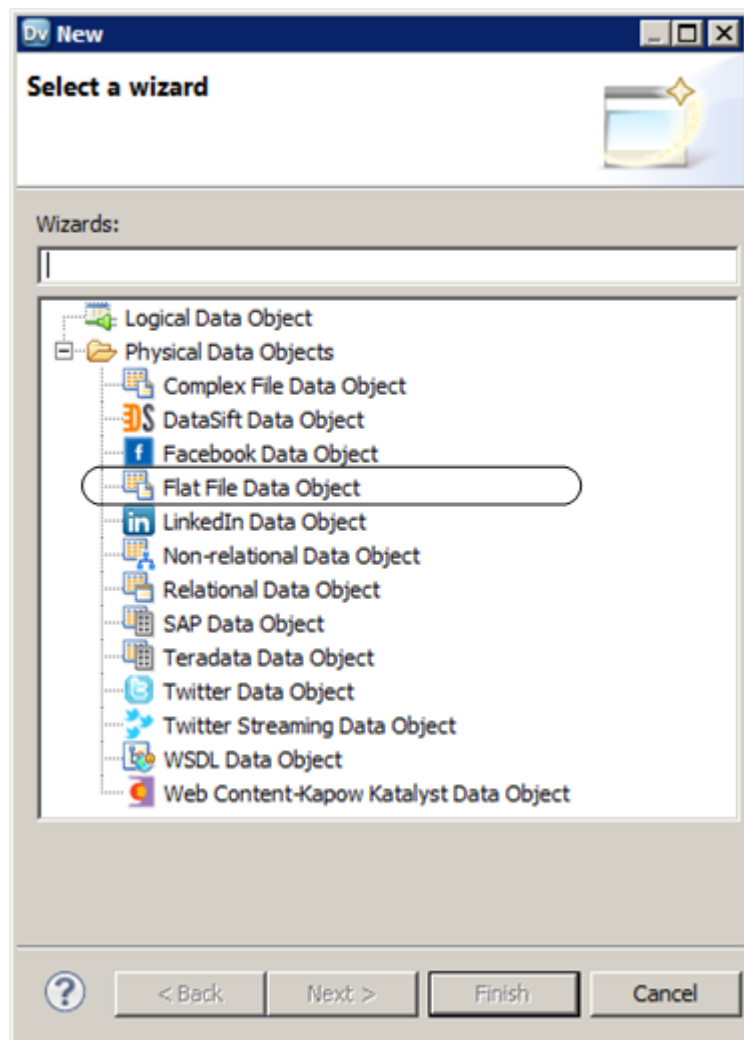
在本任务中，从包含波士顿办公室的客户数据的文件导入物理数据对象。

1. 在对象浏览器视图中，选择 Tutorial_Objects 文件夹。



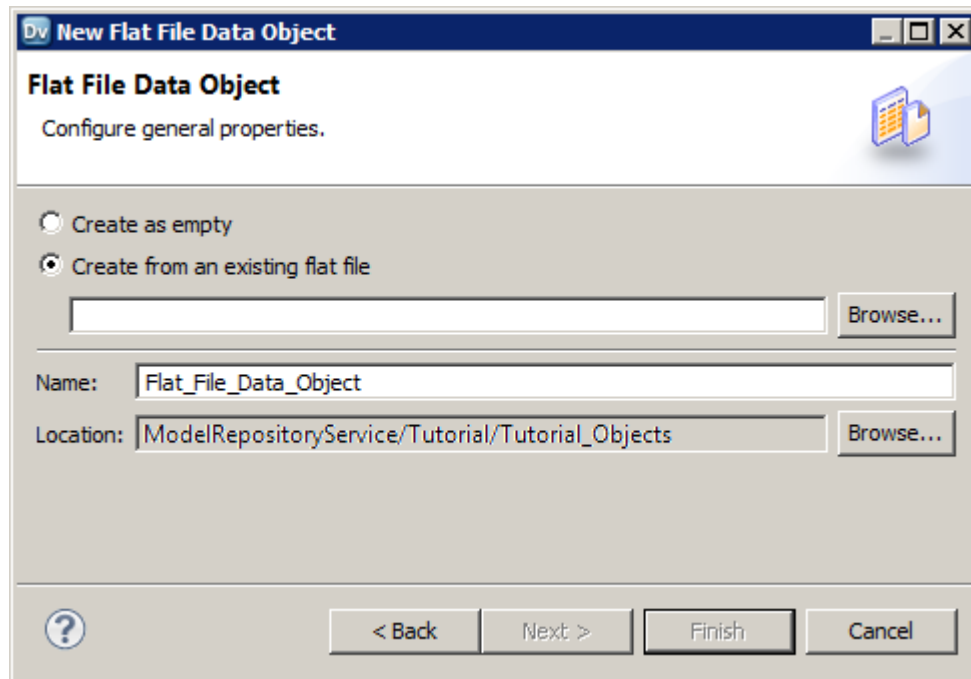
2. 右键单击 Tutorial_Objects 文件夹并选择新建 > 数据对象。

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。

New Flat File Data Object

Flat File Data Object

Configure code page and format.

Code page: MS Windows Latin 1 (ANSI), superset of Latin 1

Format

☒ Delimited (fields separated by delimiters)

☐ Fixed-width (fields aligned in columns)

Maximum rows to preview: 500

	Field1	Field2	Field3	
1	CustomerID	Customer Region	CustomerTier	La
2	10110102	Boston	Bronze	Br
3	10110105	Boston		An
4	10110106	Boston	Emerald	Bc
5	10110107	Boston	Ruby	Ch
6	10110109	Boston	Gold	Ac
7	10110111	Boston	2	Ke
8	10110117	Boston	2	Cc

?

< Back

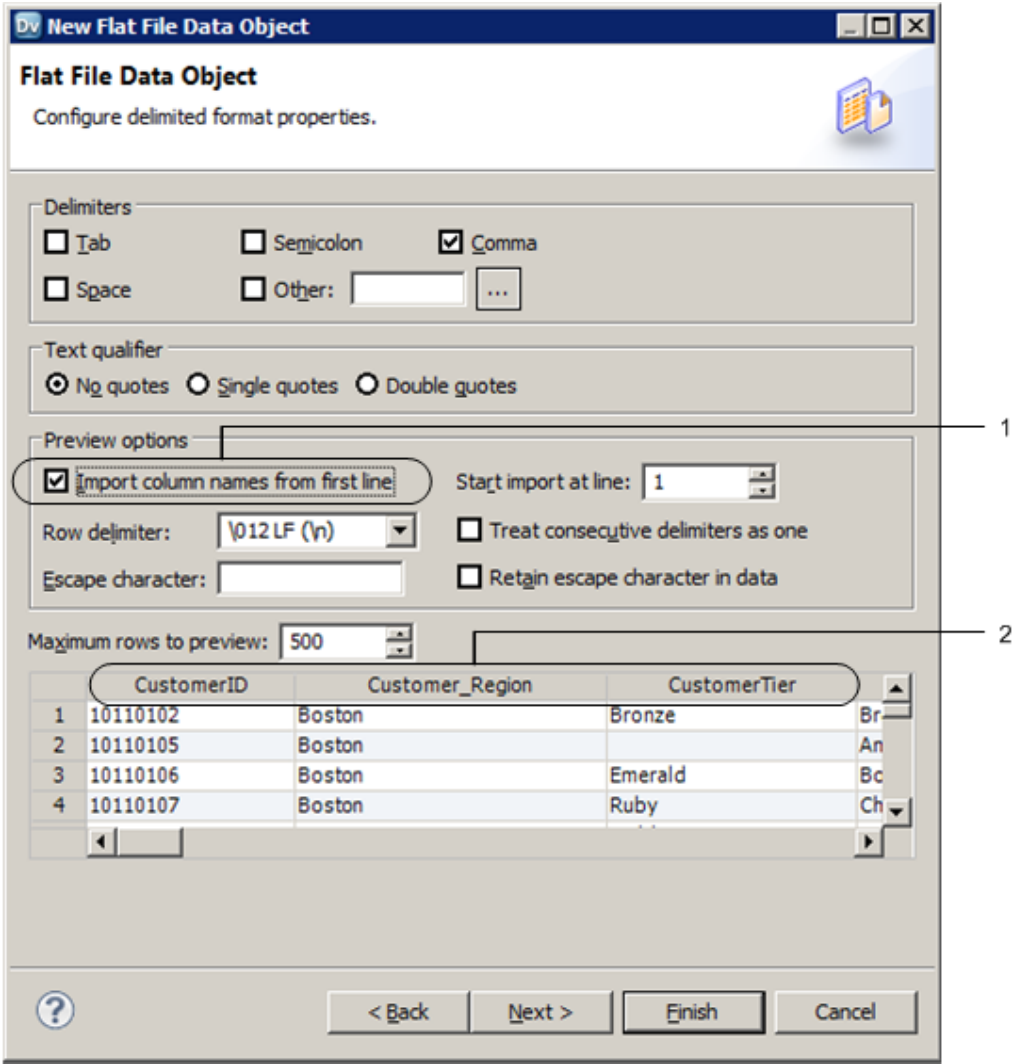
Next >

Finish

Cancel

9. 单击下一步。
10. 选择导入第一行中的列名称。

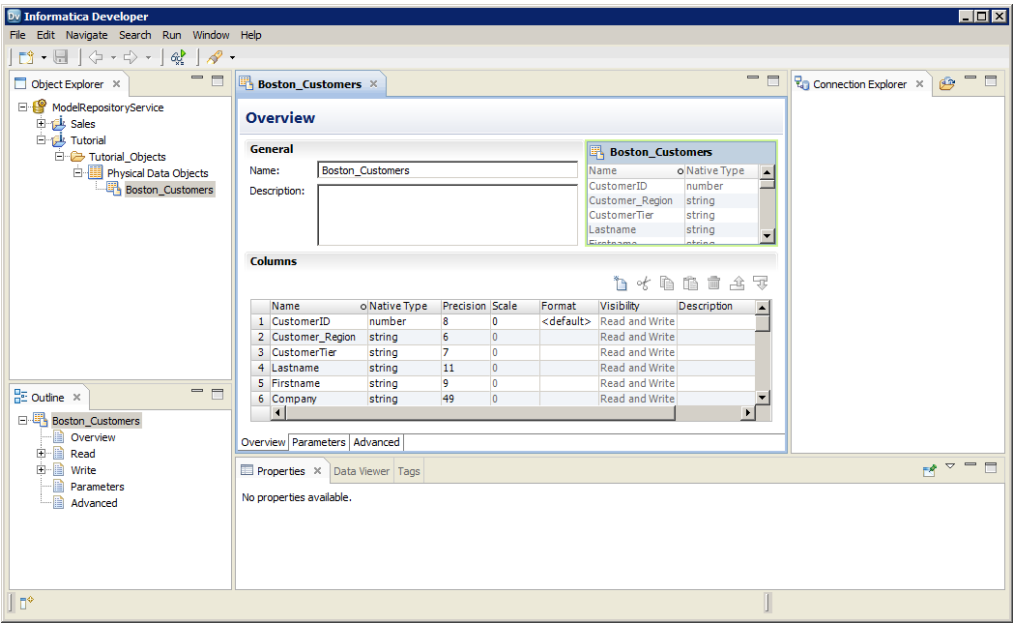
新建平面文件数据对象对话框将在平面文件数据的预览中显示列名称。



- 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。



12. 单击高级视图。
- 高级视图会显示物理数据对象的属性。
13. 在高级视图中，滚动至运行时: 读取部分。
14. 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。
15. 单击数据查看器视图。
16. 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。
17. 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, 56](#)
- [任务 1。对两个数据源执行联接分析, 57](#)
- [任务 2。查看联接分析结果, 58](#)
- [任务 3。对数据源运行配置文件, 58](#)
- [任务 4。查看列剖析结果, 58](#)
- [剖析数据课程总结, 59](#)

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，这可让您验证数据质量。

索引

C

- 创建表达式规则
 - 概览 [29](#)
- 创建结果卡
 - 概览 [32](#)
- 创建默认配置文件
 - 概览 [23](#)
- 创建数据对象
 - 概览 [20](#)
- 从列创建引用表
 - 概览 [37](#)
- 创建自定义配置文件
 - 概览 [26](#)

D

- 导入物理数据对象
 - 概览 [47](#)

P

- 剖析
 - 概览 [8](#)
- 剖析数据
 - 概览 [56](#)

S

- 设置 Analyst 工具
 - 概览 [17](#)
- 设置 Developer tool
 - 概览 [43](#)

Y

- 引用表
 - 概览 [40](#)