



Informatica®  
10.4.0

# 配置文件指南

Informatica 配置文件指南

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](mailto: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: 2020-02-05

# 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 I 部分： 配置文件简介 .....	10
第 1 章： 配置文件简介 .....	11
配置文件简介概览 .....	11
剖析进程 .....	12
剖析工具 .....	12
配置文件组件 .....	12
第 2 章： 列配置文件概念 .....	14
列配置文件概念概览 .....	14
列配置文件选项 .....	15
存储库配置文件锁定和受版本控制的配置文件管理 .....	15
结果卡 .....	15
第 3 章： 内容管理概念 .....	16
内容管理概念概览 .....	16
内容管理任务 .....	16
第 II 部分： 使用 Informatica Analyst 进行剖析 .....	18
第 4 章： Informatica Analyst 中的列配置文件 .....	19
Informatica Analyst 中的列配置文件概览 .....	19
列剖析过程 .....	20
配置文件选项 .....	20
采样选项 .....	20
向下钻取选项 .....	21
运行时环境 .....	21
本地环境 .....	21
Hadoop 环境 .....	21
Informatica Analyst 中的操作系统配置文件概览 .....	22

选择操作系统配置文件. . . . .	22
存储库资产锁定和基于团队的开发概览. . . . .	23
在 Informatica Analyst 中创建列配置文件. . . . .	23
编辑列配置文件. . . . .	24
运行配置文件. . . . .	25
同步选项. . . . .	25
在 Informatica Analyst 中同步平面文件数据对象. . . . .	25
在 Informatica Analyst 中同步关系数据对象. . . . .	27
 第 5 章： Informatica Analyst 中的规则. . . . .	28
Informatica Analyst 中的规则概览. . . . .	28
预定义规则. . . . .	28
预定义规则进程. . . . .	29
应用预定义规则. . . . .	29
表达式规则. . . . .	29
创建表达式规则. . . . .	30
使用规则规范创建表达式规则. . . . .	31
 第 6 章： Informatica Analyst 中的筛选器. . . . .	33
Informatica Analyst 中的筛选器概览. . . . .	33
创建筛选器. . . . .	33
创建简单筛选器. . . . .	34
创建高级筛选器. . . . .	35
创建 SQL 筛选器. . . . .	36
管理筛选器. . . . .	36
 第 7 章： Informatica Analyst 中的列配置文件结果. . . . .	38
Informatica Analyst 中的列配置文件结果概览. . . . .	38
摘要视图. . . . .	39
摘要视图属性. . . . .	39
摘要视图中的默认筛选器. . . . .	40
详细视图. . . . .	41
详细视图窗格. . . . .	41
统计信息. . . . .	42
数据预览. . . . .	42
数据类型. . . . .	43
离群值. . . . .	43
模式. . . . .	44
值. . . . .	45
配置文件运行的类型. . . . .	48
最新配置文件运行. . . . .	48
历史配置文件运行. . . . .	48
已合并配置文件运行. . . . .	48

选择配置文件运行. . . . .	48
比较多个配置文件结果概览. . . . .	49
比较多个配置文件结果. . . . .	49
配置文件结果比较的摘要视图 . . . . .	50
配置文件结果比较的详细视图 . . . . .	52
列配置文件向下钻取. . . . .	53
向下钻取行数据. . . . .	54
应用筛选器以向下钻取数据. . . . .	54
Analyst 工具中的内容管理. . . . .	54
批准数据类型和数据域. . . . .	54
拒绝数据类型和数据域. . . . .	55
Informatica Analyst 中的列配置文件导出文件. . . . .	55
CSV 文件格式的配置文件导出结果. . . . .	55
Microsoft Excel 格式的配置文件导出结果. . . . .	55
从 Informatica Analyst 中导出配置文件结果. . . . .	56
 第 8 章： Informatica Analyst 中的结果卡. . . . .	 57
Informatica Analyst 结果卡概览. . . . .	57
Informatica Analyst 结果卡进程. . . . .	58
在 Informatica Analyst 中创建结果卡. . . . .	58
向现有结果卡添加列. . . . .	59
运行结果卡. . . . .	60
查看结果卡. . . . .	61
编辑结果卡. . . . .	61
度量. . . . .	62
度量权重. . . . .	62
数据质量的值. . . . .	62
定义阈值. . . . .	62
度量组. . . . .	63
创建度量组. . . . .	63
将得分移至度量组. . . . .	63
编辑度量组. . . . .	64
删除度量组. . . . .	64
对列进行向下钻取. . . . .	64
趋势图表. . . . .	65
得分趋势图表. . . . .	65
成本趋势图表. . . . .	65
查看趋势图表. . . . .	66
导出趋势图表. . . . .	67
Informatica Analyst 结果卡导出文件. . . . .	68
Microsoft Excel 格式的结果卡导出结果. . . . .	68
从 Informatica Analyst 导出结果卡结果. . . . .	68
结果卡通知. . . . .	68

通知电子邮件模板. . . . .	69
设置结果卡通知. . . . .	70
配置结果卡通知的全局设置. . . . .	70
结果卡沿袭. . . . .	71
在 Informatica Analyst 中查看结果卡沿袭. . . . .	71
<b>第 III 部分：使用 Informatica Developer 进行剖析. . . . .</b>	<b>72</b>
<b>第 9 章：数据对象配置文件. . . . .</b>	<b>73</b>
Informatica Developer 中的列配置文件. . . . .	73
筛选选项. . . . .	73
采样选项. . . . .	74
在 Informatica Developer 中创建单个数据对象配置文件. . . . .	74
在 Informatica Developer 中创建多个数据对象配置文件. . . . .	75
在 Informatica Developer 中同步平面文件数据对象. . . . .	75
在 Informatica Developer 中同步关系数据对象. . . . .	77
<b>第 10 章：基于半结构化数据源的列配置文件. . . . .</b>	<b>78</b>
基于半结构化数据源的列配置文件概览. . . . .	78
JSON 和 XML 数据对象. . . . .	78
从 JSON 或 XML 数据源创建数据对象. . . . .	79
HDFS 中半结构化数据源的复杂文件数据对象. . . . .	79
HDFS 中来自 JSON 或 XML 数据源的复杂文件数据对象. . . . .	79
HDFS 中来自 Avro 或 Parquet 数据源的复杂文件数据对象. . . . .	80
创建 HDFS 连接. . . . .	80
从 HDFS 中的 JSON 或 XML 文件创建复杂文件数据对象. . . . .	80
从 Avro 或 Parquet 数据源创建复杂文件数据对象. . . . .	81
创建基于半结构化数据源的列配置文件. . . . .	82
<b>第 11 章：Informatica Developer 中的规则. . . . .</b>	<b>84</b>
Informatica Developer 中的规则概览. . . . .	84
在 Informatica Developer 中创建规则. . . . .	85
在 Informatica Developer 中应用规则. . . . .	85
<b>第 12 章：Mapplet 和映射剖析. . . . .</b>	<b>86</b>
Mapplet 和映射剖析概览. . . . .	86
对 Mapplet 或映射对象运行配置文件. . . . .	86
比较映射或 Mapplet 对象的配置文件. . . . .	87
从配置文件生成映射. . . . .	87
<b>第 13 章：Informatica Developer 中的列配置文件结果. . . . .</b>	<b>88</b>
Informatica Developer 中的列配置文件结果. . . . .	88
列值属性. . . . .	89

列模式属性. . . . .	89
列统计信息属性. . . . .	89
列数据类型属性. . . . .	90
Informatica Developer 中的内容管理. . . . .	90
批准数据类型. . . . .	91
拒绝数据类型. . . . .	91
从 Informatica Developer 中导出配置文件结果. . . . .	91
 第 14 章： Informatica Developer 中的结果卡. . . . .	 93
Informatica Developer 中的结果卡概览. . . . .	93
创建结果卡. . . . .	93
为结果卡沿袭导出资源文件. . . . .	94
查看 Informatica Developer 中的结果卡沿袭. . . . .	94
 索引. . . . .	 95

# 前言

要学习如何使用配置文件分析数据源的内容和结构，请阅读《*Informatica 配置文件指南*》。您可以确定列中源数据的特征，例如值频率、百分比和模式。本指南面向数据分析师和开发人员。

## 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](mailto: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](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](mailto: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 基于数以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](mailto: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以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 I 部分： 配置文件简介

本部分包含以下章节：

- [配置文件简介, 11](#)
- [列配置文件概念, 14](#)
- [内容管理概念, 16](#)

# 第 1 章

## 配置文件简介

本章包括以下主题：

- [配置文件简介概览, 11](#)
- [剖析进程, 12](#)
- [剖析工具, 12](#)
- [配置文件组件, 12](#)

### 配置文件简介概览

可以创建并运行配置文件，以查找应用程序、架构或企业数据源的内容、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。

可以在 Developer 工具和 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。
- 向列配置文件中添加规则。
- 对配置文件结果中的推理数据类型进行内容管理。
- 使用结果卡监视数据质量。
- 从配置文件生成映射。

## 剖析进程

开始数据集成项目时，剖析通常是第一步。可以创建配置文件以分析数据源的内容、质量和结构。作为剖析过程的一部分，可以发现数据源的元数据。

可以将不同的配置文件用于不同类型的数据分析，如列配置文件。发现并记录数据质量问题。要执行剖析，请完成以下任务：

1. 查找并分析数据源中的数据内容。包括数据类型、值频率、模式频率和数据统计信息，如最小值和最大值。
2. 审阅剖析结果。
3. 创建引用数据。
4. 向下钻取配置文件结果。
5. 记录数据问题。
6. 创建并运行规则。
7. 创建结果卡以监视数据质量。

## 剖析工具

可以使用多种 Informatica 工具来管理剖析过程。

可以使用以下工具来管理剖析过程：

Informatica Administrator

管理用户、组、权限和角色。可以管理分析服务并在 Informatica Analyst 中管理项目和对象的权限。可以在 Informatica Developer 中使用此工具控制访问权限。

Informatica Developer

在此工具中创建并运行配置文件，以查找并分析一个或多个数据源的元数据。使用向导创建配置文件。

Informatica Analyst

可以在 Analyst 工具中运行列配置文件。运行配置文件后，可以对数据源中的数据行进行向下钻取。

## 配置文件组件

配置文件具有多个组件，您可以使用这些组件有效分析数据源的内容和结构。

配置文件具有以下组件：

筛选器

用于创建满足特定条件的原始数据源的子集。随后，可以对该示例数据运行配置文件。

规则

定义您在运行配置文件时应用于数据的条件的业务逻辑。可以向配置文件中添加规则以验证数据。

标记

根据业务目的在模型存储库中定义对象的元数据。创建标记可以根据对象的业务用途来分组对象。

### 注释

有关配置文件的说明。使用注释可将配置文件的相关信息与其他 Analyst 和 Developer 工具用户共享。

### 结果卡

在配置文件结果中以图形方式显示列的有效值或规则输出。使用结果卡测量数据质量进度。

## 第 2 章

# 列配置文件概念

本章包括以下主题：

- [列配置文件概念概览, 14](#)
- [列配置文件选项, 15](#)
- [存储库配置文件锁定和受版本控制的配置文件管理, 15](#)
- [结果卡, 15](#)

## 列配置文件概念概览

列配置文件可确定数据源中列的特性，如值频率、百分比和模式。

列剖析发现数据的以下相关情况：

- 每个列中的空值、相异值和非相异值的个数，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。
- 记录的数据类型、推理的数据类型以及两者之间的潜在冲突。
- 模式离群值和值频率离群值。

创建或编辑配置文件时可以配置以下选项：

- 列配置文件选项。您可以选择要在上面运行配置文件的列，以及选择采样选项和向下钻取选项。
- 添加、编辑或删除筛选器和规则。

在配置文件结果中，可以向配置文件和配置文件中的列添加注释和标记。可以为列分配业务术语。

模型存储库会使用存储库配置文件锁定来锁定配置文件，以防止用户覆盖所做的工作。版本控制系统会保存配置文件的多个版本，并为每个版本分配一个版本号。可以先签出配置文件，然后在进行更改后再签入配置文件。重新签入配置文件之前，可以撤消配置文件的签出操作。

创建结果卡以定期查看数据质量。在为配置文件应用规则之前和之后均需创建结果卡，以便查看列中有效值的图形表示形式。

使用计划程序服务调度配置文件运行和结果卡运行，以便在特定时间或间隔运行。计划程序服务会管理配置文件、结果卡、已部署映射和已部署工作流的计划。您可以在 Informatica Administrator 中创建、管理和运行计划。

## 列配置文件选项

创建配置文件时，您可以使用配置文件向导定义筛选器、规则、向下钻取选项、采样选项和连接。这些选项将确定配置文件从源数据读取行的方式。

您可以在列配置文件、数据域发现配置文件或企业发现配置文件中定义以下选项：

- 筛选器。您可以为配置文件创建并应用筛选器。
- 规则。创建配置文件时可以添加规则。您可以重复使用在 Analyst 工具或 Developer tool 中创建的规则。
- 向下钻取选项。您可以选择读取数据源中的当前数据，或者读取暂存在剖析仓库中的配置文件数据。
- 采样选项。您可以选择一个采样选项，确定要对其运行配置文件的行数。
- 连接。您可以在本地或 Hadoop 运行时环境中运行配置文件。

## 存储库配置文件锁定和受版本控制的配置文件管理

模型存储库会锁定配置文件以防止用户覆盖所做的工作。当您开始编辑配置文件时，该配置文件会被锁定，使其他用户无法保存对其所做的更改。保存配置文件后，便会释放锁定。受版本控制的配置文件管理会创建配置文本版本，您可以查看版本历史记录。

在 Developer tool 或 Analyst 工具中编辑配置文本时，模型存储库会锁定该配置文件。如果该工具意外停止，锁定仍会保留，这样，当您再次连接到模型存储库时，便可查看已锁定的配置文件。您可以继续编辑配置文件，也可以解除配置文件锁定。

模型存储库与版本控制系统集成时，您可以管理配置文件的版本。例如，您可以签出和签入配置文件，撤消签出，查看配置文件的特定历史版本以及已签出的配置文件。有关 Analyst 工具中的存储库资产锁定和受版本控制的资产管理的信息，请参阅《Analyst 工具指南》。有关 Developer tool 中的存储库对象锁定和受版本控制的对象管理的信息，请参阅《Developer Tool 指南》。

## 结果卡

结果卡是配置文件结果中列的有效值或规则的输出的图形表示形式。使用结果卡测量数据质量进度。可以通过配置文件创建结果卡，然后监视数据质量随时间推移的进展情况。

结果卡包含多个组件，如度量、度量组和阈值。在运行配置文件后，可将源列作为度量添加到结果卡，然后为度量配置有效值。结果卡可以帮助组织通过在度量和结果卡级别跟踪错误数据的成本，从而度量数据质量的值。要度量每个度量的错误数据的成本，请为该度量分配一个成本单位，然后设置固定或可变成本。在运行结果卡后，结果卡结果将包括每个度量的错误数据成本，以及所有度量的总成本值。

使用度量组将结果卡中的相关度量分类到一个集合中。阈值以百分比的形式确定记录中的列可接受的错误数据的范围。可以为正常、可接受或不可接受的数据范围设置阈值。

在运行结果卡后，配置是否要对实时数据或暂存数据向下钻取得分度量。在运行结果卡并查看得分后，向下钻取每个度量，以确定有效数据记录和无效记录。还可以查看结果卡中每个度量或度量组的结果卡沿革。要有效跟踪数据质量，可以使用得分趋势图表和成本趋势图表。这些图表可以监视得分和错误数据的成本在一段时间内如何变化。

剖析仓库存储结果卡统计信息和配置信息。可以配置第三方应用程序来获取结果卡结果并运行报告。还可以在 Web 应用程序、门户或报告（如业务情报报告）中显示结果卡结果。

## 第 3 章

# 内容管理概念

本章包括以下主题：

- [内容管理概念概览, 16](#)
- [内容管理任务, 16](#)

## 内容管理概念概览

内容管理是验证和管理数据源的已发现元数据的过程，以便元数据适合使用和报告。

可以对以下推理配置文件结果进行内容管理：

- 数据类型
- 数据域
- 主键
- 外键

可以对推理配置文件结果进行内容管理，以使关于数据库和架构中的列、数据域和数据对象关系的元数据准确无误。随后可以在使用发现搜索跨多个存储库搜索信息时，找到最相关的元数据。还可以在企业发现结果中查看外键关系图表时，找到最相关的元数据。

可以对配置文件作为配置文件运行的一部分生成的特定元数据推理进行内容管理。例如，可以批准或拒绝列配置文件结果和数据域发现结果中的推理的数据类型。还可以批准或拒绝企业发现结果中的推理主键和外键。

## 内容管理任务

可以在配置文件运行后对配置文件结果进行内容管理。还可以扭转您在先前运行配置文件时采取的内容管理决策。

可以在 Analyst 工具中执行以下内容管理任务：

- 批准或拒绝多个列和数据域的推理的数据类型。
- 将已批准或已拒绝的数据类型还原为已推理状态。
- 将已批准或已拒绝的数据域还原为推理状态。
- 查看或隐藏已拒绝的结果行。



- 根据特定的元数据首选项（如已批准的数据类型和数据域）将列从配置文件运行中排除。

可以在 Developer 工具中执行以下内容管理任务：

- 批准或拒绝多个列的推理的数据类型。
- 将已批准或已拒绝的数据类型还原为已推理状态。
- 将已批准或已拒绝的数据域还原为推理状态。
- 查看或隐藏已拒绝的结果行。
- 批准或拒绝主键发现结果中的数据对象。
- 批准或拒绝企业发现结果，包括外键发现结果。
- 根据特定的元数据首选项（如已批准的数据类型和数据域）将列从配置文件运行中排除。

# 第 II 部分： 使用 Informatica Analyst 进行剖析

本部分包含以下章节：

- [Informatica Analyst 中的列配置文件, 19](#)
- [Informatica Analyst 中的规则, 28](#)
- [Informatica Analyst 中的筛选器, 33](#)
- [Informatica Analyst 中的列配置文件结果, 38](#)
- [Informatica Analyst 中的结果卡, 57](#)

## 第 4 章

# Informatica Analyst 中的列配置文件

本章包括以下主题：

- [Informatica Analyst 中的列配置文件概览, 19](#)
- [列剖析过程, 20](#)
- [配置文件选项, 20](#)
- [运行时环境, 21](#)
- [Informatica Analyst 中的操作系统配置文件概览, 22](#)
- [存储库资产锁定和基于团队的开发概览, 23](#)
- [在 Informatica Analyst 中创建列配置文件, 23](#)
- [编辑列配置文件, 24](#)
- [运行配置文件, 25](#)
- [同步选项, 25](#)

## Informatica Analyst 中的列配置文件概览

创建配置文件时，您应选择要运行的配置文件所基于的数据对象中的列。您可以配置采样选项和向下钻取选项，以提高剖析速度。您可以选择运行时环境。创建配置文件时，您可以向该配置文件添加规则和筛选器。在运行配置文件后，可以检查剖析统计信息，以了解数据。

可以剖析最多包含 1000 列的宽表和平面文件。创建或运行配置文件时，您可为配置文件选择所有列，也可选择各个列。您可以选择对所有列进行向下钻取，并可查看这些列的值频率。

可以在 Informatica Analyst 中使用以下方法创建列配置文件：

- 右键单击库工作区中的数据对象来创建配置文件。
- 使用默认选项创建默认列配置文件。
- 自定义配置文件设置来创建自定义配置文件。

**注意：**您可以查看并运行基于 Avro、JSON、Parquet 和 XML 数据源的配置文件。您可以在 Informatica Developer 中创建并编辑基于 Avro、JSON、Parquet 和 XML 数据源的配置文件。

# 列剖析过程

在列剖析过程中，可以选择包括所有源列进行剖析，或者选择特定列进行剖析。您也可以接受默认配置文件选项，或配置采样选项、向下钻取选项和运行时环境。

以下步骤描述了列剖析过程：

1. 为列配置文件选择名称、说明和位置。
2. 选择要运行配置文件的导入数据对象或外部源。
3. （可选）预览源数据。
4. 选择要运行配置文件的列。
5. 确定要使用默认选项创建配置文件还是更改默认选项。可以配置的选项包括采样选项、向下钻取选项以及运行时环境。
6. （可选）在创建配置文件时添加规则和筛选器。
7. 运行配置文件。

**注意：**对于列名称以及剖析多语言和 Unicode 数据，请考虑以下规则和准则：

- 可以从不同的源剖析多语言数据，并基于浏览器中的区域设置查看配置文件结果。Analyst 工具将根据浏览器的区域设置更改“日期时间”、“数值”和“十进制”数据类型。
- 对多语言数据进行排序。您可以对多语言数据进行排序。Analyst 工具将根据浏览器的区域设置显示排序顺序。
- 要剖析 DB2 数据库中的 Unicode 数据，请设置数据库中的 DB2CODEPAGE 数据库环境变量，并重新启动数据集成服务。

## 配置文件选项

配置文件选项包括数据采样选项和数据向下钻取选项。可以在为数据对象创建或编辑列配置文件时配置这些选项。

您可以在**发现**工作区中配置配置文件选项。可以选择使用默认的列选项、采样选项和向下钻取选项创建配置文件。使用向下钻取选项可在实时数据和暂存数据之间进行选择。

### 采样选项

采样选项可以确定 Analyst 工具选择对其运行配置文件的行数。可以在定义配置文件或运行配置文件时配置采样选项。

下表介绍了配置文件的各个采样选项：

选项	说明
所有行	选择数据对象中的所有行。
对前 <number> 行进行采样	希望对其运行配置文件的行数。Analyst 工具将从源的第一行开始选择行。最多可以选择 2,147,483,647 行。
对 <number> 行随机采样	随机采样算法将以随机方式在数据对象中选择行来运行配置文件。最多可以选择 2,147,483,647 行。

选项	说明
随机采样(自动)	随机采样大小根据数据对象中的行数计算而得。
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	在下次运行配置文件时，从数据类型和数据域推理中排除已批准的数据类型或数据域。

选择对随机的行样本运行配置文件后，随机采样算法将以随机方式在数据对象中选择行来运行配置文件。为列配置文件选择随机采样选项时，Analyst 工具将对暂存的数据执行向下钻取。这会影响向下钻取性能。为数据域发现配置文件选择随机采样选项时，Analyst 工具将对实时数据执行向下钻取。

## 向下钻取选项

可以在定义或编辑配置文件时配置向下钻取选项。

下表介绍了配置文件的各个向下钻取选项：

选项	说明
实时	对实时数据进行向下钻取，以读取数据源中的当前数据。
暂存	对暂存数据进行向下钻取，以读取暂存在剖析仓库中的配置文件数据。
选择列	标识要向下钻取但并未选择进行剖析的列。

## 运行时环境

您可以选择本地或 Hadoop 作为列配置文件的运行时环境。在 Hadoop 运行时环境中，可以选择 Blaze 或 Spark 引擎。选择运行时环境后，Informatica Analyst 会在配置文件定义中设置该运行时环境。

### 本地环境

在本地运行时环境中运行配置文件时，Analyst 工具会向剖析服务模块提交配置文件作业。剖析服务模块随后将配置文件作业拆分成一组映射。数据集成服务会在运行数据集成服务的同一台计算机上运行这些映射，并将配置文件结果写入剖析仓库。默认情况下，所有配置文件都在本地运行时环境中运行。

您可以使用本地源在本地环境中创建并运行配置文件。本地数据源是非 Hadoop 源，例如平面文件、关系源或大型机源。您还可以使用本地环境中的 Hive 或 HDFS 数据源，在映射规范或逻辑数据源中运行配置文件。

### Hadoop 环境

您可以选择 Hadoop 选项，以便在 Hadoop 运行时环境中运行配置文件。

选择 Hadoop 选项后，可以选择一个 Hadoop 连接。数据集成服务将配置文件逻辑推送到 Hadoop 群集上的 Blaze 引擎来运行配置文件。

在 Hadoop 环境中运行配置文件时，Developer tool 会向剖析服务模块提交配置文件作业。剖析服务模块随后将配置文件作业拆分成一组映射。数据集成服务会通过 Hadoop 连接将映射推送至 Hadoop 环境。Blaze 引擎会处理映射，并且数据集成服务会将配置文件结果写入剖析仓库。

## Sqoop 数据源的列配置文件

您可以在使用 Sqoop 的数据对象上运行列配置文件。您可以选择 Hadoop 运行时环境来运行列配置文件。

在逻辑数据对象或自定义数据对象上运行列配置文件时，您可以配置 num-mappers 参数以实现并行处理并优化性能。您还必须配置 split-by 参数，以指定 Sqoop 必须据其拆分工作单元的列。

请使用以下语法：

```
--split-by <column_name>
```

如果主键的值并没有在最小值和最大值范围内均匀分布，则可以将 split-by 参数配置为指定另一个具有平衡的数据分布的列来拆分工作单元。

如果没有定义 split-by 列，Sqoop 会根据以下条件拆分工作单元：

- 如果数据对象包含单个主键，Sqoop 会将主键用作 split-by 列。
- 如果数据对象包含复合主键，则 Sqoop 会默认处理不包含 split-by 参数的复合主键。有关详细信息，请参阅 Sqoop 文档。
- 如果数据对象包含具有相同列的两个表，您必须使用表限定名称定义 split-by 列。例如，如果表名称是 CUSTOMER，列名称是 FULL\_NAME，请将 split-by 列定义如下：  
--split-by CUSTOMER.FULL\_NAME
- 如果数据对象不包含主键，m 参数和 num-mappers 参数的值默认为 1。

如果您使用 Teradata 提供技术支持的 Cloudera 连接器或 Hortonworks Connector for Teradata，并且 Teradata 表不包含主键，则需要配置 split-by 参数。

# Informatica Analyst 中的操作系统配置文件概览

您可以在 Analyst 工具中选择操作系统配置文件。选择一个操作系统配置文件后，数据集成服务即会根据操作系统配置文件用户的权限创建并运行列配置文件、企业发现配置文件和结果卡。

该 Analyst 工具会使用默认配置文件来运行配置文件和结果卡。如果只有一个操作系统配置文件，则默认情况下即会选择该操作系统配置文件。如果具有多个操作系统配置文件，您可以选择操作系统配置文件之一。

## 选择操作系统配置文件

您可以在 Informatica Analyst 中选择操作系统配置文件。数据集成服务使用操作系统配置文件用户的权限来运行剖析作业。

1. 在 Informatica Analyst 标头区域中，单击 **<用户名> > 设置**。  
此时将显示**设置**对话框。
2. 选择操作系统配置文件。单击**保存**。

# 存储库资产锁定和基于团队的开发概览

模型存储库会锁定配置文件，以防止用户覆盖其他用户所做的工作。如果模型存储库与版本控制系统集成在一起，它会保存多个资产版本并为版本分配版本号。可以签出和签入配置文件，还可以撤消签出。可以查看所签出配置文件的特定版本。

当您开始在 Analyst 工具中编辑配置文件时，模型存储库会锁定此配置文件，使其他用户无法对其进行编辑。保存此配置文件时，锁定仍会保留。关闭此配置文件时，模型存储库会解除配置文件锁定。

模型存储库会通过受版本控制的资产管理来保护配置文件不被开发团队的其他成员覆盖。当您试图编辑另一用户已经签出的配置文件时，您会收到一条通知，告知您哪位用户已经签出此配置文件。可以在只读模式下打开已签出的配置文件，也可以使用其他名称来保存此配置文件。

可以在“配置文件属性”对话框中选择配置文件的一个版本来查看此版本的配置文件定义。可以在“操作”菜单中访问“配置文件属性”选项。有关存储库资产锁定和受版本控制的资产管理的详细信息，请参阅《Analyst 工具指南》。

## 在 Informatica Analyst 中创建列配置文件

既可创建自定义配置文件，也可创建默认配置文件。创建自定义配置文件时，您可以配置列、采样行和向下钻取选项。创建默认配置文件时，将对整个数据集的所有数据域运行列配置文件和数据域发现。

1. 在**发现**工作区中，单击**配置文件**，或从表头区域中选择**新建 > 配置文件**。

**注意：**可以右键单击**库**工作区中的数据对象并创建一个配置文件。在此配置文件中，配置文件名称、位置名称和数据对象将从数据对象属性中提取。可以创建默认配置文件，也可以通过自定义相关设置来创建自定义配置文件。

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**屏幕中，输入配置文件的名称和可选说明。在“位置”字段中，选择要在其中创建配置文件的项目或文件夹。单击**下一步**。
4. 在**选择源**屏幕中，单击**选择**以选择数据对象，或单击**新建**以导入数据对象。单击**下一步**。
  - 在**选择数据对象**对话框中，选择数据对象。单击**确定**。  
“属性”窗格将显示所选数据对象的属性。“数据预览”窗格将显示数据对象中的列。
  - 在**新建数据对象**对话框中，可以选择要为其创建配置文件的链接、架构、表或视图，选择位置，然后创建用于导入数据对象的文件夹。单击**确定**。
5. 在**选择源**屏幕中，选择要对其运行配置文件的列。或者，选择**名称**以选择所有列。单击**下一步**。  
默认情况下将选择所有列。Analyst 工具将显示每个列的列属性，例如，名称、数据类型、精度、小数位数、可空性和主键的参与方。
6. 在**指定设置**屏幕中，选择要运行列配置文件还是数据域发现或同时运行列配置文件和数据域发现。默认情况下将选择列配置文件选项。
  - 选择**运行列配置文件**可运行列配置文件。
  - 选择**运行数据域发现**可执行数据域发现。在**数据域**窗格的**编辑数据域发现的列选择**对话框中，选择要发现的数据域，接着选择遵从性条件，然后选择要执行数据域发现的列。
  - 选择**运行列配置文件和运行数据域发现**可同时运行列配置文件和数据域发现。在**数据域**窗格中选择数据域选项。

**注意:** 默认情况下, 所选列将用于列配置文件和数据域发现。单击**编辑**可选择或取消选择用于数据域发现的列。

- 选择“数据”、“列”或“数据和列”以对其运行数据域发现。
- 选择采样选项。可以在**运行配置文件**窗格中选择**所有行(完整分析)**、**先采样**、**随机采样**或**随机采样(自动)**作为采样选项。此选项将应用于列配置文件和数据域发现。
- 选择向下钻取选项。可以在**向下钻取**窗格中选择**实时**或**暂存**向下钻取选项, 也可以选择**关闭禁用**向下钻取。或者, 单击**选择列**, 以选择要进行向下钻取的列。可以选择跳过对数据类型或数据域已经过批准的列进行数据类型和数据域推理。
- 选择**本地**或**Hadoop**选项作为运行时环境。如果选择 Hadoop 选项, 请单击**选择**以在**选择 Hadoop 连接**对话框中选择 Hadoop 连接。

7. 单击**下一步**。

此时将打开**指定规则和筛选器**选项卡。

8. 在**指定规则和筛选器**屏幕中, 可以执行以下任务:

- 创建、编辑或删除规则。您可以将现有规则应用到配置文件。
- 创建、编辑或删除筛选器。

**注意:** 基于此配置文件创建结果卡时, 可以重用为配置文件创建的筛选器。

9. 单击**保存并完成**以创建配置文件, 或单击**保存并运行**以创建并运行配置文件。

## 编辑列配置文件

可以在运行列配置文件后对其进行更改。

1. 在**库**工作区中, 选择包含配置文件的项目, 或者在**资产**窗格中选择配置文件。

2. 单击配置文件名称。

摘要视图将显示在**发现**工作区中。

3. 如果启用了版本控制系统, 请单击**操作 > 签出**以签出配置文件。

4. 单击**操作 > 编辑配置文件**。

此时将显示**配置文件**向导。

5. 根据要进行的更改, 选择以下页面选项之一:

- **指定常规属性**。更改基本属性, 例如名称、说明和位置。
- **选择源**。选择其他匹配的数据源和列以对其运行配置文件。
- **指定设置**。选择要运行列配置文件还是同时运行列配置文件和数据域发现。选择要发现的数据域, 并修改数据域发现、采样和向下钻取选项。
- **指定规则和筛选器**。创建、编辑或删除规则和筛选器。

6. 单击**保存并完成**以完成配置文件编辑, 或单击**保存并运行**以编辑并运行配置文件。

7. 如果启用了版本控制系统, 必须执行以下任务:

- 单击**保存并完成**以完成配置文件编辑。
- 在摘要视图中, 单击**签入**以签入配置文件。
- 单击**操作 > 运行配置文件**, 以运行该配置文件。



# 运行配置文件

运行配置文件，以针对内容和结构分析数据源，然后选择要用于向下钻取的列和规则。可以针对列和规则对实时数据或暂存数据进行向下钻取。执行初始配置文件运行后，您可以仅对某个列或规则运行配置文件，而不对所有源列运行配置文件。

1. 在**库**工作区中，在“项目”窗格中选择包含配置文件的项目或文件夹，或者在“资产”窗格中选择配置文件。
2. 单击**操作 > 打开**。  
摘要视图将显示在**发现**工作区中。
3. 单击**操作 > 运行配置文件**。  
Analyst 工具将执行配置文件运行，并在摘要视图中显示配置文件结果。
4. 在摘要视图中，单击一个列以查看列结果。  
此时将显示详细视图。

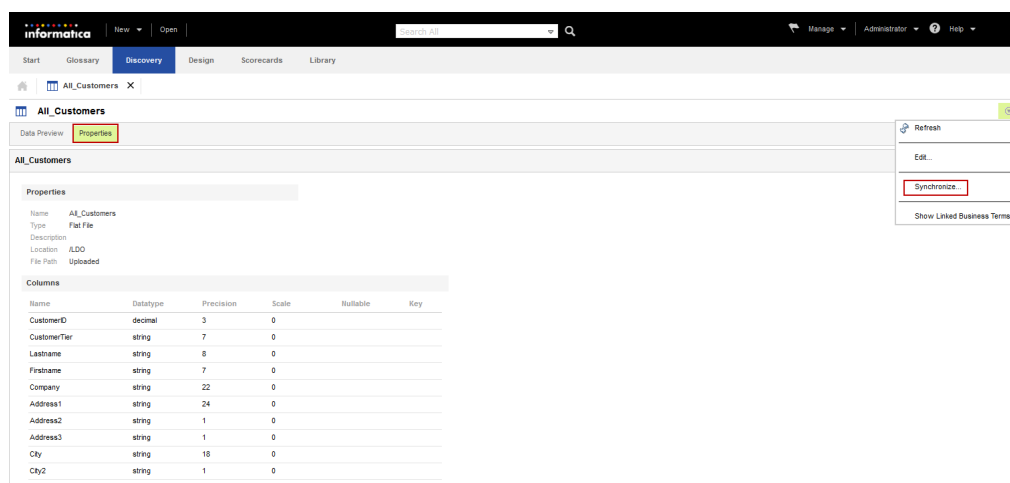
## 同步选项

更改外部数据源的元数据时，默认情况下不会更新模型存储库中的数据对象元数据。可以使用“同步”选项将数据对象元数据与数据源元数据同步。可以对列配置文件、企业发现配置文件和结果卡使用“同步”选项。外部数据源可以是关系数据源或平面文件数据源。

### 在 Informatica Analyst 中同步平面文件数据对象

可以将外部平面文件数据源的更改与其在 Analyst 工具中的数据对象同步。使用**同步平面文件**向导同步数据对象。

1. 打开**库**工作区。
2. 在**项目**部分中，选择项目中的平面文件数据对象。  
Analyst 工具会在**数据预览**选项卡中显示平面文件的数据预览。
3. 单击**属性**选项卡。
4. 从“操作”菜单中，单击**同步**。  
下图显示了“属性”选项卡和“操作”菜单中的“同步”选项：

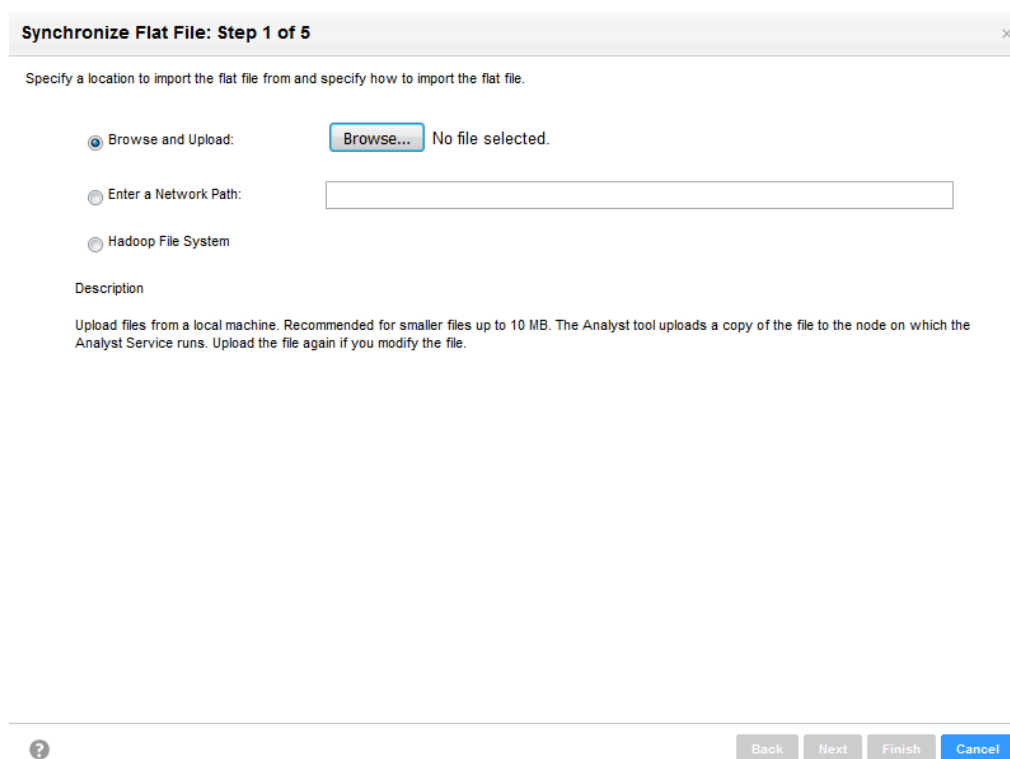


将显示同步平面文件向导。

5. 选择浏览某个位置或输入网络路径以导入平面文件。

- 要浏览位置，请单击**选择文件**以从计算机可访问的目录中选择平面文件。
- 要输入网络路径，请选择**输入网络路径**并配置文件路径和文件名。

下图显示了“同步平面文件”向导：



6. 单击**下一步**。

7. 选择导入带分隔符或固定宽度的平面文件。

- 要导入带分隔符的平面文件，请接受**带分隔符**选项。
- 要导入固定宽度的平面文件，请选择**固定宽度**选项。

8. 单击**下一步**。

9. 为带分隔符或固定宽度的平面文件配置平面文件选项。
10. 单击**下一步**。
11. (可选) 更改列属性。
12. 单击**下一步**。
13. 接受默认名称或为平面文件输入其他名称。
14. (可选) 输入说明。
15. 单击**完成**。  
此时将显示一条同步消息, 提示您确认操作。
16. 单击**是**可同步平面文件。  
将显示一条消息指明同步已完成。要查看元数据更改的详细信息, 请单击**显示详细信息**。
17. 单击**确定**。

## 在 Informatica Analyst 中同步关系数据对象

您可以将外部关系数据源的更改与其表数据对象进行同步。外部数据源更改包括添加、更改和删除源列和规则列。

1. 打开**库**工作区。
2. 在**项目**部分中, 选择项目中的表数据对象。  
Analyst 工具会在**数据预览**选项卡上显示表的数据预览。
3. 单击**属性**选项卡。
4. 从“操作”菜单中, 单击**同步**。  
此时将显示一条消息, 提示您确认操作。
5. 要完成同步过程, 请单击**是**。  
将显示同步状态消息。
6. 将显示一条消息指明同步已完成。  
要查看元数据更改的详细信息, 请单击**显示详细信息**。
7. 单击**确定**。

## 第 5 章

# Informatica Analyst 中的规则

本章包括以下主题：

- [Informatica Analyst 中的规则概览, 28](#)
- [预定义规则, 28](#)
- [表达式规则, 29](#)

## Informatica Analyst 中的规则概览

规则是业务逻辑，用于定义在运行列配置文件时对源数据应用的条件。可以在配置文件中添加规则以验证数据。

在不同情况下，您可能需要使用规则。可以添加规则，以清理一个或多个数据列。可以添加查找规则，以提供源数据无法提供的信息。可以添加规则，为数据质量或数据集成项目验证清理规则。

创建或编辑列配置文件时，可以创建规则并将其添加到配置文件中，或者也可以将现有规则应用到配置文件。可以在列配置文件中使表达式规则或预定义规则。

运行配置文件后，Analyst 工具会在摘要视图中显示规则列的配置文件结果。您可以在详细视图中查看规则的列结果。规则的输出可以是一个或多个虚拟列。虚拟列存在于配置文件结果中。Analyst 工具将对虚拟列运行配置文件。例如，您使用某一自定义规则来拆分某一列，该列在 FIRST\_NAME 和 LAST\_NAME 虚拟列中包含名字和姓氏。Analyst 工具将对 FIRST\_NAME 和 LAST\_NAME 列运行配置文件。

**注意：**如果您删除了其他对象类型引用的某一规则对象，则 Analyst 工具将显示一条消息，列出这些对象类型。在删除规则之前，请确定删除它会造成什么影响。

## 预定义规则

预定义规则是在 Developer tool 中创建的规则，或 Developer tool 和 Analyst 工具随附提供的规则。将预定义规则应用到列配置文件，以修改或验证源数据。

预定义规则使用转换来定义规则逻辑。可以将预定义规则与多个配置文件配合使用。在模型存储库中，一条预定义规则就是一个 Maplet，其中包含用于定义规则逻辑的输入组、输出组和转换。

## 预定义规则进程

可以使用**新建规则向导**将预定义规则应用于配置文件。

要应用预定义规则，可以执行以下步骤：

1. 打开某一配置文件。
2. 选择一个预定义规则。
3. 审阅该规则的参数。
4. 选择输入列。如果想将该规则应用于多个列，可以选择多个列。
5. 配置剖析选项。

## 应用预定义规则

在应用预定义规则时，首先选择该规则，然后为该规则配置输入列和输出列。应用一个预定义规则，以使用一条被升级为可重用规则的规则，或者使用由开发人员创建的规则。

1. 在**库**工作区中，选择包含配置文件的项目，或者在**资产**窗格中选择配置文件。
2. 单击**操作 > 打开**，以打开配置文件。  
摘要视图将显示在**发现**工作区中。
3. 单击**操作 > 编辑配置文件**。  
此时将显示**配置文件**向导。
4. 单击**指定规则和筛选器**。
5. 在**指定规则和筛选器**屏幕的**规则**面板中，单击**操作 > 应用现有规则**。  
此时将显示**应用规则向导**对话框。
6. 选择一个规则，然后单击**下一步**。
7. 单击**添加**。  
此时将显示**选择输入端口的列**对话框。
8. 选择一个字段和一个输入列。单击**确定**。  
输入列和输出列将显示在**应用规则向导**对话框中。
9. 在**应用规则向导**对话框中，单击**确定**。  
该规则将显示在**指定规则和筛选器**屏幕中。

## 表达式规则

表达式规则使用表达式函数和列来定义规则逻辑。可以在 Analyst 工具中创建表达式规则，然后将它们添加到列配置文件中。

可以使用表达式规则更改或验证列配置文件中的列的值。可以创建一个或多个表达式规则在配置文件中。表达式函数与用于转换源数据的 SQL 函数相似。可以使用以下类型的函数创建表达式规则逻辑：

- 字符
- 转换
- 数据清理

- 日期
- 编码
- 财务
- 数值
- 科学计数
- 特殊
- 测试

可以使用以下方法创建表达式规则：

- 配置文件向导。创建或编辑列配置文件时，可以在配置文件向导中创建和应用表达式规则。通过将规则升级为可重用规则，您可以在多个配置文件中使用该规则。
- 规则规范。可以在 Analyst 工具中配置规则规范，并在列配置文件中使用该规则规范。配置规则规范时，请将业务规则的要求转换为一个或多个规则语句。规则语句表示用于确定数据集是否符合业务规则的逻辑。从规则规范生成 Mapplet，然后在 Developer tool 中创建的列配置文件中使用该 Mapplet。

可以使用表达式编辑器来添加表达式函数、将列配置为函数的输入、验证表达式，以及配置返回类型、精度和小数位数。创建并验证表达式规则后，可以编辑输出规则列的精度值。默认情况下，输出规则列的精度值设置为 10。输出规则列超出所设置的精度值时，精度值会被截断。

表达式规则的输出是虚拟列，该列使用规则名称作为列名称。Analyst 工具将对该虚拟列运行列配置文件。例如，您使用某一表达式规则来验证邮政编码。如果邮政编码有效，则该规则返回 1；如果邮政编码无效，则返回 0。Informatica Analyst 将对该规则的 1 和 0 输出值运行列配置文件。

## 创建表达式规则

可以使用**配置文件**向导来创建表达式规则，然后将其添加到配置文件。创建表达式规则，以验证配置文件中各列的值。

1. 打开某一配置文件。
2. 在摘要视图中，单击**操作 > 编辑配置文件**，以打开**配置文件**向导。
3. 单击**指定规则和筛选器**。
4. 在“规则”窗格中，单击**操作 > 添加规则**。

此时将显示**新建规则**对话框。

5. 在**新建规则**对话框中，输入规则的名称和可选说明。您可以在“函数”面板或“列”面板中创建规则。
  - 在“函数”面板中，选择函数类别，然后单击向右箭头 (>>) 按钮。在对话框中，指定参数并单击**确定**。函数将与列和值一起显示在“表达式”面板中。
  - 在“列”面板中，选择列，然后单击向右箭头 (>>) 按钮。该列将显示在“表达式”面板中。您可以添加函数、表达式和值，以创建规则。
6. 要验证规则，请单击**验证**。
7. 或者，选择将规则升级为可重用规则，然后配置项目和文件夹位置。如果将规则升级为可重用规则，则您或其他用户可以将该规则在其他配置文件中用作预定义规则。
8. 单击**确定**。

此时将显示**指定规则和筛选器**屏幕，并在“规则”窗格中列出规则。

## 使用规则规范创建表达式规则

可以在 Informatica Analyst 中使用规则规范创建表达式规则。可以将规则添加到列配置文件中来验证数据。

1. 在标题区域中，单击**新建 > 规则规范**。

此时将显示**新建规则规范**向导。
2. 在**新建规则规范**向导中，输入规则的名称和可选说明。
3. 在**位置**字段中，单击**浏览**以选择要将规则保存到的项目或文件夹。
4. 单击**继续**。

规则规范将显示在**设计**工作区中。


5. 要输入规则的属性，请选择该规则中的顶级八角形，然后单击**属性**。
6. 要配置主规则集，请单击该规则中的下一级矩形。
7. 要输入规则集的输入，请单击**属性 > 输入**。

此时将显示**输入管理**对话框。

8. 在**输入管理**对话框中，单击**添加输入**，然后键入输入的名称、数据类型、最大长度和说明。（可选）可以输入多个输入。
9. 单击**确定**。


输入将显示在**属性**部分中。

10. 要定义规则逻辑，请单击**规则逻辑**，输入运算符和条件，然后在**操作**列表中选择操作。
11. （可选）根据需输入多个规则集。

12. 要验证规则，请单击**验证** () 图标。

13. 要在列配置文件中保存并使用规则规范，请单击**保存并完成**。

14. 要保存并继续处理该规则，请单击**保存并继续**。

15. 要在 Developer tool 中使用该规则规范，请单击**生成规则** () 图标以生成 Mapplet。

Analyst 工具将在模型存储库中创建 Mapplet。使该 Mapplet 成为有效的规则，然后在 Developer tool 中创建的列配置文件中使用该 Mapplet。



## 第 6 章

# Informatica Analyst 中的筛选器

本章包括以下主题：

- [Informatica Analyst 中的筛选器概览, 33](#)
- [创建筛选器, 33](#)
- [管理筛选器, 36](#)

## Informatica Analyst 中的筛选器概览

可以创建一个筛选器，以便能够创建符合筛选条件的原始数据源子集。随后，可以对筛选后的数据运行配置文件。

您可以创建筛选器，以查看满足筛选条件的配置文件结果。您可以基于摘要视图中可用的默认筛选器查看配置文件结果。

## 创建筛选器

可以创建一个筛选器，以便能够创建符合筛选条件的原始数据源子集。

1. 打开某一配置文件。
2. 在摘要视图中，单击**操作 > 编辑配置文件**。  
此时将显示**配置文件**向导。
3. 单击**指定规则和筛选器**。
4. 在**筛选器**窗格中，单击**操作 > 添加筛选器**。  
此时将显示**新建筛选器**对话框。
5. 创建简单筛选器、高级筛选器或 SQL 筛选器。  
**注意：**对于日期列上的简单或高级筛选器，请以 YYYY/MM/DD HH:MM:SS 格式提供条件。  
**数据预览**窗格将显示满足筛选条件的原始数据源的子集。
6. 单击**确定**。  
此时将显示**指定规则和筛选器**屏幕，并在**筛选器**窗格中列出筛选器。

# 创建简单筛选器

可以使用 =、!=、>、< 等条件运算符创建简单的筛选器。使用筛选器可创建原始数据源的子集。

1. 在**新建筛选器**对话框中，单击**简单**。

下图显示了**新建筛选器**对话框中可用来创建简单筛选器的选项：

2. 输入名称和可选说明。
3. 选择一列。
4. 选择条件运算符。
5. 输入一个值。
6. （可选）单击加号 (+) 图标添加更多筛选器。
7. 单击**确定**。

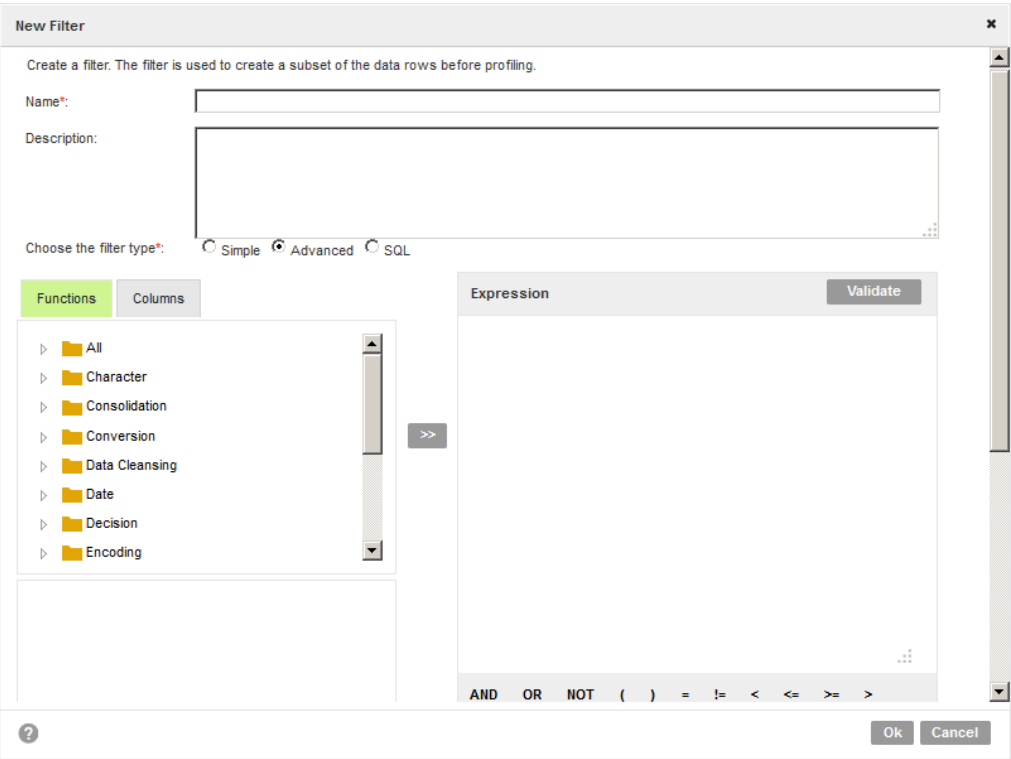
此时将显示**指定规则**和**筛选器**页面，并在“筛选器”窗格中列出筛选器。

# 创建高级筛选器

您可以使用诸如 AND、OR 和 NOT 等表达式创建高级筛选器，以创建原始数据源的子集。

1. 在**新建筛选器**对话框中，单击**高级**。

下图显示了**新建筛选器**对话框中的高级筛选器选项。



2. 输入高级筛选器的名称和可选说明。
3. 您可以在“函数”面板或“列”面板中创建高级筛选器。
  - 在“函数”面板中，选择函数类别，然后单击向右箭头 (>>) 按钮。  
在对话框中，指定参数并单击**确定**。函数将与列和值一起显示在“表达式”面板中。
  - 在“列”面板中，选择列，然后单击向右箭头 (>>) 按钮。该列将显示在“表达式”面板中。  
您可以添加函数、表达式和值，以创建高级筛选器。
4. 要验证高级筛选器，请单击**验证**。
5. 单击**确定**。

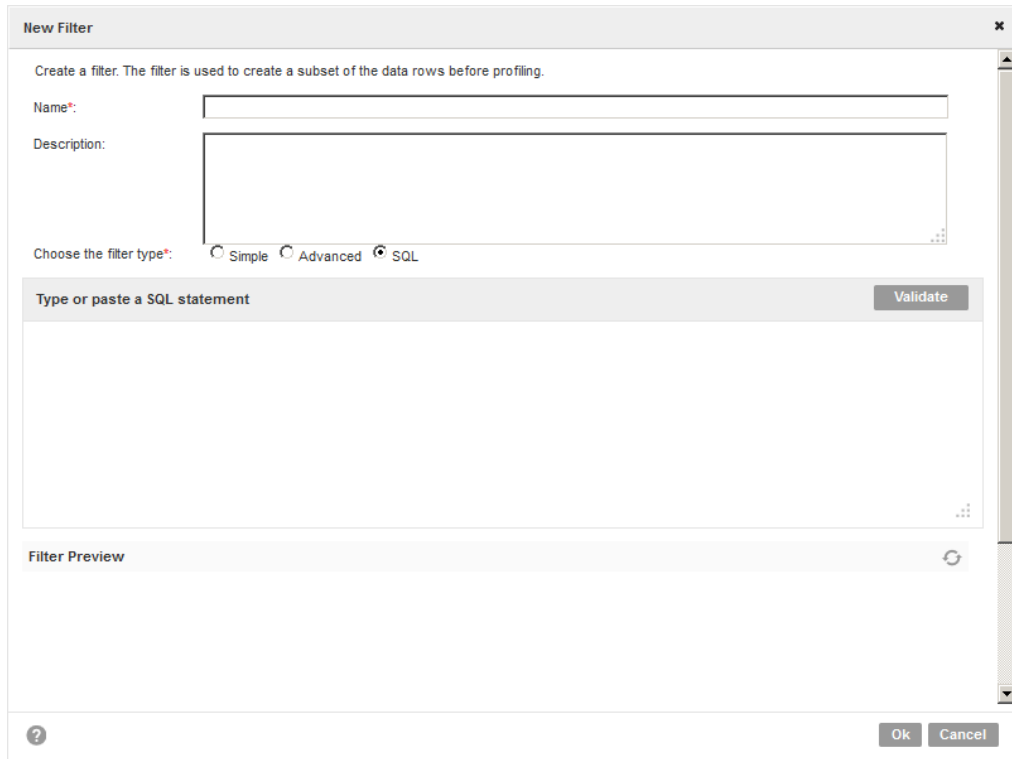
此时将显示**指定规则**和**筛选器**屏幕，并在“筛选器”窗格中列出筛选器。

## 创建 SQL 筛选器

您可以使用 SQL 查询创建 SQL 筛选器。可以为关系数据源创建 SQL 筛选器。

1. 在**新建筛选器**对话框中，单击 **SQL**。

下图显示了**新建筛选器**对话框中的 SQL 筛选器选项。



The screenshot shows the 'New Filter' dialog box. At the top, it says 'Create a filter. The filter is used to create a subset of the data rows before profiling.' Below this are fields for 'Name\*' and 'Description:'. Underneath, there are radio buttons for 'Simple', 'Advanced', and 'SQL', with 'SQL' being selected. Below the radio buttons is a section titled 'Type or paste a SQL statement' with a 'Validate' button. At the bottom, there is a 'Filter Preview' section with a refresh icon. The dialog box has 'Ok' and 'Cancel' buttons at the bottom right.

2. 输入 SQL 筛选器的名称和可选说明。
3. 在文本框中，键入或粘贴 SQL 查询。
4. 单击**验证**以验证 SQL 查询。
5. 单击**确定**。

此时将显示**指定规则和筛选器**页面，并在“筛选器”窗格中列出 SQL 筛选器。

## 管理筛选器

可以编辑和删除筛选器。

1. 在**库**工作区中，选择包含配置文件的项目，或者在**资产**窗格中选择要筛选的配置文件。
2. 打开某一配置文件。
3. 在摘要视图中，单击**操作 > 编辑配置文件**，以打开**配置文件**向导。
4. 单击**指定规则和筛选器**。
5. 在“筛选器”窗格中，选择筛选器，然后单击**操作 > 编辑筛选器**。  
此时将显示**编辑筛选器**对话框。
6. 编辑筛选器设置，然后单击**确定**。

7. 要删除筛选器，请选择该筛选器，然后单击**操作 > 删除筛选器**。

## 第 7 章

# Informatica Analyst 中的列配置文件结果

本章包括以下主题：

- [Informatica Analyst 中的列配置文件结果概览, 38](#)
- [摘要视图, 39](#)
- [详细视图, 41](#)
- [统计信息, 42](#)
- [配置文件运行的类型, 48](#)
- [比较多个配置文件结果概览, 49](#)
- [列配置文件向下钻取, 53](#)
- [Analyst 工具中的内容管理, 54](#)
- [Informatica Analyst 中的列配置文件导出文件, 55](#)

## Informatica Analyst 中的列配置文件结果概览

查看配置文件结果可了解并分析数据的内容、结构和质量。您可以在摘要视图中查看配置文件中的所有列和规则。可以在详细视图中查看列或规则的详细属性。

您可以在**发现**工作区下查看配置文件结果。视图标题显示了配置文件类型、配置文件中的列数和规则数以及采样数据与创建日期和时间。

在摘要视图中，每个列的属性可以显示为值、水平条形图或百分比。您可以查看列属性，例如，空值、相异值、非相异值、模式、数据类型和数据域。您可以基于默认筛选器在摘要视图中查看配置文件结果。

在详细视图中，您可以在各个窗格中查看空值、相异值、非相异值、推理的数据类型、推理的数据域、推理的模式、值、业务术语，以及预览数据。

您可以查看最新运行、历史运行与合并的运行的配置文件结果。还可以比较两次配置文件运行的配置文件结果，并在摘要视图和详细视图中查看结果。您可以查看配置文件统计信息并管理数据。配置文件统计信息包括列和规则的值、模式、数据类型、离群值和统计信息。您可以对数据执行数据发现和向下钻取。

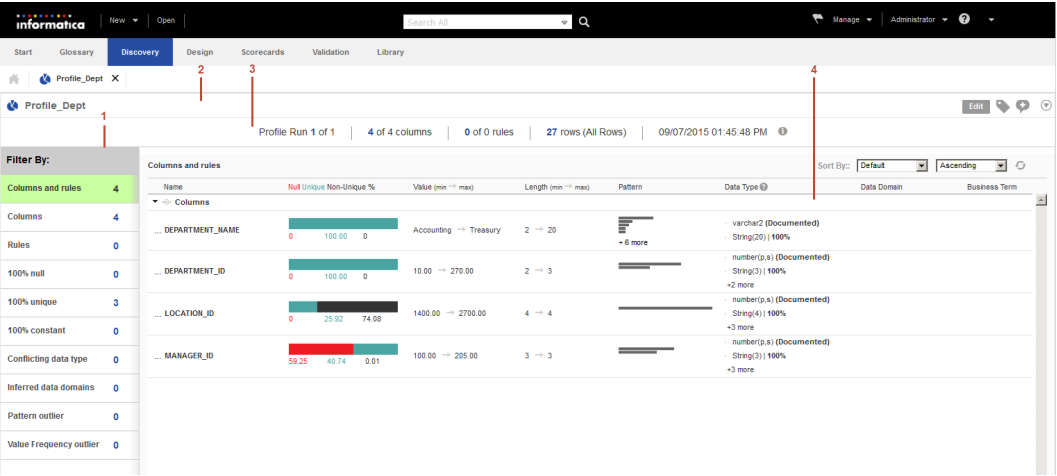
**注意：**您可以查看并运行基于 Avro、JSON、Parquet 和 XML 数据源的配置文件。还可以查看最新运行、历史运行与合并的运行的配置文件结果，以及比较两次配置文件运行的配置文件结果。

您可以将值频率、模式频率、向下钻取数据、注释、标记和业务术语导出到 CSV 文件。可以将配置文件摘要信息导出到 Microsoft Excel 文件，以便在一个文件中查看所有数据并进行进一步分析。您可以在配置文件结果中查看规则信息。显示的配置文件结果取决于配置文件配置和采样选项。

# 摘要视图

配置文件结果摘要以网格格式显示在摘要视图中。您可以在摘要视图使用默认筛选器查看特定统计信息。例如，选择“规则”时，摘要视图将显示配置文件中的所有规则。

下图显示了摘要视图的图形视图示例：



1. 默认筛选器。在摘要视图中，可以基于默认筛选器查看配置文件结果。
2. 配置文件表头。可以在表头中查看配置文件名。可以使用“编辑”按钮编辑配置文件，使用标记和注释图标来添加或编辑标记和注释，以及从“操作”菜单中选择相应选项。
3. 摘要视图表头。可以在摘要视图表头中查看配置文本特定信息。可以查看配置文件运行编号、配置文件总运行次数、列数、规则数以及配置文件中的行数。
4. 摘要视图。可以查看配置文件中所有列和规则的属性。

在摘要视图中，可以运行或编辑配置文件，检测模式离群值或值频率离群值，向结果卡添加列，选择配置文件运行，比较两次配置文件运行，将配置文件结果或数据域发现结果导出到 Microsoft Excel 电子表格，验证多个列的推理结果，添加或删除注释和标记，或者查看配置文件属性。

## 摘要视图属性

摘要视图显示了配置文件中所有列和规则的属性。摘要视图包含属性的可视化表示形式。您可以单击每个摘要属性，对属性值进行排序。

下表介绍了配置文件结果摘要属性：

属性	说明
名称	显示配置文件中的列或规则的名称。
空值 相异 非相异百分比	以百分比形式显示列或规则输出的空值、相异值和非相异值。值可以显示为水平条形图。
模式	将列中的多个模式显示为水平条形图。将鼠标指针悬停于条形图上方时，列中的模式字符和相似模式的个数可以显示为百分比。
值	显示列或规则输出的最小值和最大值。
长度	显示列或规则输出中的值的最小长度和最大长度。

属性	说明
数据类型	<p>显示列或规则的已记录的数据类型。在您将鼠标指针悬停于字段上方时显示推理的数据类型。Analyst 工具可以推理以下数据类型：</p> <ul style="list-style-type: none"> <li>- 字符串</li> <li>- 变长字符型</li> <li>- 小数</li> <li>- 整型</li> <li>- 日期</li> </ul> <p>您也可以基于推理的数据类型查看遵从性百分比。</p> <p><b>注意:</b> Analyst 工具无法从精度高于 38 的数值列的值派生数据类型。Analyst 工具无法从精度高于 255 的字符串列的值派生数据类型。如果您基于日期列创建了年份值早于 1800 年的列配置文件，则推理的数据类型可能会以固定长度字符串的形式显示。根据需要更改 InferDateTimeConfig.xml 中最小年份参数的默认值。</p>
数据域	显示与列关联的数据域的名称，以及遵从性百分比和遵从行数。
业务术语	显示分配给列的业务术语。

## 摘要视图中的默认筛选器

您可以基于默认筛选器在摘要视图中查看配置文件结果。

摘要视图默认显示所有源列、虚拟列和规则列的配置文件结果。“筛选依据”窗格会显示可将默认筛选器应用到的列的数量。

在摘要视图中，您可以使用以下默认筛选器选项来查看配置文件结果：

默认筛选器选项	说明
列和规则	显示源列和规则列的配置文件结果。您可以展开和折叠源列与规则列来查看结果。
列	显示源列的配置文件结果。
规则	显示规则列的配置文件结果。
100% 空值	显示值全部为空值的列的配置文件结果。
全部相异	显示值全部为相异值的列的配置文件结果。
100% 常量	显示所有记录都具有相同值的列的配置文件结果。例如，如果“国家/地区”列仅包含值“USA”，则“100% 常量”筛选器将包括该列的配置文件结果。
冲突数据类型	显示已记录的数据类型与推理的数据类型不匹配的列的配置文件结果。例如，筛选器将显示列 CustomerTier，因为该列的已记录的数据类型为“整数 (2)”，而推理的数据类型为“字符串”。
推理的数据域	显示推理的数据域与配置的数据域相同的列的配置文件结果。
模式离群值	显示具有模式离群值的列的配置文件结果。
值频率离群值	显示具有值离群值或频率离群值的列的配置文件结果。

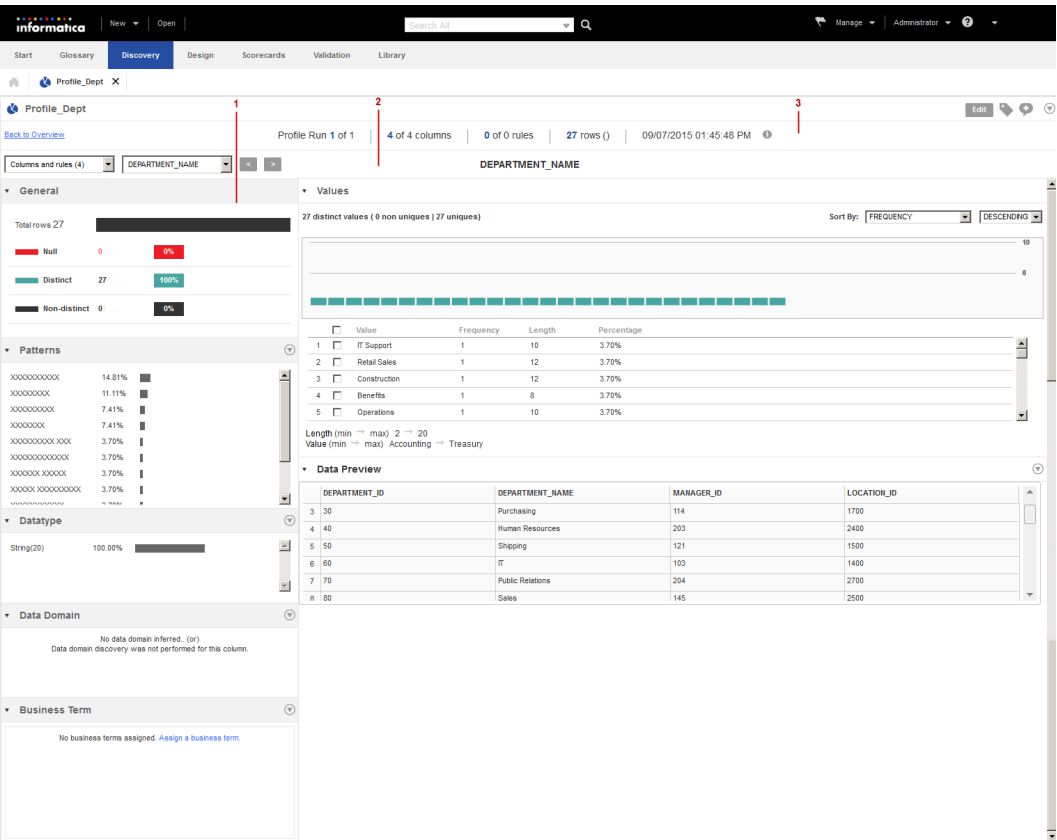


# 详细视图

列结果将显示在详细视图中。可以详细查看列属性。

单击摘要视图中的某个列后，即会显示该列的详细视图。

下图是详细视图中列属性的一个图形视图示例：



1. 窗格。可以在窗格中查看常规属性、列值、数据预览、推理的模式、推理的数据类型、推理的数据域以及业务术语。
2. 列详细信息表头。可以通过在下拉列表中选择列或者使用导航按钮来查看列结果。
3. 摘要视图表头。可以在摘要视图表头中查看配置文本特定信息。可以查看配置文件运行、列数、规则、配置文件运行中的行数以及配置文件运行的时间和日期。

在详细视图中，可以运行或编辑配置文件，向结果卡添加列，选择配置文件运行，比较两次配置文件运行，将配置文件结果导出到 Microsoft Excel 电子表格，将值频率、模式频率、数据类型、所选值的向下钻取数据或所选模式的向下钻取数据导出到 csv 文件，将注释和标记添加到列或从列中将其删除，以及查看配置文件属性。

使用每个窗格中的“操作”菜单可对列属性执行进一步的操作。可以折叠或展开窗格。

## 详细视图窗格

详细视图会在各个窗格中显示如下列属性：相异值、非相异值和空值的个数与百分比以及模式、推理的数据类型、推理的数据域、值、数据预览和链接的业务术语。

单击列或规则时，将打开该列或规则的详细视图。

下表介绍了详细视图中的窗格：

窗格	说明
常规	以各种颜色显示包含空值、相异值和非相异值的行的数量。您可以查看以百分比表示的值。您可以通过迷你图查看每次连续的配置文件运行的常规值增加和减少情况。走势图会在折线图中显示最近连续五次运行配置文件时空值、相异值或非相异值的数量变化。将指针移动到每个配置文件运行对应的走势图时，您可以查看值数量和值百分比。可以向列添加标记和注释。
模式	显示列值的模式。模式在列中出现的频率显示为水平条形图，并以百分比表示。您可以对模式进行向下钻取，还可以将模式添加到引用表或使用所选模式创建数据域。
数据类型	显示列的推理的数据类型。数据类型在列中出现的频率显示为水平条形图，并以百分比表示。您可以对数据类型进行向下钻取，还可以批准、拒绝或充值所选的推理的数据类型。 <b>显示拒绝项</b> 选项显示了拒绝的推理的数据类型。
数据域	显示列的推理的数据域。您可以对数据域进行向下钻取，了解遵从行数、不遵从行数或具有空值的行数。可以批准、拒绝或重置数据域值。 <b>显示拒绝项</b> 选项显示了拒绝的数据域。您可以验证数据域值。
业务术语	显示列的已分配业务术语。您可为列分配或取消分配业务术语。
值	以图形表示形式显示列中的所有值以及频率、长度和百分比。您可以对每个值进行向下钻取。您可以将值添加到引用表，还可以创建值频率规则以及创建数据域。
数据预览	显示所选模式、数据类型、数据域或值的向下钻取数据。

# 统计信息

您可以查看配置文件中的列和规则的如下统计信息：值、模式、数据类型、数据域和离群值。

您可以在摘要视图中查看配置文件统计信息，而在摘要视图和详细视图中查看列统计信息。可以查看最新配置文件运行、历史配置文件运行与合并的配置文件运行的统计信息。还可以比较两次配置文件运行的配置文件结果，并在摘要视图和详细视图中查看配置文件和列的统计信息。

## 数据预览

您可以在“数据预览”窗格中查看所选模式、数据类型、数据域或值的向下钻取数据。

您可以在详细视图中查看“数据预览”窗格。单击摘要视图中的某列时即会显示详细视图，而且默认情况下会收起“数据预览”窗格。要查看列数据，请单击**操作 > 显示预览**。

下表说明了“数据预览”窗格中操作菜单中的选项：

选项	说明
添加到筛选器	创建向下钻取筛选器来筛选向下钻取数据，从而对分析结果子集中数据的不规范问题进行分析。
保存筛选器	保存向下钻取筛选器。
显示预览	显示源行。
导出数据	将向下钻取结果导出到 CSV 文件或 Microsoft Excel 文件。

## 数据类型

数据类型包括配置文件结果中每个列的所有推理的数据类型。

可以在摘要视图和详细视图查看数据类型。在摘要视图中，您可以查看已记录的数据类型和推理的数据类型。**冲突数据类型**筛选器可显示已记录的数据类型和推理的数据类型之间存在冲突的列。在详细视图中，您可以查看列的推理的数据类型。数据类型在列中出现的频率显示为水平条形图，并以百分比表示。您可以向下钻取、批准、拒绝或重置选定的推理的数据类型。“显示拒绝项”选项显示了拒绝的推理的数据类型。

下表介绍了数据类型的属性：

属性	说明
数据类型	显示配置文件中列的已记录的和推理的数据类型的列表。
频率	显示数据类型在列中出现的次数，以数字表示。
百分比	显示数据类型在列中出现的百分比。
向下钻取	根据列数据类型向下钻取到特定的源行。 <b>注意:</b> 如果选择了多个推理的数据类型，则无法执行向下钻取操作。
状态	指示数据类型的状态。状态有“已推理”、“已批准”或“已拒绝”。 <b>已推理</b> 指示 Analyst 工具推理的列数据类型。 <b>已批准</b> 指示列的已批准数据类型。批准数据类型后，该数据类型随即提交到模型存储库。 <b>已拒绝</b> 指示列的已拒绝数据类型。

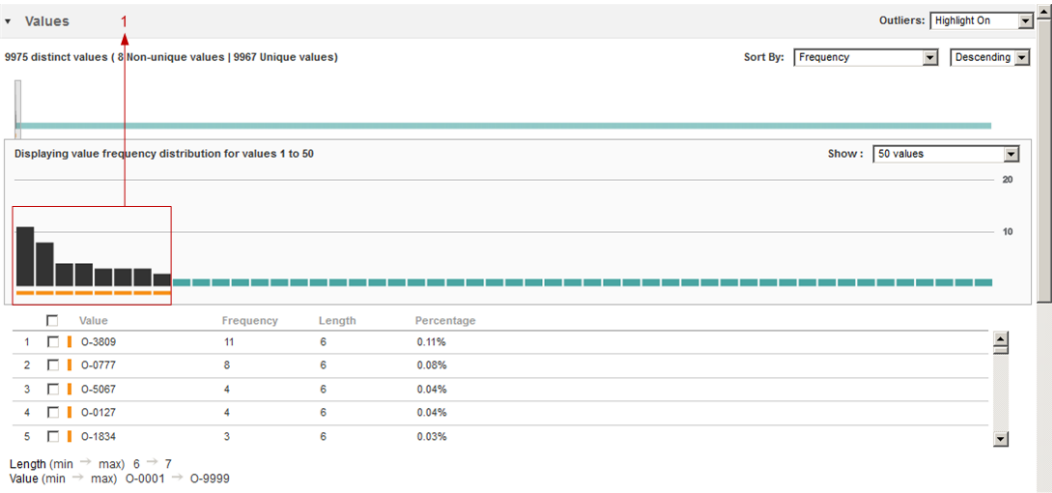
## 离群值

离群值是指配置文件结果中的列的模式、值或频率不在预期值范围内。

数据集成服务的剖析插件将运行某种算法，以确定不在列中大多数值的范围内的值。任何不在列中大多数值的预期范围内的模式、值或频率将被视为离群值。

默认情况下，Analyst 工具不会确定配置文件结果中的离群值。在摘要视图中，可以运行离群值以查看离群值结果。“模式”离群值筛选器会基于列中的模式显示离群值。“值频率”离群值筛选器会基于列中的值或频率显示离群值。离群值检测会在后台进行，以便您能在摘要视图中执行其他操作。

在详细视图中，如果从列表中选择了**突出显示**选项，则您可以在“值”窗格中查看离群值。离群值以带橙色下划线的竖条显示。要仅查看离群值，必须从列表中选择**筛选器**选项。



1. 离群值。离群值以带橙色下划线的竖条显示。

## 运行离群值

运行离群值可确定列中不在预期值范围内的模式、值或频率。

1. 在摘要视图中，单击**操作 > 检测离群**。  
筛选依据窗格中的“模式离群值”和“值频率离群值”会从 N/A 变为检测到的离群值数量。
2. 在筛选依据窗格中，单击**模式离群值**。  
具有模式离群值的列将显示在摘要视图中。
3. 在筛选依据窗格中，单击**值频率离群值**。  
具有值离群值或频率离群值的列将显示在摘要视图中。
4. 在详细视图中，从离群值下拉列表中选择**突出显示**。  
在“值”窗格中，离群值将显示为带橙色下划线的竖条。
5. 单击“离群值”下拉列表中的**筛选器**，以只查看离群值。

## 模式

可以在摘要视图和详细视图中查看列值的模式以及这些模式的出现频率。

在摘要视图中，列中的多个模式可以显示为水平条形图。将鼠标指针悬停于条形图上方时，列中的模式字符和相似模式的个数可以显示为百分比。在详细视图中，模式在列中出现的频率可以显示为水平条形图，并以百分比表示。您可以执行向下钻取，还可以将模式添加到引用表或使用所选模式创建数据域。

默认情况下，剖析仓库最多可以存储 16,000 个唯一的最高频率值，包括配置文件结果的空值。如果配置文件结果中至少有一个空值，则 Analyst 工具可能会将空值显示为模式。

**注意:** Analyst 工具无法为精度高于 38 的数字列派生模式。Analyst 工具无法为精度高于 255 的字符串列派生模式。

下表说明了列模式的属性：

属性	说明
模式	显示配置文件中列的模式。
频率	显示模式在列中出现的次数，以数字表示。
百分比	显示模式在列中出现的百分比。

下表介绍了模式字符及其表示的含义：

字符	说明
“B” 或 “b” 或 “ ”	表示空格。
“C” 或 “c”	表示任何字符。
“L” 或 “l”	表示任何小写字母字符。
“T” 或 “t”	表示 Tab 键。
“U” 或 “u”	表示任何大写字母字符。
9	表示任何数字字符。Informatica Analyst 以 “9” 的格式单独显示最多 3 个字符。Analyst 工具将 3 个以上字符显示为一个以括号括起的值。例如，格式 “9(8)” 表示包含 8 位数的数值。
“X” 或 “x”	表示任何字母字符。Informatica Analyst 以 “X” 的格式单独显示最多 3 个字符。Analyst 工具将 3 个以上字符显示为一个以括号括起的值。例如，格式 “X(6)” 可以表示值 “Boston”。 <b>注意：</b> 模式字符 X 不区分大小写，并且可以表示源数据中的大写字符或小写字符。
“P” 或 “p”	表示 “(”，即左括号。
“Q” 或 “q”	表示 “)”，即右括号。

**注意：**列模式可以包含特殊字符。例如，~、[、]、=、-、?、=、{、\*、-、>、< 和 \$。

## 值

您可以查看列的值，以及值在列中出现的频率。

可以在摘要视图中查看列中的最小值和最大值。在详细视图中，您可以查看列的值属性。

### 摘要视图中的值

您可以在摘要视图中查看最新配置文件运行、历史配置文件运行与合并的配置文件运行的所有列和规则的最小值和最大值。

示例

假设某家零售店的数据库的“员工”表中名为“员工 ID”的列填充了范围在 100 到 250 之间的员工 ID，并且该数据库还包含姓名（例如 Bob 和 Robert），则当您为“员工”表运行列配置文件时，摘要视图中的“员工 ID”的“值”列将显示 “100 --> Robert”。

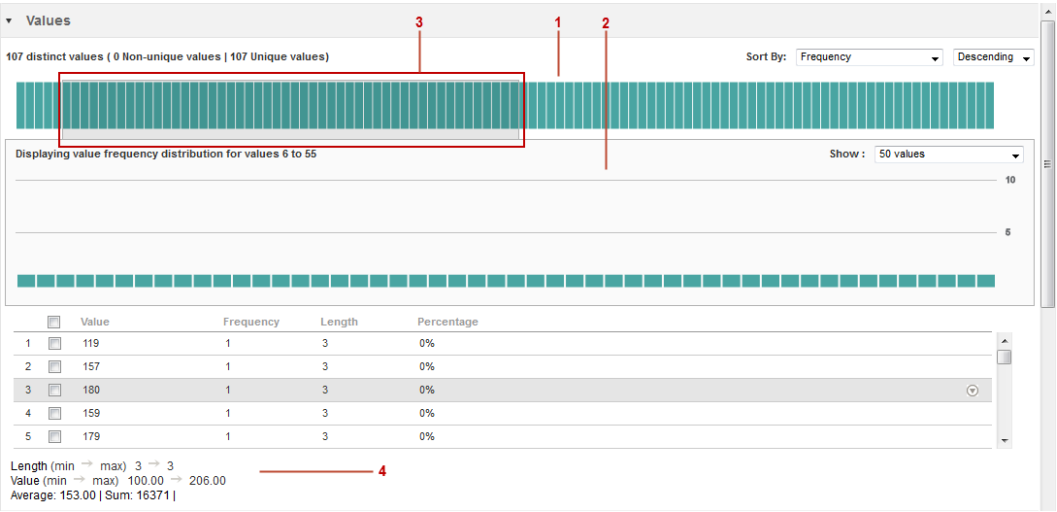
详细视图中的值

详细视图中的列值包括列的值，以及值在列中出现的频率。

值窗格以图形表示形式显示列值。您可以查看每个值的频率、长度和百分比。可以基于值或频率对值进行排序。可以对数据执行向下钻取以及将值添加到引用表，还可以创建值频率规则或创建数据域。 您可以查看用红色竖条表示的空值，用黑色竖条表示的值频率，以及用橙色竖条突出显示的离群值。您可以突出显示离群值、禁用离群值或将结果筛选为只显示列中的离群值。  
“值”窗格包含图形布局和价值部分。

图形布局分为两个面板。

下图显示了详细视图中的“值”窗格：



1. 上方面板。您可以查看用垂直条形图表示的各种值。您可以按频率和值对各种值进行排序。您可以按照升序或降序顺序对值进行排序。您可以查看用橙色竖条突出显示的离群值。
2. 下方面板。您可以在下方面板中查看滑块中的值，其中的每个值都用竖条表示。可以对值执行向下钻取以及将值添加到引用表，还可以创建值频率规则，以及为值创建数据域。您可以同时查看 50、75 或 100 个值。
3. 滑块。在上方面板中，您可以将滑块滑动至各个值上方。下方面板会显示滑块中的值。
4. 值属性。值属性部分会显示值和属性。

下表介绍了图形布局中的面板：

面板	说明
上方面板	以垂直条形图形式显示所有值。上方面板最多可以显示 16,000 个值。 可以使用滑块查看一系列值。
下方面板	显示您在上方面板中选择的批次的值。默认情况下，Analyst 工具显示 50 个值。您可以选择同时查看 75 或 100 个值。

下表介绍了值部分中的列值的属性：

属性	说明
值	显示您在上方面板中选择的批次的值列表。 <b>注意:</b> Analyst 工具会从列值中排除 CLOB、BLOB、“原始”和“二进制”等数据类型。
频率	显示值在列中出现的次数，以数字表示。
长度	显示列值的长度。
百分比	显示值在列中出现的百分比。

下表介绍了所选列的统计信息：

统计信息	说明
长度(最小值 - 最大值)	显示列中最短值和最长值的长度。
值(最小值 - 最大值)	显示列中的最小值和最大值。
平均值	显示列的值的平均值。
总和	显示列中所有值的总和。

### 配置文件结果比较的详细视图中的值

配置文件结果比较的详细视图中的“值”窗格显示了如下值属性：非重复值的个数、最小值、最大值、最大长度、最小长度、平均值、标准偏差以及值总和。

配置文件结果比较的列的详细视图在水平条形图中显示值属性、值和值频率。

下表介绍了比较两次配置文件运行的结果时详细视图中的列值的属性：

属性	说明
非重复值的个数	显示列中非重复值的个数。
最小值	显示列中的最小值。
最大值	显示列中的最大值。
长度(最小 - 最大)	显示列中最短值和最长值的长度。
平均值	显示列的值的平均值。
标准偏差	显示所有列值之间的标准偏差或变异性。
总和	显示列中所有值的总和。

# 配置文件运行的类型

可以查看最新配置文件运行、历史配置文件运行与合并的配置文件运行的配置文件结果。您可以在摘要视图中查看配置文件运行结果。

## 最新配置文件运行

可以在摘要视图中查看最新配置文件运行的配置文件结果。

在以下情况下，可以在摘要视图中查看最新配置文件运行的配置文件结果：

- 创建、保存和运行配置文件。
- 在**库**工作区中打开之前运行的配置文件。
- 对于合并的配置文件运行，在摘要视图或详细视图中单击**返回最新的配置文件运行**链接。
- 对于历史配置文件运行，在摘要视图或详细视图中单击**返回最新的配置文件运行**链接。
- 在**选择配置文件运行**对话框中选择最新配置文件运行，然后单击**确定**。

## 历史配置文件运行

可以在摘要视图中查看先前的配置文件运行的配置文件结果。

剖析仓库会保存配置文件的所有运行的配置文件结果。您可以通过在“选择配置文件运行”对话框中选择先前版本的配置文件运行来查看该配置文件运行的结果。

## 已合并配置文件运行

可以在摘要视图中查看配置文件中每个列的最新配置文件结果。

在合并的配置文件运行中，您可以查看配置文件中每个列的最新结果。如果在**选择配置文件运行**对话框中选择已合并配置文件运行，则剖析仓库会从所有配置文件运行中检索最新列结果。您可以在摘要视图中查看结果，摘要视图标题会显示增量配置文件运行。

### 示例

作为数据分析人员，您可以查看配置文件中每个列的最新结果。例如，您可以选择列 1、2 和 3 以执行配置文件运行 A，并选择列 3、4 和 5 以执行配置文件运行 B。要查看所有列的最新结果，您可以在“选择配置文件运行”对话框中选择合并的配置文件运行。摘要视图会显示运行 A 返回的列 1 和 2 的结果，并显示运行 B 返回的列 3、4 和 5 的结果。

## 选择配置文件运行

您可以选择历史配置文件运行、最新配置文件运行或合并的配置文件运行，以查看配置文件结果。您可以在摘要视图中查看配置文件结果，而在详细视图中查看列结果。

1. 在**库**工作区中，选择包含配置文件的项目或文件夹，或者在**资产**窗格中选择配置文件。

2. 单击**操作 > 打开**，以打开配置文件。

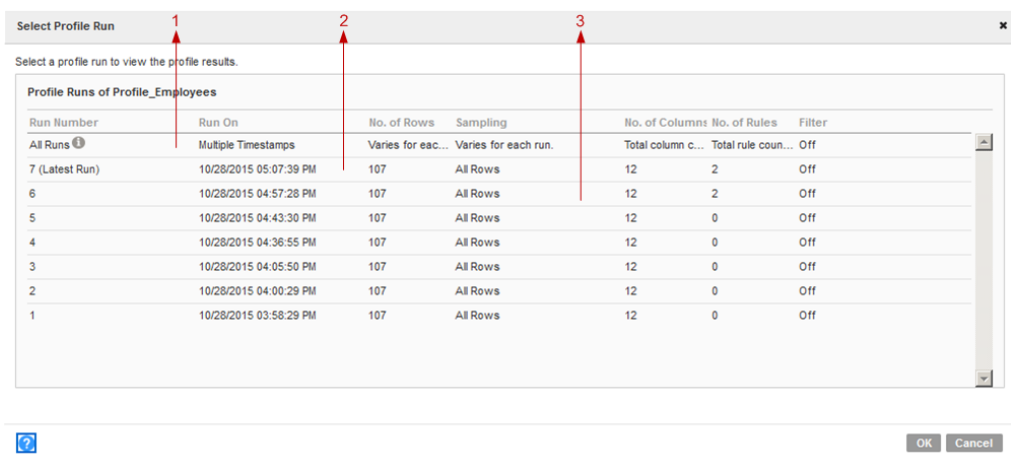
摘要视图将显示在**发现**工作区中。

3. 在摘要视图中，单击**操作 > 选择配置文件运行**。

此时将显示**选择配置文件运行**对话框。

下图显示了**选择配置文件运行**对话框。





1. 合并的配置文件运行。选择此配置文件运行时，可以在摘要视图中查看每列的最新配置文件结果。
  2. 最新配置文件运行。选择此配置文件运行时，您可以在摘要视图中查看配置文件的最新配置文件结果。
  3. 历史配置文件运行。选择此配置文件运行时，您可以在摘要视图中查看先前的配置文件运行的历史配置文件结果。
  4. 在**选择配置文件运行**对话框中，选择其中一次配置文件运行以查看其配置文件结果：
    - 要查看最新配置文件运行的配置文件结果，请选择最新配置文件运行，然后单击**确定**。
    - 要查看历史配置文件运行的配置文件结果，请选择最新配置文件运行以外的配置文件运行，然后单击**确定**。
    - 要查看合并的配置文件运行的配置文件结果，请单击**全部运行**，然后单击**确定**。每个列的最新配置文件结果将显示在摘要视图中。
- Analyst 工具将执行配置文件运行，并在摘要视图中显示配置文件结果。
5. 在摘要视图中，单击一个列以查看列结果。  
此时将显示详细视图。

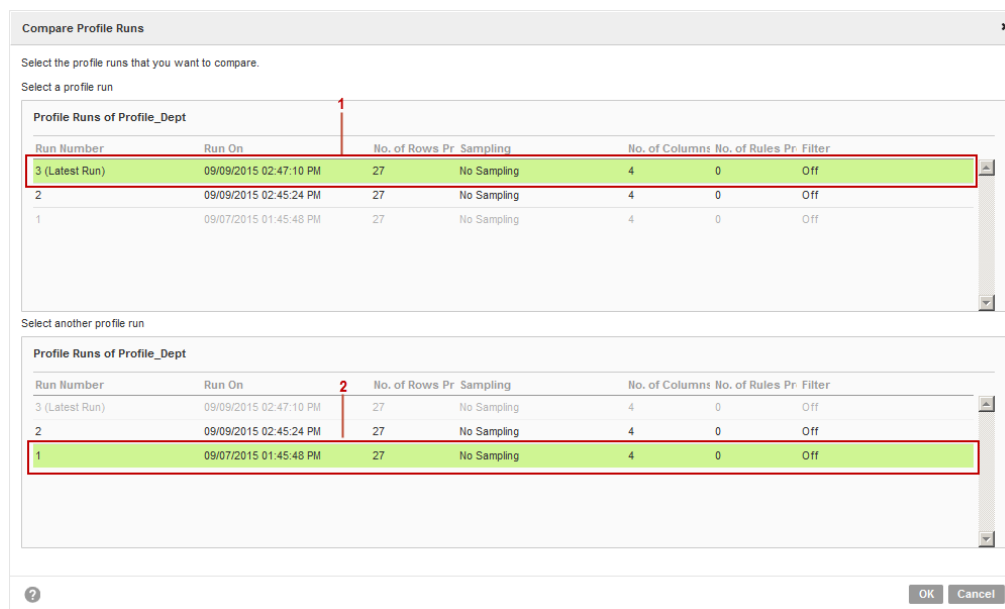
## 比较多个配置文件结果概览

可以比较两次配置文件运行的配置文件结果。您可以在摘要视图中查看比较结果，在详细视图中查看列结果。  
在摘要视图中，您可以查看两次配置文件运行的所有列的比较结果。

### 比较多个配置文件结果

比较两次配置文件运行后，您可以在摘要视图中查看配置文件结果比较。

1. 在摘要视图中，单击**操作 > 比较配置文件运行**。  
下图显示了**比较配置文件运行**对话框。

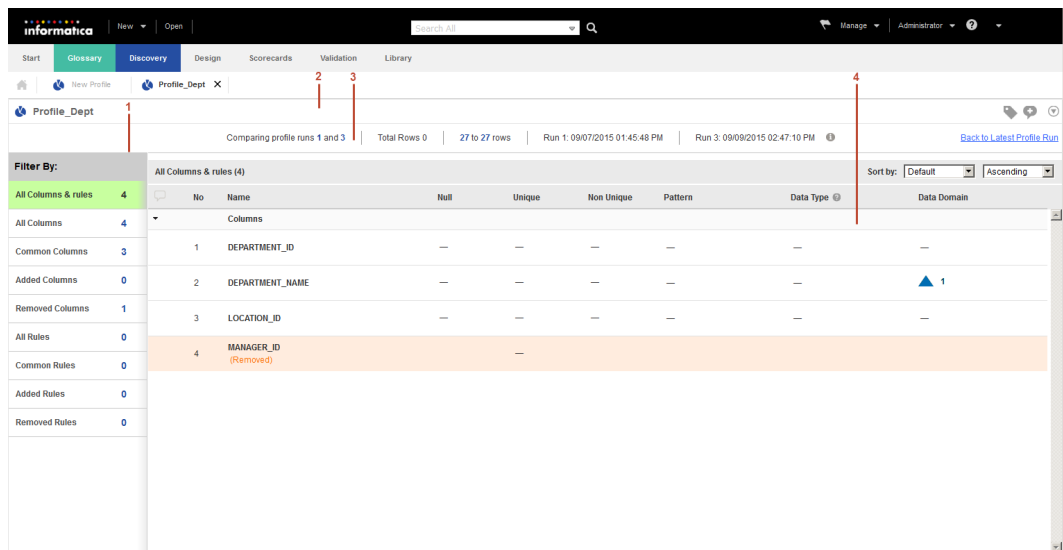


1. 运行 A。选择一次配置文件运行作为“运行 A”。
  2. 运行 B。选择一次配置文件运行作为“运行 B”。
- 此时将显示**比较配置文件运行**对话框。
2. 从**运行 A** 窗格中选择一个配置文件，并从**运行 B** 窗格中选择另一个配置文件。
  3. 单击**确定**。
- 摘要视图将显示配置文件结果的合并视图。

## 配置文件结果比较的摘要视图

比较两次配置文件运行后，您可以在摘要视图中查看网格格式的结果。您可以在摘要视图使用默认筛选器查看特定统计信息。

下图显示了摘要视图中两次配置文件运行的配置文件比较结果：



1. 默认筛选器。在摘要视图中，您可以基于默认筛选器查看配置文件比较结果。
2. 配置文件表头。可以在表头中查看配置文件名。
3. 摘要视图表头。可以在摘要视图表头中查看配置文本特定信息。可以查看所比较的配置文件运行、配置文件各次运行相比行数的增加或减少、配置文件中的行数，以及配置文件运行的时间和日期。
4. 摘要视图。可以查看两次配置文件运行中列之间的比较结果。

## 配置文件结果比较的摘要视图属性

配置文件结果比较的摘要视图属性包括相异值、非相异值和空值的个数与百分比以及模式、推理的数据类型、推理的数据域和链接的业务术语。摘要视图包含属性的可视化表示形式。您可以单击每个摘要属性，对属性值进行排序。

在摘要视图中，数据集成服务按升序为所有列和规则分配编号。

**注意:** 带有计数的向上箭头指示其中一次配置运行的属性值比另一次增加了。带有计数的向下箭头指示其中一次配置运行的属性值比另一次减少了。

下表介绍了配置文件结果比较的摘要属性：

属性	说明
编号	显示列或规则的编号。
名称	显示配置文件中的列或规则的名称。
空值	显示空值的增加或减少。
相异	显示相异值的增加或减少。
非相异	显示非相异值的增加或减少。
模式	显示两次配置文件运行之间的模式差异。
数据类型	显示两次配置文件运行之间列或规则的推理的数据类型的差异。
数据域	显示两次配置文件运行之间与列或规则关联的推理的数据域的差异。

## 摘要视图中配置文件结果比较的默认筛选器

您可以基于默认筛选器在摘要视图中查看配置文件结果。

在摘要视图中，您可以查看源列和虚拟列。规则的输出在摘要视图中显示为虚拟列。更改规则的输出端口并将配置文件运行与历史运行进行对比时，历史规则输出列会显示在**移除的规则**筛选器中，并且新规则输出列会显示在**添加的规则**筛选器中。如果更改单个输出规则的规则逻辑，或者如果更改配置文件运行中的多个规则输出的输入，并将其与历史运行进行比较，**添加的规则**和**移除的规则**筛选器输出不会发生更改。筛选器输出不会发生更改是因为筛选器仅将对列的名称更改视为对筛选器的有效输入。

您可以使用以下默认筛选器选项来查看满足特定条件的配置文件结果：

默认筛选器选项	说明
所有列和规则	显示源列、虚拟列和规则列的配置文件结果。您可以展开和折叠源列与规则列来查看结果。
所有列	显示源列和虚拟列的配置文件结果。
共同列	显示在两个配置文件运行结果中都可用的列。
添加的列	显示在最新配置文件运行中可用的列。例如，如果将运行 5 与运行 3 相比较，则“添加的列”会显示在运行 5 中可用而在运行 3 中不可用的列。
移除的列	显示在历史配置文件运行中可用的列。例如，如果将运行 5 与运行 3 相比较，则“移除的列”会显示在运行 3 中可用而在运行 5 中不可用的列。
所有规则	显示所有规则列的配置文件结果。
添加的规则	显示在最新配置文件运行中可用的规则。例如，如果将运行 5 与运行 3 相比较，则“添加的规则”会显示在运行 5 中可用而在运行 3 中不可用的规则。
移除的规则	显示在历史配置文件运行中可用的规则。例如，如果将运行 5 与运行 3 相比较，则“移除的规则”会显示在运行 3 中可用而在运行 5 中不可用的规则。

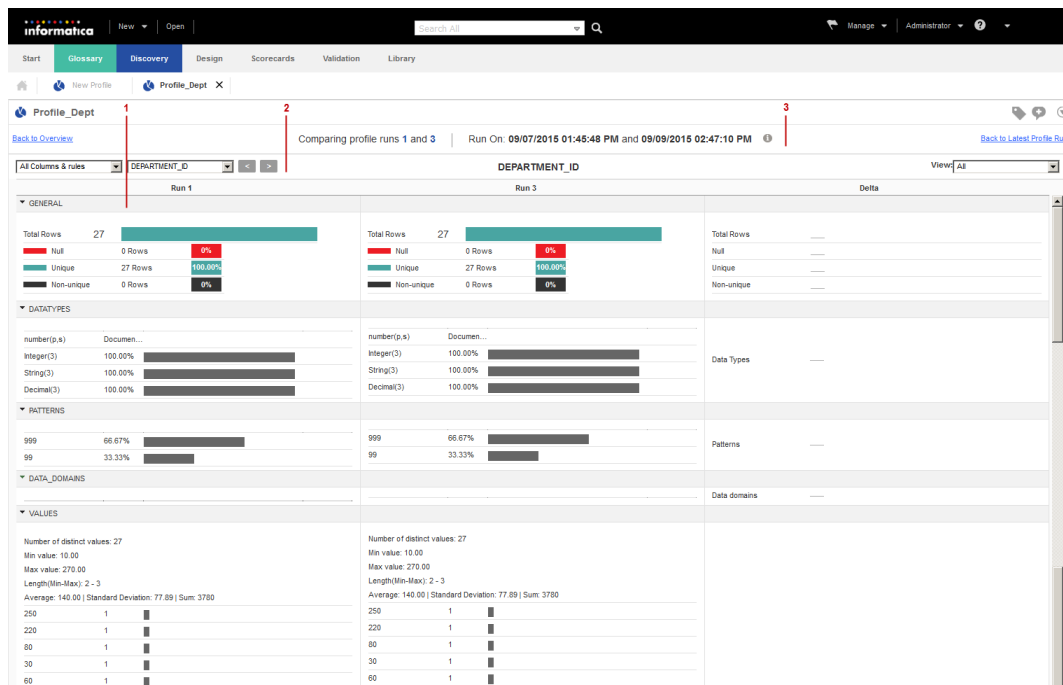
摘要视图默认显示所有源列和虚拟列的配置文件结果。

## 配置文件结果比较的详细视图

列结果以网格格式显示在详细视图中。列详细信息包含常规信息，例如，相异值、非相异值、空值、模式、数据类型、数据域、业务术语、值和数据预览。

单击列名称时，将显示该列的详细视图。您可以在单独的列中查看运行 A 和运行 B 的列结果，两列数据的比较结果将显示在增量列中。

下图显示了详细视图中某个列的配置文件结果比较：



1. 窗格。在窗格中，可以查看配置文件结果以及两次配置文件运行中列的统计信息，还可查看两次配置文件运行中列的增量信息。
2. 配置文件表头。可以通过在下拉列表中选择列或者使用导航按钮来查看列结果。可以查看列名称，还可使用“视图”下拉列表中的选项来查看特定结果。
3. 摘要视图表头。可以在摘要视图表头中查看配置文本特定信息。可以查看所比较的配置文件运行以及配置文件运行的时间和日期。

## 配置文件结果比较的详细视图窗格

详细视图详细地显示了两次配置文件运行的配置文件结果和某个列的比较结果。

详细视图显示了运行 A 和运行 B 的列结果，而数据比较会显示在增量列中。要查看其他列结果，您可以从筛选器下拉列表中选择筛选器，或从列下拉列表中选择列。

# 列配置文件向下钻取

在列配置文件中，使用向下钻取选项可根据列值向下钻取到数据源中的特定行。可以选择读取数据源中的当前数据进行向下钻取，或者读取暂存在剖析仓库中的配置文件数据。当向下钻取到暂存配置文件数据中的特定行时，Analyst 工具将创建一个向下钻取筛选器以用于匹配列值。在向下钻取后，可以编辑、重新调用、重置和保存向下钻取筛选器。

可以选择要进行向下钻取的列，即使您并未选择这些列进行剖析也可以这样做。可以选择读取数据源中的当前数据进行向下钻取，或者读取暂存在剖析仓库中的配置文件数据。在对某一列值执行向下钻取后，可将所选值或模式的向下钻取数据导出到位于所选位置的 CSV 文件。虽然 Informatica Analyst 仅显示向下钻取数据的前 200 个值，但该工具会将所有值导出到 CSV 文件。

## 向下钻取行数据

运行配置文件后，可以向下钻取到与列值、数据类型或模式匹配的特定行。

1. 运行配置文件。  
配置文件结果将显示在摘要视图中。
2. 在摘要视图中，单击一个列名称。  
列结果将显示在详细视图中。
3. 在详细视图中，右键单击**值**窗格中的值，然后选择**向下钻取**。  
**数据预览**窗格将显示向下钻取数据。

## 应用筛选器以向下钻取数据

可以按照迭代方式筛选向下钻取数据，以便能够针对配置文件结果的子集分析数据不规范性。

1. 在**值**选项卡上选择一个列值。
2. 右键单击并选择**向下钻取**。  
向下钻取结果将显示在**数据预览**窗格中。
3. 要添加筛选条件，请右键单击**数据预览**窗格中的某一列值，然后选择**添加到筛选器**。  
此时将显示**向下钻取筛选器**对话框，其中包含筛选条件。
4. 添加所需的筛选条件，然后单击**确定**。  
无法将向下钻取筛选器应用于推理数据类型。
5. 要保存筛选器，请单击**操作 > 保存筛选器**。
6. 要清除向下钻取筛选器，请单击**操作 > 刷新**。
7. 要将向下钻取数据导出到 Microsoft Excel 电子表格，请单击**操作 > 导出数据**。

## Analyst 工具中的内容管理

内容管理是验证和管理数据源的已发现元数据的过程，以便元数据适合使用和报告。在 Analyst 工具中管理元数据时，可以批准、拒绝和重置配置文件结果中的推理的数据类型或数据域。

您可为一个列批准一个数据类型和一个数据域。可以对列隐藏拒绝的数据类型或数据域。批准或拒绝推理的数据类型或数据域后，您可以重置数据类型或数据域以还原已推理状态。

## 批准数据类型和数据域

配置文件结果包含数据源中每个列的推理的数据类型和数据域。您可以在 Analyst 工具中选择并批准每个列的单个数据类型和单个数据域。

1. 运行配置文件。  
配置文件结果将显示在摘要视图中。
2. 在摘要视图中，单击一个列名称。  
列结果将显示在详细视图中。
3. 在详细视图中，在**数据类型**窗格中选择数据类型，或者在**数据域**窗格中选择数据域。

4. 单击**操作 > 批准**。
5. 要还原数据类型或数据域的已推理状态，请选择该数据类型或数据域，然后单击**操作 > 重置**。

## 拒绝数据类型和数据域

在详细视图中，您可以拒绝数据类型和数据域，还可以显示或隐藏拒绝的数据类型和数据域。

1. 运行配置文件。  
配置文件结果将显示在摘要视图中。
2. 在摘要视图中，单击一个列名称。  
列结果将显示在详细视图中。
3. 在详细视图中，在**数据类型**窗格中选择数据类型，或者在**数据域**窗格中选择数据域。
4. 单击**操作 > 拒绝**。  
Analyst 工具会从数据类型列表中删除拒绝的数据类型。
5. 要查看拒绝的数据类型，请单击**操作 > 显示拒绝项**。

# Informatica Analyst 中的列配置文件导出文件

可以将列配置文件结果导出到 CSV 文件或 Microsoft Excel 文件，这取决于您选择的是部分配置文件结果还是完整的结果摘要。

可将所选值和模式的值频率、模式频率、数据类型或向下钻取数据导出到 CSV 文件。可以将所有列的剖析结果摘要导出到 Microsoft Excel 文件。使用数据集成服务权限**向下钻取和导出结果**来确定是由用户还是由组来导出配置文件结果。

## CSV 文件格式的配置文件导出结果

可以导出值频率、模式频率、数据类型或向下钻取数据，以在文件中查看数据。Analyst 工具会将信息保存在 CSV 文件中。

在导出推理列模式时，Analyst 工具将导出不同格式的列模式。例如，在导出推理列模式 X(5) 时，Analyst 工具将在 CSV 文件中显示以下格式的列模式：XXXXX。

## Microsoft Excel 格式的配置文件导出结果

在导出完整配置文件结果摘要时，Analyst 工具会将信息保存至一个 Microsoft Excel 文件内的多个工作表中。Analyst 工具将以“xlsx”格式保存该文件。

下表描述了导出文件中的每个工作表上显示的信息：

选项卡	说明
列配置文件	摘要信息将在配置文件运行后从摘要视图中导出。示例包括列名称、规则名称、相异值数量、空值数量、推理的数据类型，以及上次运行配置文件的日期和时间。
值	列和规则的值以及这些值在每一列显示的频率。

选项卡	说明
模式	运行配置文件的列和规则的值模式以及这些模式显示的频率。
数据类型	列的所有数据类型、每种数据类型的频率、百分比值以及数据类型的状态（例如，“已推理”、“已批准”或“已拒绝”）。
统计信息	每一列和每条规则的统计信息。例如平均值、长度、最高值、最低值和标准偏差。
属性	“属性”视图信息，包括配置文件名称、类型、采样策略和行计数。

## 从 Informatica Analyst 中导出配置文件结果

可将配置文件的结果导出到“.csv”或“.xlsx”文件中，以便在一个文件中查看数据。

1. 在库工作区中，选择包含配置文件的项目或文件夹。
2. 单击配置文件将其打开。  
配置文件结果将显示在摘要视图中。
3. 在摘要视图中，单击**操作 > 导出数据**。  
此时会显示**将数据导出到文件**对话框。
4. 在**将数据导出到文件**对话框中，输入文件名。或者，使用默认文件名。
5. 选择**所有(摘要、值、模式、统计信息、属性)**或**数据域发现结果**，然后选择**代码页**。单击**确定**。  
数据将导出到 Microsoft Excel 电子表格。
6. 单击摘要视图中的列。  
列结果将显示在详细视图中。
7. 在详细视图中，单击**操作 > 导出数据**。  
此时会显示**将数据导出到文件**对话框。
8. 在**将数据导出到文件**对话框中，输入文件名。或者，使用默认文件名。
9. 选择以下选项之一：
  - 所有（摘要、值、模式、统计信息、属性）
  - 选定列的值频率。
  - 选定列的模式频率。
  - 选定列的数据类型。
  - 选定值的向下钻取数据。
  - 选定模式的向下钻取数据。
  - 选定数据类型的向下钻取数据。
10. 输入文件格式。对于**全部**选项，格式为 **Excel**；对于其余选项，格式为 **CSV**。可以选择将字段名称导出为文件中的第一行。
11. 选择文件的代码页。
12. 单击**确定**。  
数据将导出到文件。



## 第 8 章

# Informatica Analyst 中的结果卡

本章包括以下主题：

- [Informatica Analyst 结果卡概览, 57](#)
- [Informatica Analyst 结果卡进程, 58](#)
- [在 Informatica Analyst 中创建结果卡, 58](#)
- [向现有结果卡添加列, 59](#)
- [运行结果卡, 60](#)
- [查看结果卡, 61](#)
- [编辑结果卡, 61](#)
- [度量, 62](#)
- [度量组, 63](#)
- [对列进行向下钻取, 64](#)
- [趋势图表, 65](#)
- [Informatica Analyst 结果卡导出文件, 68](#)
- [结果卡通知, 68](#)
- [结果卡沿袭, 71](#)

## Informatica Analyst 结果卡概览

结果卡是配置文件中某个列的有效值的图形表示形式。可以创建结果卡，然后向下钻取实时数据或暂存数据。

使用结果卡测量数据质量进度。例如，可以创建一个结果卡，用于在应用数据质量规则之前度量数据质量。应用数据质量规则后，可以创建另一个结果卡，以比较这些规则对数据质量的影响。

结果卡将各列的值频率显示为得分。得分可以反映各列中有效值的百分比。运行配置文件后，可将配置文件中的列作为度量添加到结果卡中。可以创建度量组，以便能将相关度量分组到一个实体中。可以定义阈值，指定记录中的列可接受的错误数据的范围；还可以为每个度量分配度量权重。在运行结果卡时，Analyst 工具将为每个度量组生成加权平均值。要进一步评估数据质量，还可以为每个度量分配一个固定或可变成本。在运行结果卡时，Analyst 工具将为每个度量计算错误数据成本总和，然后显示总成本。

创建或编辑结果卡时，可以基于源数据创建结果卡筛选器。使用结果卡筛选器，可以基于筛选条件重新计算度量得分。要标识有效数据记录和无效记录，可以向下钻取每个度量。可以使用趋势图表来跟踪度量得分，以及度量中无效数据的成本在一段时间内如何变化。可以在结果卡中重复使用配置文件筛选器。

在 Analyst 工具中启用版本控制系统时，可以为结果卡创建多个版本并查看结果卡的版本历史记录。默认情况下，结果卡在创建后会签出。您必须签入结果卡，其他用户才能编辑结果卡。

可以在**结果卡**工作区中查结果卡仪表板。在结果卡仪表板中，您可以查看具有结果卡的数据对象、项目中的结果卡、过去六个月的结果卡运行趋势以及一个月内所有结果卡运行的良好、可接受和不可接受的度量汇总。

您可以在 Informatica Analyst 中配置和管理结果卡的电子邮件通知。请使用电子邮件服务管理电子邮件通知。电子邮件服务是一项系统服务，可在 Informatica Administrator 中配置。

## Informatica Analyst 结果卡进程

可以在 Developer tool 和 Analyst 工具中创建和编辑结果卡。可以在 Analyst 工具中运行结果卡。可以针对数据对象中的当前数据或暂存在剖析仓库中的数据运行结果卡。

可以在**结果卡**工作区中查看结果卡。在运行结果卡后，可以在**结果卡**面板上查看得分。可以选择数据对象，然后从结果卡中的某一得分导航到该数据对象。Analyst 工具将在另一个选项卡中打开该数据对象。

在使用结果卡时，可以执行以下任务：

1. 在 Developer tool 或 Analyst 工具中创建结果卡，然后从配置文件中添加列。
2. 在 Analyst 工具中打开结果卡。
3. 在运行配置文件后，将配置文件的列作为度量添加到结果卡中。
4. 或者，基于源数据创建结果卡筛选器。
5. 或者，为每个度量配置无效数据的成本。
6. 运行结果卡以生成各列的得分。
7. 查看结果卡，以查看记录中每一列的得分。
8. 对列进行向下钻取以获得得分。
9. 编辑结果卡。
10. 为结果卡中的每个度量设置阈值。
11. 创建组，以添加或移动结果卡中的相关度量。
12. 根据需要编辑或删除组。
13. 查看每个得分的得分趋势图表，以监视得分如何随着时间推移而变化。
14. 或者，查看每个度量的成本趋势图表，以监视数据质量的值。
15. 查看每个度量或度量组的结果卡沿革。
16. 查看您拥有读取权限的结果卡的合并信息。

## 在 Informatica Analyst 中创建结果卡

创建结果卡，然后将配置文件中的列添加到该结果卡中。必须首先运行配置文件，然后才能将列添加到结果卡中。

1. 在**库**工作区中，选择包含配置文件的项目或文件夹。
2. 单击配置文件将其打开。

配置文件结果将显示在**发现**工作区的摘要视图中。

3. 单击**操作 > 添加到结果卡**。

此时将显示**添加到结果卡**向导。

4. 在**添加到结果卡**屏幕中，可以选择创建新结果卡，也可以编辑现有结果卡以向预定义结果卡添加列。默认情况下将选择**新建结果卡**选项。单击**下一步**。
5. 在**第 2 步(共 8 步)**屏幕中，输入结果卡的名称。还可以选择输入结果卡的说明。选择用于保存结果卡的项目和文件夹。单击**下一步**。

默认情况下，结果卡向导将选择配置文件中定义的列和规则。无法添加未包括在配置文件中的列。
6. 在**第 3 步(共 8 步)**屏幕中，选择要作为度量添加到结果卡的列和规则。或者，单击左侧列表头中的复选框，以选择所有列。或者，选择**列名称**，以对列名称进行排序。单击**下一步**。
7. 在**第 4 步(共 8 步)**屏幕中，可以将筛选器添加到度量。

可以将为配置文件创建的筛选器应用于度量，或创建新筛选器。在**度量筛选器**窗格中选择一个度量，然后单击**管理筛选器**图标以打开**编辑筛选器: 列名称**对话框。在**编辑筛选器: 列名称**对话框中，可以选择执行以下任务之一：

  - 选择为配置文件创建的筛选器。单击**下一步**。
  - 选择现有筛选器。单击编辑图标，以在**编辑筛选器**对话框中编辑筛选器。单击**下一步**。
  - 单击加号 (+) 图标，以在**新建筛选器**对话框中创建筛选器。单击**下一步**。

或者，您可以选择将选定的筛选器应用于结果卡中的所有度量。

筛选器将显示在**度量筛选器**窗格中。
8. 在**第 4 步(共 8 步)**屏幕中，单击**下一步**。
9. 在**第 5 步(共 8 步)**屏幕中，选择**度量**窗格中的每个度量以执行下列任务：
  - 配置有效值。在**正在使用的得分: 值**窗格中，选择**可用值**窗格中的一个或多个值，然后单击向右箭头按钮，将这些值移至**有效值**窗格中。此时将在**可用值**窗格顶部显示某一度量的有效值的总数。
  - 配置度量阈值。在**度量阈值**窗格中，为**正常**、**可接受**和**不可接受**得分设置阈值。
  - 配置无效数据的成本。要为度量的成本分配常量值，请选择**固定成本**。要将数字列作为可变成本附加到度量，请选择**可变成本**，然后单击**选择列**，以选择某一数字列。或者，单击**更改成本单位**，以更改成本的单位。如果不希望为度量配置无效数据的成本，请选择**无**。
10. 单击**下一步**。
11. 在**第 6 步(共 8 步)**屏幕中，可以选择要将度量添加到的度量组，或创建新度量组。要创建新度量组，请单击组图标。单击**下一步**。
12. 在**第 7 步(共 8 步)**屏幕中，指定组中度量的权重以及组的阈值。
13. 在**第 8 步(共 8 步)**屏幕中，选择**本地**或**Hadoop**运行时环境选项来运行结果卡。如果选择 Hadoop 选项，请单击**浏览**选择 Hadoop 连接，以在 Blaze 引擎上运行配置文件。
14. 单击**保存**以保存结果卡，或者单击**保存并运行**以保存并运行结果卡。

结果卡将显示在**结果卡**工作区中。

## 向现有结果卡添加列

运行配置文件后，可向现有结果卡添加列。

1. 单击某一配置文件以打开它。

配置文件结果将显示在摘要视图中。
2. 选择一列。单击**操作 > 添加到结果卡**。

此时将显示**添加到结果卡**向导。

**注意:** 在将列添加到结果卡中之前, 使用以下规则和准则:

- 如果列名称与结果卡名称均匹配, 则无法将列添加到结果卡中。
  - 即使更改列名称, 也无法将同一列两次添加到一个结果卡中。
3. 选择**现有结果卡**, 以将列添加到预定义结果卡中。单击**下一步**。
  4. 在**第 2 步(共 7 步)** 屏幕中, 选择要添加列的结果卡。单击**下一步**。  
您可以查看与结果卡关联的现有度量和度量组。
  5. 在**第 3 步(共 7 步)** 屏幕中, 选择要作为度量添加到结果卡的列和规则。或者, 单击左侧列表头中的复选框, 以选择所有列。单击**列名称**, 以对列名称进行排序。单击**下一步**。
  6. 在**第 4 步(共 7 步)** 屏幕中, 可以为度量创建筛选器。也可以将为配置文件创建的筛选器应用到度量。
  7. 在**第 5 步(共 7 步)** 屏幕中, 可以执行以下任务:
    - 在**度量**窗格中, 选择各个度量并配置其他窗格中的度量值。
    - 在**正在使用的得分:值**窗格中, 选择**可用值**窗格中的多个值, 然后单击向右箭头按钮, 将这些值移至**有效值**窗格中。  
此时将在**可用值**窗格顶部显示某一度量的有效值的总数。
    - 在**度量阈值:** 窗格中, 可以为**正常、可接受和不可接受**得分设置阈值。
    - 在**无效数据的成本**中, 可以:
      - 选择每个度量, 然后为度量配置无效数据的成本。
      - 选择**固定成本**选项, 以为度量的成本分配常量值。可以单击**更改成本单位**, 以更改成本的单位。
      - 选择**可变成本**选项, 以将数字列作为可变成本附加到度量。可以单击**选择列**以选择数字列。
  8. 单击**下一步**。
  9. 在**第 6 步(共 7 步)** 屏幕中, 可以执行以下任务:
    - 选择要添加度量的度量组。
    - 在**默认 - 度量**窗格中, 可以双击默认度量权重 0, 以更改该值。
    - 在**度量阈值:**窗格中, 可以为**正常、可接受和不可接受**得分设置阈值。
  10. 单击**下一步**。
  11. 在**第 7 步(共 7 步)** 屏幕中, 选择运行时环境。
  12. 单击**保存**以保存结果卡, 或者单击**保存并运行**以保存并运行结果卡。

## 运行结果卡

可以运行结果卡, 以生成各列的得分。

1. 在**资产**面板中, 选择要运行的结果卡。
2. 单击该结果卡以打开它。  
结果卡显示在**结果卡**工作区中。
3. 单击**操作 > 运行结果卡**。
4. 从**度量**窗格中选择得分, 然后从**列**窗格中选择要对其进行向下钻取的列。
5. 在**向下钻取**选项中, 选择对实时数据或暂存数据进行向下钻取。  
要获得最佳性能, 请对实时数据进行向下钻取。

6. 单击**运行**。

## 查看结果卡

运行结果卡可以查看每个度量的得分。结果卡将以百分比和条形图的形式显示得分。可以查看有效数据或无效数据。还可以查看结果卡信息，如度量权重、度量组得分、得分趋势和数据对象的名称。

1. 运行结果卡以查看得分。
2. 选择包含要查看的得分的度量。
3. 单击**操作 > 向下钻取**，以查看相应列的有效数据行或无效数据行。  
默认情况下，Analyst 工具将在**向下钻取**部分中显示无效数据行。

## 编辑结果卡

可以在结果卡中编辑度量的有效值。必须运行结果卡，然后才能对其进行编辑。

1. 在**库**工作区中，在**资产**窗格中单击要编辑的结果卡。  
结果卡将显示在**结果卡**工作区中。
2. 如果启用了版本控制系统，则单击**操作 > 签出**。
3. 单击**操作 > 编辑 > 常规**。  
此时将显示**编辑结果卡**对话框。
4. 在**常规**选项卡中，您可以根据需要，编辑结果卡的名称和说明。
5. 单击**度量**选项卡。
6. 选择**度量**窗格中的一个得分，然后在**正在使用的得分: 值**窗格中，配置所有值的列表中的有效值。
7. 在**度量阈值**窗格中，根据需要更改得分阈值。
8. 检查每个度量的无效数据的成本，然后根据需要更改。
9. 单击**结果卡筛选器**选项卡。
10. 您可以添加、编辑或删除筛选器。
11. 单击**度量组**选项卡。
12. 您可以创建、编辑或删除度量组。  
还可以在**度量组**选项卡上编辑度量权重和度量阈值。
13. 单击**通知**选项卡。
14. 根据需要更改结果卡通知设置。  
可以为度量和度量组设置全局和自定义设置。
15. 单击**运行时环境**选项卡。  
您可以选择**本地**或 **Hadoop** 作为运行时环境。
16. 单击**保存**将更改保存到结果卡中；或者单击**保存并运行**保存更改并运行结果卡。
17. 单击**签入**。

# 度量

度量是数据源的某一列，或者是构成结果卡的某一规则的输出。在创建结果卡时，可为每个度量分配一个权重。可以创建度量组，以将结果卡中的相关度量分类到一个集合中。

## 度量权重

在创建结果卡时，可为每个度量分配一个权重。权重的默认值为 0。

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。

例如，为度量 M1 分配权重 W1，然后为度量 M2 分配权重 W2。Analyst 工具将使用下面的公式来计算加权平均值：

$$(M1 \times W1 + M2 \times W2) / (W1 + W2)$$

## 数据质量的值

源数据中数据质量的度量是管理组织中的数据资产的关键信息。无效数据的成本以度量的方式在结果卡中表示，可以帮助组织在监视源数据的数据质量过程中派生相应的值。作为数据分析人员，您可能想将某个值（如货币单位或任何自定义单位）与度量和度量组关联起来。随后，可以运行结果卡，以查看源数据中无效数据的总成本。

可以根据业务需要定义某一度量的成本单位。还可以在创建或编辑结果卡时，为每个度量配置可变或固定成本。

### 固定成本

固定成本是可以分配给结果卡中某一度量的常量值。既可选择预定义成本单位，也可以创建符合业务需要的自定义成本单位。

### 可变成本

可变成本是根据数据源中某一数字列中的值分配给某一度量的值。数据集成服务将根据该数字列或分配给成本的虚拟列来计算度量的可变成本。

#### 示例

作为抵押贷款管理人员，您需要为客户提供缴款书，以便客户能够提交抵押贷款付款。您可以使用结果卡来度量客户地址的准确性，以确保交付缴款书。您可能想为“地址准确性”度量的“每月付款金额”列设置可变成本。运行结果卡，以计算如果客户没有按时支付每月应付金额，抵押贷款组织亏损的总成本。

## 定义阈值

可以为结果卡中的每个得分设置阈值。阈值以百分比的形式指定记录中列可接受的错误数据的范围。可以为正常、可接受或不可接受的数据范围设置阈值。可以在将列添加到结果卡或编辑结果卡时，为每个列定义阈值。

为结果卡中的列定义阈值之前，需要完成以下先决条件任务之一：

- 打开配置文件，然后在**添加到结果卡**对话框中将该配置文件中的列添加到结果卡中。
  - 或者，在**库工作区**中单击结果卡，选择**操作 > 编辑**，在**编辑结果卡**对话框中编辑结果卡。
1. 在**添加到结果卡**对话框或**编辑结果卡**对话框中，选择**度量**窗格中的每个度量。
  2. 在**度量阈值**窗格中，输入代表不可接受范围上界以及正常范围下界的阈值。  
可以设置最多两个小数位的阈值。
  3. 单击**下一步**或**保存**。

# 度量组

可以创建度量组，将结果卡中的相关得分分类到一个集合中。默认情况下，Analyst 工具会对默认度量组中的所有得分进行分类。

在创建度量组后，可以将得分移出默认度量组，移至其他度量组中。可以编辑度量组以更改其名称和说明，包括默认度量组。可以删除不再使用的度量组。无法删除默认度量组。

## 创建度量组

可以创建度量组，以将结果卡中的相关得分添加到该组中。

1. 在**库**工作区中，在**资产**窗格中单击要编辑的结果卡。  
结果卡将显示在**结果卡**工作区中。
2. 单击**操作 > 编辑**。  
此时将显示**编辑结果卡**窗口。
3. 单击**度量组**选项卡。  
**度量组**面板中将显示默认组，而**度量**面板中则显示该默认组中的得分。
4. 单击**新建组**图标以创建度量组。  
此时将显示**度量组**对话框。
5. 输入名称和可选说明。
6. 单击**确定**。
7. 单击**保存**以将更改保存到结果卡中。

## 将得分移至度量组

在创建度量组后，可以将相关得分移至该度量组。

1. 在**库**工作区中，在**资产**窗格中单击要编辑的结果卡。  
结果卡将显示在**结果卡**工作区中。
2. 单击**操作 > 编辑**。  
此时将显示**编辑结果卡**窗口。
3. 单击**度量组**选项卡。  
**度量组**面板中将显示默认组，而**度量**面板中则显示该默认组中的得分。
4. 从**度量**面板中选择度量，然后单击**移动度量**图标。  
此时将显示**移动度量**对话框。  
**注意:** 要选择多个得分，请按住 Shift 键。
5. 选择要向其中移动得分的度量组。
6. 单击**确定**。

## 编辑度量组

可以编辑度量组，以更改名称和说明。可以更改默认度量组的名称。

1. 在**库**工作区中，在**资产**窗格中单击要编辑的结果卡。  
结果卡将显示在**结果卡**工作区中。
2. 单击**操作 > 编辑**。  
此时将显示**编辑结果卡**窗口。
3. 单击**度量组**选项卡。  
**度量组**面板中将显示默认度量组，而**度量**面板中将显示该默认度量组中的度量。
4. 在**度量组**面板上，单击**编辑组**图标。  
此时将显示**编辑**对话框。
5. 输入名称和可选说明。
6. 单击**确定**。

## 删除度量组

可以删除不再有效的度量组。在删除度量组时，可以选择将该度量组中的得分移至默认度量组。无法删除默认度量组。

1. 在**库**工作区中，在**资产**窗格中单击要编辑的结果卡。  
结果卡将显示在**结果卡**工作区中。
2. 单击**操作 > 编辑**。  
此时将显示**编辑结果卡**窗口。
3. 单击**度量组**选项卡。  
**度量组**面板中将显示默认度量组，而**度量**面板中将显示该默认度量组中的度量。
4. 在**度量组**面板中选择度量组，然后单击**删除组**图标。  
此时将显示**删除组**对话框。
5. 在删除度量组之前，请选择相应选项以删除度量组中的度量，或者选择相应选项以将度量移至默认度量组。
6. 单击**确定**。

## 对列进行向下钻取

在某一得分的列上进行向下钻取，以选择在查看有效数据行或无效数据行时显示的列。选择对其进行向下钻取的列将显示在**向下钻取**面板中。

1. 运行结果卡以查看得分。
2. 选择包含您要查看的得分的列。
3. 单击**操作 > 向下钻取**，以查看相应列的有效数据行或无效数据行。
4. 单击**操作 > 向下钻取列**。  
相应的列将显示在选定得分对应的**向下钻取**面板中。默认情况下，Analyst 工具将显示相应列的有效数据行。或者，单击**无效**以查看无效数据行。



# 趋势图表

可以使用趋势图表来监视度量得分和度量中无效数据的成本在一段时间内如何变化。

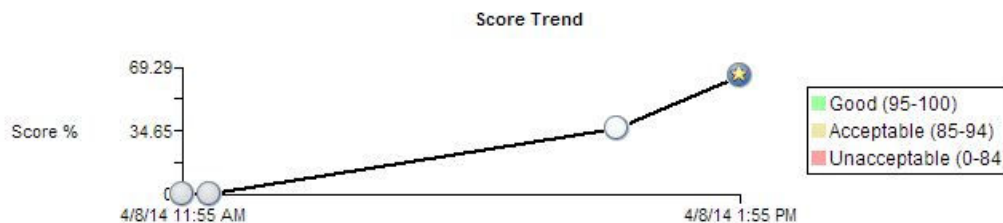
趋势图表包含得分图形和成本图形，在纵轴上绘制了得分值或成本值，而在横轴上则绘制了所有结果卡运行。默认情况下，趋势图表将显示来自最近 10 次结果卡运行的数据。可以在趋势图表中查看相应度量的总行数和无效行数。趋势图表还会显示得分和成本趋势是保持不变还是根据上次结果卡运行上移或下移。

Analyst 工具会将历史结果卡运行数据用于每个日期和最新有效得分值，以计算得分。Analyst 工具将使用该图表中的最新阈值设置来描绘得分点的颜色。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。在导出结果卡时，Analyst 工具会将趋势图表信息（其中包括得分和成本信息）包括在导出的文件中。

## 得分趋势图表

得分趋势图表以图形方式表示了度量得分如何随着多次配置文件运行而变化。得分趋势图表在纵轴上绘制度量得分值，而在横轴上绘制所有结果卡运行。

下图显示了得分趋势图表的一个示例：



### 示例

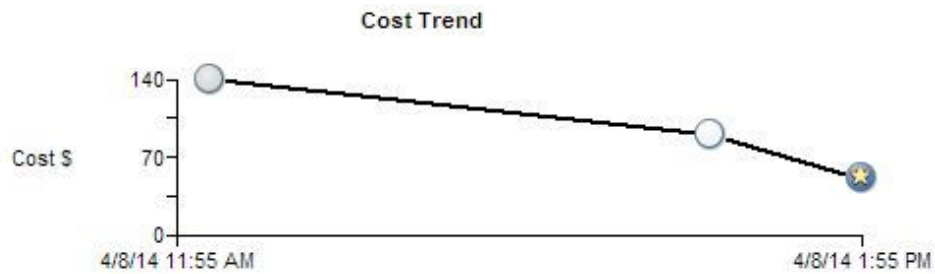
作为数据分析人员，您可以监视数据质量，以分析映射和其他进程更改是否使得数据质量得分提高。在度量数据质量中的更改之后，即可报告组织的数据质量变化，以供分析和使用。例如，在多次结果卡运行结束时，“社会保障号码”列中有效值的百分比可能已从 84 移至 90。您可以使用可视化图表的形式报告数据质量中的这一变化，以供快速分析。

## 成本趋势图表

成本趋势图表以图形方式表示了度量中无效数据的成本如何随着多次配置文件运行而变化。成本趋势图表可以度量数据质量在组织中的影响。成本趋势图表在纵轴上绘制成本值，而在横轴上绘制所有结果卡运行。还可以查看成本趋势图表下某一网格中的度量的无效数据以及无效值的总成本。

成本趋势图表可以帮助您跟踪无效数据对高价值记录的影响。有时，在使用固定成本计算无效数据时，可能会错过了解无效数据对高价值记录的影响的机会。发生这种问题的原因在于趋势图表可能会显示随着多次结果卡的运行，得分的提高以及总体成本的下降。不过，结果卡中显示的数据质量问题较少的情况可能存在于高价值记录中。

下图显示了成本趋势图表的一个示例：



#### 示例

在某一金融机构中，您有几位高余额客户，在银行内存有大量存款和投资（如 1,000 万美元）。您还拥有大量低余额客户。得分趋势图表可以显示在一段时间内得分的提高。不过，一些高余额客户帐户的地址或性别不正确，可能会影响与该组织最宝贵的客户之间的关系。您可以将“帐户余额”列设置为可变成本列，用于计算无效数据。如果由于该列而使无效数据的成本较高，则可认为总值处于风险之中，并应立即采取更正操作。

## 查看趋势图表

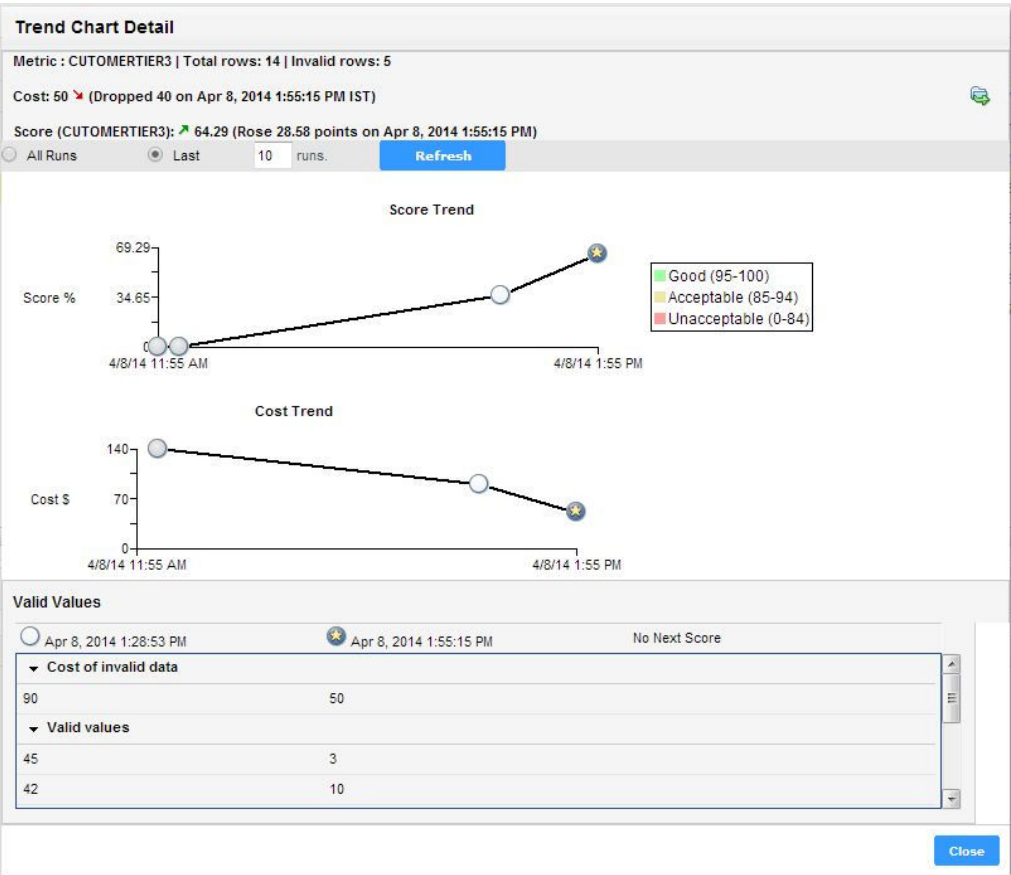
可以查看每个度量的趋势图表，以监视得分或无效数据的成本如何随着时间推移而变化。

1. 在**库**工作区中，选择包含结果卡的项目或文件夹。
2. 单击该结果卡以打开它。  
结果卡显示在**结果卡**工作区中。
3. 在**结果卡**视图中，选择某一度量。

4. 单击**操作 > 显示趋势图表**。

此时将显示**趋势图表详细信息**对话框。

下图显示了**趋势图表详细信息**对话框：



可以查看随时间改变的得分和成本值。在对话框顶部，可以查看总行数和无效行数。Analyst 工具使用每个日期的历史结果卡运行数据和最新有效得分值计算得分。在得分趋势图表和成本趋势图表下，可以查看度量的有效值以及无效数据的成本。

## 导出趋势图表

可将得分趋势图表和成本趋势图表导出到“.xlsx”文件，以便在一个文件中查看数据。

1. 打开结果卡。
2. 选择一个度量，然后单击**操作 > 显示趋势图表**。  
此时将显示**趋势图表详细信息**对话框。
3. 单击**导出数据**图标。  
此时会显示**将数据导出到文件**对话框。
4. 输入文件名称。或者，使用默认文件名。  
默认文件格式为 Microsoft Excel。
5. 选择文件的代码页。
6. 单击**确定**。

# Informatica Analyst 结果卡导出文件

可将结果卡结果导出到 Microsoft Excel 文件。Analyst 工具将以“xlsx”格式保存该文件。

导出结果卡时，可以选择将结果卡摘要、趋势图表、无效行和结果卡属性导出到 Microsoft Excel 文件。或者，也可以仅将结果卡摘要、趋势图表和结果卡属性导出到 Microsoft Excel 文件。

## Microsoft Excel 格式的结果卡导出结果

在导出结果卡结果时，Analyst 工具会将信息保存至 Microsoft Excel 文件内的多个工作表中。结果卡摘要、趋势图表、无效行和结果卡属性在文件中显示为工作表。Analyst 工具将以“xlsx”格式保存该文件。

下表描述了导出文件中的每个工作表上显示的信息：

选项卡	说明
结果卡摘要	导出的结果卡结果的摘要信息。这些信息包括结果卡名称、每列的总行数、无效的行数、得分和度量权重。
趋势图表	得分的趋势图表。
无效行数	每列的无效行的详细信息。Analyst 工具最多可以将 100 行导出到工作表。在将数据导出到文件对话框中选择 <b>数据 &gt; 全部</b> 选项时，将显示“无效行”工作表。
属性	结果卡属性，如名称、类型、说明和位置。

## 从 Informatica Analyst 导出结果卡结果

可将结果卡结果导出到“.xlsx”文件，以便在一个文件中查看数据。

1. 打开结果卡。
2. 单击**操作 > 导出数据**。  
此时将显示**将数据导出到文件**对话框。
3. 输入文件名称。或者，使用默认文件名。  
默认文件格式为 Microsoft Excel。
4. 选择文件的代码页。
5. 单击**确定**。

## 结果卡通知

配置结果卡通知设置，以使 Analyst 工具能在特定度量得分、度量组得分或度量成本跨越阈值时发送电子邮件。度量得分或度量组得分可能会跨越阈值，也可能会保持在特定的得分范围内，如“不可接受”、“可接受”和“正常”。度量成本值可以超过您设置的成本阈值上限和下限。

可以为单个度量得分、度量组和度量成本配置电子邮件通知。如果将全局设置用于得分，则 Analyst 工具将在特定度量得分从得分范围“正常”跨越阈值到达“可接受”以及从“可接受”跨越阈值到达“错误”时，发送通知电子邮件。如果得分在结果卡连续运行期间保持在“不可接受”得分范围内，也会收到通知电子邮件。如果将全

局设置用于度量成本，则 Analyst 工具将在选定度量中无效数据的成本跨越阈值上限和下限时，发送通知电子邮件。

可以自定义通知设置，以使结果卡用户在得分从“不可接受”移至“可接受”以及从“可接受”移至“正常”得分范围时，能够收到电子邮件通知。可以选择如果度量得分或度量成本在每次结果卡运行时都保持在特定范围内，是否发送电子邮件通知。可以在通知设置中根据可以设置成本阈值的度量来查看每个度量的无效数据的当前成本。

配置结果卡以发送电子邮件通知之前，管理员必须在 Administrator 工具中配置电子邮件服务。

## 通知电子邮件模板

可以设置 Analyst 工具作为结果卡通知的组成部分发送给收件人的电子邮件的邮件文本和结构。电子邮件模板包含可选的介绍性文本部分、只读邮件正文部分，以及可选的结束文本部分。

下表介绍了电子邮件模板中的标记：

标记	说明
ScorecardName	结果卡的名称。
ObjectURL	指向结果卡的超级链接。需要提供用户名和密码。
MetricGroupName	度量所属的度量组的名称。
CurrentWeightedAverage	当前结果卡运行中度量组的加权平均值。
CurrentRange	当前结果卡运行中度量组的得分范围，如“不可接受”、“可接受”、和“正常”。
PreviousWeightedAverage	上一次结果卡运行中度量组的加权平均值。
PreviousRange	上一次结果卡运行中度量组的得分范围，如“不可接受”、“可接受”、和“正常”。
MetricName	度量的名称。
MetricGroupName	度量组的名称。
CurrentScore	最新结果卡运行的得分。
CurrentRange	在最新结果卡运行中，当前得分所处的得分范围。
PreviousScore	上一次结果卡运行的得分。
PreviousRange	上一次结果卡运行的得分范围。
CurrentCost	在最新结果卡运行中，度量中无效数据的成本。
PreviousCost	在上一次结果卡运行中，度量中无效数据的成本。
ColumnName	为度量分配的源列的名称。
ColumnType	源列的类型。
RuleName	规则名称。

标记	说明
RuleType	规则的类型。
DataObjectName	源数据对象的名称。

## 设置结果卡通知

既可以在度量级别也可以在度量组级别设置结果卡通知。全局通知设置适用于不包含单个通知设置的度量和度量组。

1. 在 Analyst 工具中运行结果卡。
2. 单击**操作 > 编辑**。
3. 单击**通知**选项卡。
4. 选择**启用通知**，以开始配置结果卡通知。
5. 选择某一度量或度量组。
6. 单击**通知**复选框，以便为该度量或度量组启用全局设置。
7. 选择**使用自定义设置**以更改该度量或度量组的设置。

可以选择在得分处于**不可接受**、**可接受**和**正常**范围内以及跨越阈值时发送通知电子邮件。还可以在度量成本跨越阈值上限或下限时发送通知电子邮件。

8. 要编辑结果卡通知的全局设置，请单击**编辑全局设置**图标。  
在可以编辑设置（包括电子邮件模板）的位置，将显示**编辑全局设置**对话框。

## 配置结果卡通知的全局设置

如果选择全局结果卡通知设置，则 Analyst 工具会在得分处于**不可接受**范围内时向目标用户发送电子邮件。还可以将通知设置配置为在度量得分或度量成本超过阈值时发送电子邮件。可以为结果卡配置电子邮件模板，包括电子邮件地址和邮件文本。

1. 在 Analyst 工具中运行结果卡。
2. 单击**操作 > 编辑 > 通知**，以打开**编辑结果卡**对话框。
3. 选择**启用通知**，以开始配置结果卡通知。
4. 单击**编辑全局设置**图标。  
在可以编辑设置（包括电子邮件模板）的位置，将显示**编辑全局设置**对话框。
5. 使用**得分**和**得分移动**复选框选择想在何时为度量得分发送电子邮件通知。
6. 使用**成本趋势**复选框选择想在何时为度量成本发送电子邮件通知。
7. 在**电子邮件收件人**字段中，输入收件人的电子邮件 ID。使用分号分隔多个电子邮件 ID。  
默认发件人电子邮件 ID 是在域 SMTP 属性中配置的**发件人电子邮件地址**。
8. 输入电子邮件主题的文本。
9. 在**正文**字段中，添加电子邮件的介绍性文本和结束文本。
10. 要应用全局设置，请选择**将设置应用于所有度量和度量组**。
11. 单击**确定**。

# 结果卡沿袭

结果卡沿袭显示数据的来源、介绍路径，并显示数据如何针对度量或度量组流动。可以使用结果卡沿袭来分析度量或度量组中出现不可接受的得分差异的根本原因。可以在 Analyst 工具中查看结果卡沿袭。

要查看结果卡沿袭，请完成以下任务：

1. 在 Informatica Administrator 中，将 Metadata Manager 服务与分析服务关联起来。
2. 选择某一项目，然后使用 Developer 工具中的“导出资源文件供 Metadata Manager 使用”选项或 infacmd 工具的 exportResources 命令，将该项目中的结果卡对象导出到 XML 文件。
3. 在 Metadata Manager 中，使用导出的 XML 文件创建资源，然后加载该资源。

**注意：**在 Metadata Manager 中创建并加载的资源文件的名称必须使用以下命名约定：<MRS 名称>\_<项目名称>。有关如何创建并加载资源文件的更多信息，请参阅《Informatica PowerCenter Metadata Manager 用户指南》。

4. 在 Analyst 工具中，打开结果卡，然后选择度量或度量组。
5. 查看结果卡沿袭。

## 在 Informatica Analyst 中查看结果卡沿袭

可以查看度量或度量组的结果卡沿袭图表。首先必须在 Metadata Manager 中加载结果卡沿袭和元数据，然后才能在 Analyst 工具中查看结果卡沿袭图表。

1. 在库工作区中，在资产窗格中单击要查看的结果卡。  
结果卡将显示在结果卡工作区中。
2. 在结果卡视图中，选择某一度量或度量组。
3. 右键单击然后选择显示沿袭。

此时将在新窗口中显示结果卡沿袭图表。

**重要说明：**如果没有使用导出的结果卡对象 XML 文件在 Metadata Manager 中创建并加载资源，则会看到一条错误消息，指出目录中的资源不可用。有关为结果卡沿袭导出 XML 文件的更多信息，请参阅[“为结果卡沿袭导出资源文件”](#)页面上 94。

# 第 III 部分：使用 Informatica Developer 进行剖析

本部分包含以下章节：

- [数据对象配置文件, 73](#)
- [基于半结构化数据源的列配置文件, 78](#)
- [Informatica Developer 中的规则, 84](#)
- [Mapplet 和映射剖析, 86](#)
- [Informatica Developer 中的列配置文件结果, 88](#)
- [Informatica Developer 中的结果卡, 93](#)



## 第 9 章

# 数据对象配置文件

本章包括以下主题：

- [Informatica Developer 中的列配置文件, 73](#)
- [在 Informatica Developer 中创建单个数据对象配置文件, 74](#)
- [在 Informatica Developer 中创建多个数据对象配置文件, 75](#)
- [在 Informatica Developer 中同步平面文件数据对象, 75](#)
- [在 Informatica Developer 中同步关系数据对象, 77](#)

## Informatica Developer 中的列配置文件

可以使用列配置文件来分析数据源中各列的特性，如值百分比和值模式。可以添加筛选器以确定配置文件在运行时读取的行。配置文件不处理不满足筛选条件的行。

可以发现关于对其运行配置文件的列的以下类型的信息：

- 某个值在列中出现的次数。
- 某一列中每个值出现的频率，用百分比或行数表示。
- 某一列中各个值的字符模式。
- 统计信息，如某一列中各个值的最大长度和最小长度，以及第一个值和最后一个值。
- 推理的数据类型、频率、数据域发现的遵从性条件以及数据类型推理状态。

可以为模型存储库中的映射、Mapplet 或对象的数据对象定义列配置文件。存储库中的对象可以位于一个或多个数据对象配置文件中，也可以位于企业发现配置文件中。

您可以为列配置文件选择采样选项、向下钻取选项和运行时环境。可以向列配置文件添加规则和筛选器。

### 筛选选项

您可以添加高级筛选器或 SQL 筛选器来确定在您运行配置文件时列配置文件使用的行。配置文件不会处理不满足筛选条件的行。

# 采样选项

配置采样选项，以确定在剖析操作期间配置文件读取的行数。

下表介绍了采样选项：

属性	说明
所有行	选择数据对象中的所有行。
第一行	希望对其运行配置文件的行数。Developer tool 将从源的第一行开始选择行。最多可以选择 2,147,483,647 行。
以下项的随机采样	随机采样算法将以随机方式在数据对象中选择行来运行配置文件。最多可以选择 2,147,483,647 行。
随机采样(自动)	随机采样大小根据数据对象中的行数计算而得。
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	在下次运行配置文件时，从数据类型和数据域推理中排除已批准的数据类型或数据域。

选择对随机的行样本运行配置文件后，随机采样算法将以随机方式在数据对象中选择行来运行配置文件。为列配置文件选择随机采样选项时，Developer tool 将对暂存的数据执行向下钻取。这会影响向下钻取性能。为数据域发现配置文件选择随机采样选项时，Developer tool 将对实时数据执行向下钻取。

# 在 Informatica Developer 中创建单个数据对象配置文件

您可以为数据对象中的一个或多个列创建单个数据对象配置文件，然后将该配置文件对象存储在模型存储库中。

1. 在**对象浏览器**视图中，选择要剖析的数据对象。
2. 单击**文件 > 新建 > 配置文件**以打开配置文件向导。
3. 选择**配置文件**，然后单击**下一步**。
4. 输入配置文件的名称并验证项目位置。如果需要，浏览到新位置。
5. 或者，输入配置文件的文本说明。
6. 验证所选数据对象的名称是否显示在**数据对象**部分。
7. 单击**下一步**。
8. 配置要执行的配置文件操作。可以配置以下操作：
  - 列剖析
  - 主键发现
  - 功能相关性发现
  - 数据域发现

**注意:** 要启用配置文件操作，请为该操作选择**已在“运行配置文件”操作中启用**。默认情况下将启用列剖析。
9. 查看配置文件的选项。

可以编辑所有配置文件类型的列选择。查看列配置文件的筛选器和采样选项。可以查看主键、功能相关性和数据域发现的推理选项。还可以查看数据域发现的数据域选择。

10. 查看向下钻取选项，并根据需要进行编辑。默认情况下，**启用行向下钻取**选项为选中状态。您可以编辑列配置文件的向下钻取选项。这些选项还可确定向下钻取操作是从数据源读取还是从暂存数据读取，以及配置文件是否存储上一次配置文件运行的结果数据。
11. 在**运行设置**部分，选择运行时环境。选择**本机**或**Hadoop**作为运行时环境。选择 Hadoop 选项后，可以选择 Hadoop 连接。
12. 单击**完成**。

## 在 Informatica Developer 中创建多个数据对象配置文件

对多个数据对象运行多个数据对象配置文件时，Developer tool 将使用默认的列剖析选项为一个或多个数据对象生成列配置文件。您也可以选择创建一个企业发现配置文件，以对多个数据对象运行一个配置文件。

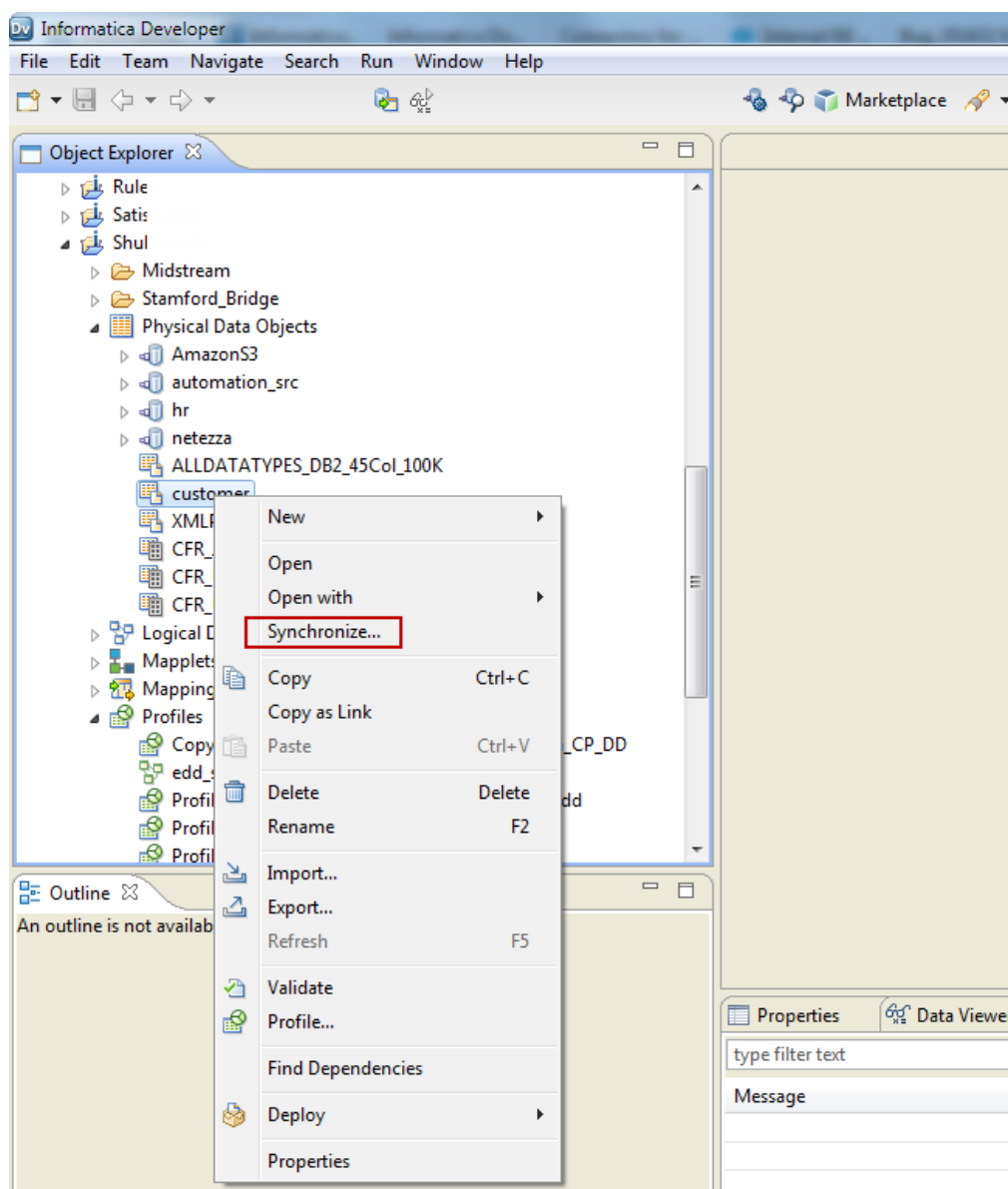
1. 在**对象浏览器**视图中，选择要剖析的数据对象。
2. 单击**文件 > 新建 > 配置文件**，以打开**新建配置文件**向导。
3. 在**新建**向导中，选择**多个配置文件**选项，然后单击**下一步**。
4. 在**多个配置文件**窗口中，选择要在其中创建配置文件的位置。可以在与其剖析对象相同的位置创建各个配置文件，也可以为配置文件指定一个通用位置。
5. 验证所选数据对象的名称是否显示在**数据对象**部分。  
或者，单击**添加**以添加其他数据对象。
6. 或者，指定要剖析的行数，然后选择是否在向导完成时运行配置文件。
7. 单击**下一步**。
8. 在**验证环境**部分中，选择**本地**。  
**注意：**仅选择本地选项以运行多个数据对象配置文件。在 Hadoop 运行时环境中，要在 Blaze 引擎上运行多个数据对象，可以选择企业发现配置文件。
9. 单击**完成**。
10. 或者，输入要添加到配置文件名称的前缀和后缀字符串。
11. 单击**确定**。

## 在 Informatica Developer 中同步平面文件数据对象

可以将外部平面文件数据源的更改与其在 Informatica Developer 中的数据对象同步。使用**同步平面文件**向导同步数据对象。

1. 在**对象浏览器**视图中，选择平面文件数据对象。
2. 右键单击，然后选择**同步**。

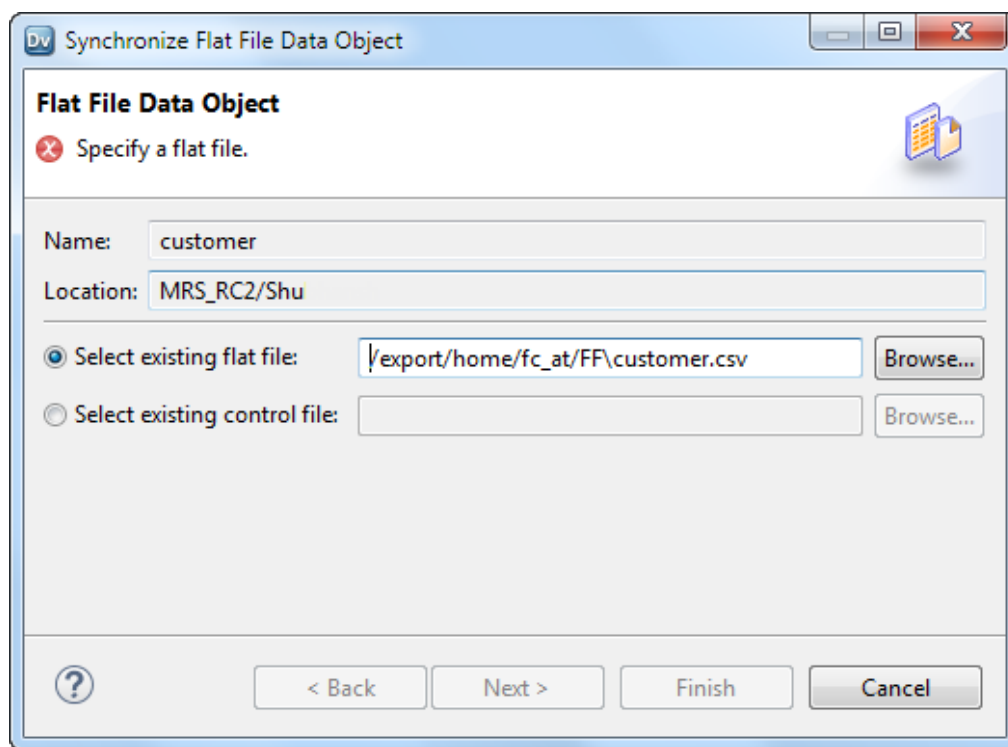
下图显示了数据对象的“同步”选项：



此时将显示同步平面文件数据对象向导。

3. 在同步平面文件数据对象向导中，验证选择现有平面文件字段中的平面文件路径。

下图显示了“同步平面文件数据对象”向导：



4. 单击**下一步**。
5. 或者，选择代码页、格式、带分隔符的格式属性以及列属性。
6. 单击**完成**，然后单击**确定**。

## 在 Informatica Developer 中同步关系数据对象

可以将关系数据源的外部数据源更改与其在 Informatica Developer 中的数据对象同步。外部数据源更改包括添加、更改和删除列以及规则更改。

1. 在**对象浏览器**视图中，选择关系数据对象。
2. 右键单击，然后选择**同步**。  
此时将显示一条消息，提示您确认操作。
3. 要完成同步进程，请单击**确定**。  
此时将显示同步过程状态消息。
4. 显示**同步完成**消息时，请单击**确定**。  
该消息将显示数据对象的元数据更改的摘要。

## 第 10 章

# 基于半结构化数据源的列配置文件

本章包括以下主题：

- [基于半结构化数据源的列配置文件概览, 78](#)
- [JSON 和 XML 数据对象, 78](#)
- [HDFS 中半结构化数据源的复杂文件数据对象, 79](#)
- [创建 HDFS 连接, 80](#)
- [从 HDFS 中的 JSON 或 XML 文件创建复杂文件数据对象, 80](#)
- [从 Avro 或 Parquet 数据源创建复杂文件数据对象, 81](#)
- [创建基于半结构化数据源的列配置文件, 82](#)

## 基于半结构化数据源的列配置文件概览

您可以基于 Avro、JSON、Parquet 和 XML 数据源创建数据对象，然后在数据对象上创建列配置文件。

Avro、JSON、Parquet 和 XML 格式均为半结构化数据源。要使用半结构化数据源创建列配置文件，您可以执行以下任务：

1. 基于半结构化数据源创建物理数据对象。
2. 在该物理数据对象上创建并运行列配置文件。

您可以为 JSON 或 XML 数据源创建平面文件数据对象。您可以为采用 Hadoop 分布式文件系统 (HDFS) 的 Avro、JSON、Parquet 和 XML 数据源创建复杂文件数据对象。

## JSON 和 XML 数据对象

您可以从 JSON 或 XML 数据源创建平面文件数据对象或复杂文件数据对象。您可以在数据对象上创建并运行列配置文件。

创建含有 JSON 或 XML 数据源路径的文本文件，然后将该文本文件用作数据源，以此来创建平面文件数据对象。您也可以将多个 JSON 或多个 XML 数据源的文件路径添加到文本文件中。

您可以使用复杂文件读取器，从 JSON 或 XML 数据源创建复杂文件数据对象。复杂文件读取器会为数据处理器转换提供输入，然后由数据处理器转换解析文件并将源数据转换为逗号分隔值平面记录。

**注意：**Developer tool 不支持使用 UTF-8 编码的 JSON 数据源。

## 从 JSON 或 XML 数据源创建数据对象

您可以从 JSON 或 XML 数据源创建平面文件数据对象或复杂文件数据对象。

1. 在 Developer tool 的**对象浏览器**视图中，选择要在其中创建数据对象和列配置文件的项目。
2. 单击**文件 > 新建 > 数据对象**。  
此时将显示**新建**对话框。
3. 您可以选择创建平面文件数据对象或复杂文件数据对象。
  - 要创建平面文件数据对象，请执行以下任务：
    1. 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。
    2. 选择**从现有平面文件创建**，然后单击**浏览**以选择文本文件。单击**下一步**。
    3. 验证代码页是否为 **MS Windows Latin 1 (ANSI)**，**Latin 1 的超集**，并验证格式是否设置为带分隔符。单击**下一步**。
    4. 验证分隔符是否设置为**逗号**。单击**完成**。
  - 要创建复杂文件数据对象，请执行以下任务：
    1. 选择**物理数据对象 > 复杂文件数据对象**，然后单击**下一步**。  
此时将显示**新建复杂文件数据对象**对话框。
    2. 输入数据对象的名称。选择**文件**作为访问类型。
    3. 单击**浏览**以选择 JSON 或 XML 文件。单击**完成**。  
Developer 服务器位于 Linux 中时，必须将数据源的文件路径更新为服务器中的位置。要更新文件路径，请选择复杂文件数据对象，单击**数据对象操作**选项卡中的**读取**，然后在**数据对象操作详细信息**窗格中的**高级**选项卡中添加文件路径。

数据对象将显示在项目文件夹中。

## HDFS 中半结构化数据源的复杂文件数据对象

您可以对使用 HDFS 的 Avro、JSON、Parquet 或 XML 文件创建并运行列配置文件。要读取 HDFS 格式的 JSON 或 XML 文件，您需要使用复杂文件读取器将 JSON 或 XML 输入传递给数据处理器进行转换。

## HDFS 中来自 JSON 或 XML 数据源的复杂文件数据对象

您可以从 JSON 或 XML 文件创建复杂文件数据对象您可以在数据对象上创建并运行列配置文件。

请先与 HDFS 建立连接，然后再为 HDFS 格式的 JSON 或 XML 文件创建数据对象。

从 HDFS 格式的 JSON 和 XML 文件创建数据对象时，您可以使用以下方法之一：

- 基于 JSON 或 XML 文件创建复杂文件数据对象。
- 基于含有多个 JSON 或多个 XML 文件的文件夹创建复杂文件数据对象。

创建数据对象后，您可以在数据对象上创建并运行列配置文件。

## HDFS 中来自 Avro 或 Parquet 数据源的复杂文件数据对象

您可以从 HDFS 格式的 Avro 或 Parquet 数据源创建复杂文件数据对象。您可以使用该数据对象创建并运行列配置文件。

可以从 Avro 或 Parquet 文件创建复杂文件数据对象，也可以基于包含多个 Avro 或多个 Parquet 文件的文件夹创建。可以从 Avro 和 Parquet 数据源创建访问类型为文件或连接且资源格式为二进制、Avro 或 Parquet 的复杂文件数据对象。您需要先创建 HDFS 连接，然后再从 Avro 和 Parquet 数据源创建复杂文件数据对象。

**注意：**您只能为平面结构化 Avro 和 Parquet 数据源选择 **Avro** 或 **Parquet** 资源格式。

从 HDFS 格式的 Avro 和 Parquet 文件创建数据对象时，您可以选择以下选项之一：

- 将访问类型选择为文件，将资源格式选择为二进制。
- 将访问类型选择为文件，将资源格式选择为 Avro 或 Parquet。
- 将访问类型选择为连接，将资源格式选择为 Avro 或 Parquet。

## 创建 HDFS 连接

在 Informatica Developer 中配置 HDFS 连接，以便基于 HDFS 格式的 Avro、JSON、Parquet 和 XML 数据源创建列配置文件。创建 HDFS 连接后，您可以创建复杂文件数据对象。

1. 单击**窗口 > 首选项**。
2. 选择 **Informatica > 连接**。
3. 展开域。
4. 选择连接类型**文件系统 > Hadoop 文件系统**，然后单击**添加**。
5. 输入连接名称。
6. 或者，也可以输入连接说明。
7. 单击**下一步**。
8. 配置连接属性。
9. 单击**测试连接**以验证与 HDFS 的连接。
10. 单击**完成**。

## 从 HDFS 中的 JSON 或 XML 文件创建复杂文件数据对象

可以从使用 HDFS 的 JSON 或 XML 源文件创建复杂文件数据对象，然后基于该数据对象创建列配置文件。

1. 在 Developer tool 的**对象浏览器**视图中，选择要在其中创建物理数据对象和列配置文件的项目。
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 > 复杂文件数据对象**，然后单击**下一步**。  
此时将显示**新建复杂文件数据对象**对话框。
4. 输入数据对象的名称。选择**连接**作为访问类型。



5. 可以从 JSON 或 XML 文件创建数据对象，也可以基于包含多个 JSON 文件或多个 XML 文件的文件夹创建。
  - 要从 JSON 或 XML 文件创建复杂文件数据对象，请执行以下步骤：
    1. 单击**浏览**以选择连接。
    2. 在**添加资源**对话框中，单击**添加**以选择 JSON 或 XML 文件。
    3. 单击**完成**。  
数据对象将显示在项目文件夹中。
  - 要基于包含多个 JSON 文件或多个 XML 文件的文件夹创建复杂文件数据对象，请执行以下步骤：
    1. 单击**浏览**以选择连接。
    2. 在**添加资源**对话框中，单击**添加**以选择文件夹中的 JSON 或 XML 文件。
    3. 单击**完成**。  
数据对象将显示在项目文件夹中。
    4. 选择项目文件夹中的数据对象，然后单击**高级 > 运行时: 读取 > 源文件目录**。
    5. 删除文件路径中的文件名，而保留其中的文件夹名称。

## 从 Avro 或 Parquet 数据源创建复杂文件数据对象

可以从 Avro 和 Parquet 数据源创建访问类型为**文件**或**连接**的复杂文件数据对象。可以基于该数据对象创建列配置文件。

1. 在**对象浏览器**视图中，选择一个项目。
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 > 复杂文件数据对象**，然后单击**下一步**。  
此时将显示**新建复杂文件数据对象**对话框。
4. 输入数据对象的名称。
5. 您可以选择访问类型为**连接**或**文件**。
  - 如果选择**连接**作为访问类型，请执行以下步骤：
    1. 单击**浏览**选择 HDFS 连接。
    2. 在**选择连接**对话框中，选择数据源，然后单击**确定**。
    3. 在**新建复杂文件数据对象**对话框中，单击**完成**。  
数据对象将显示在项目文件夹中。
  - 如果选择**文件**作为访问类型并选择**二进制**作为资源格式，请执行以下步骤：
    1. 单击**浏览**选择本地计算机上的 Avro 或 Parquet 文件。
    2. 在**新建复杂文件数据对象**对话框中，单击**完成**。  
数据对象将显示在项目文件夹中。
    3. 选择项目文件夹中的数据对象，然后单击**数据对象操作**视图。
    4. 在**数据对象操作**视图中，单击**读取 > 高级**选项卡。
    5. 在**高级**选项卡中，在**文件路径**字段中输入 Linux 或 Windows 计算机上数据源的文件路径。
    6. 将文件格式输入为**自定义输入**。

7. 在 Avro 数据源的**输入格式**字段中输入 **com.informatica.avro.AvroToXML**，以及在 Parquet 数据源的**输入格式**字段中输入 **com.informatica.parquet.ParquetToXML**。添加输入格式时，数据处理器转换会在运行时将 Avro 或 Parquet 格式的数据源处理并转换成 XML 格式数据源。
- 如果选择**文件**作为访问类型并选择 **Avro** 或 **Parquet** 作为资源格式，请执行以下步骤：
  1. 单击**浏览**选择本地计算机上的 Avro 或 Parquet 文件。
  2. 在**新建复杂文件数据对象**对话框中，单击**完成**。  
数据对象将显示在项目文件夹中。
  3. 创建数据对象后，导航到**数据对象操作 > 读取 > 高级选项卡**，然后确认**文件路径**字段中的文件路径是否与 Linux 或 Windows 计算机中的数据源对应。

**注意:** 您只能为平面结构化 Avro 和 Parquet 数据源选择 **Avro** 或 **Parquet** 资源格式。

您可选择具有多个 Avro 或多个 Parquet 文件的文件夹来创建数据对象。创建数据对象后，导航到**数据对象操作 > 读取 > 高级选项卡**，然后验证**文件路径**字段中的文件路径是否指向 Linux 或 Windows 计算机中的数据源文件夹。

## 创建基于半结构化数据源的列配置文件

基于 Avro、JSON、Parquet 或 XML 数据源创建平面文件数据对象或复杂文件数据对象后，您可以基于该数据对象创建和运行列配置文件。

1. 在**对象浏览器**视图中，为 Avro、JSON、Parquet 或 XML 文件选择数据对象。
2. 单击**文件 > 新建 > 配置文件**。  
此时将显示**新建**对话框。
3. 选择**配置文件**。单击**下一步**。  
此时将显示**新建配置文件**对话框。
4. 在**新建配置文件**对话框中，为配置文件添加名称和可选说明。
5. 选择**处理扩展文件格式**选项。单击**下一步**。

下图显示的是选择了**处理扩展文件格式**选项的**新建配置文件**向导：

1. 处理扩展文件格式。选择此选项可处理半结构化数据源。

**注意:** 为“资源格式”选择 **Avro** 或 **Parquet** 时，Avro 和 Parquet 数据源不会显示**处理扩展文件格式**选项。

6. 在**单个数据对象配置文件**页面中，根据需要选择**列选择**和**数据域发现**下的列和选项。单击**完成**。

**注意:** 如果 Developer tool 安装在 Linux 计算机上，而且 JSON 或 XML 物理数据对象为含有文本文件的平面文件数据对象，请执行以下任务：

1. 在**概览**选项卡中，更新**精度**值，以便将字符数加入服务器中数据源的文件路径。
  2. 在平面文件数据对象上创建配置文件后，将数据源的文件路径更新为服务器中的相应位置。要更新文件路径，请单击**高级**选项卡中的**运行时: 读取 > 源文件目录**，然后添加文件路径。
7. 右键单击配置文件，然后选择**运行配置文件**。

此时将显示配置文件结果。

## 第 11 章

# Informatica Developer 中的规则

本章包括以下主题：

- [Informatica Developer 中的规则概览, 84](#)
- [在 Informatica Developer 中创建规则, 85](#)
- [在 Informatica Developer 中应用规则, 85](#)

## Informatica Developer 中的规则概览

规则是业务逻辑，用于定义在运行列配置文件时对源数据应用的条件。可以在配置文件中添加规则以验证数据。可以在列配置文件中使用的经验证为有效的规则、预定义规则或可重用规则的 Maplet。

可以通过以下方法在列配置文件中使用的规则：

- 在 Developer tool 中，创建 Maplet 并使其成为有效的规则。此规则在 Analyst 工具中显示为可重用规则。可以在 Analyst 工具和 Developer tool 中将规则应用到列配置文件。
- 可以在列配置文件中使用的预定义规则。Informatica 随 Developer tool 和 Analyst 工具一起提供了预定义规则。
- 在 Analyst 工具中，创建规则规范并生成 Maplet。可以在 Analyst 工具中将规则规范应用到列配置文件。在 Developer tool 中，使该 Maplet 成为有效的规则。该规则将显示为可重用规则，您可以在列配置文件中使用该规则。

**注意：**在 Developer tool 中，无法在列配置文件中添加、编辑或删除规则规范。

规则必须符合以下要求：

- 必须包含一个输入转换和一个输出转换。不能在规则中使用数据源。
- 可以包含表达式转换、查找转换和被动数据质量转换。不能包含其他任何类型的转换。例如，由于规则是主动转换，因而无法包含匹配转换。
- 它不指定输入组之间的基数。

# 在 Informatica Developer 中创建规则

需要将 Mapplet 作为规则进行验证才能在 Developer 工具中创建规则。

在 Developer 工具中创建 Mapplet。

1. 右键单击 Mapplet 编辑器。
2. 选择**验证为 > 规则**。

# 在 Informatica Developer 中应用规则

可以将规则添加到已保存的列配置文件中。无法将规则添加到配置为用于联接分析的配置文件中。

1. 浏览**对象浏览器**视图并查找所需的配置文件。
2. 右键单击配置文件并选择**打开**。  
配置文件将在编辑器中打开。
3. 单击**定义**选项卡，然后选择“规则”。
4. 单击**添加**。  
此时将打开**应用规则**对话框。
5. 单击**浏览**找到要应用的规则。  
从存储库项目中选择一个规则，然后单击**确定**。
6. 单击**输入值**下的**值**列以选择规则的输入端口。
7. 或者，单击**输出值**下的**值**列以编辑规则输出端口的名称。  
规则将显示在**定义**选项卡中。

## 第 12 章

# Mapplet 和映射剖析

本章包括以下主题：

- [Mapplet 和映射剖析概览, 86](#)
- [对 Mapplet 或映射对象运行配置文件, 86](#)
- [比较映射或 Mapplet 对象的配置文件, 87](#)
- [从配置文件生成映射, 87](#)

## Mapplet 和映射剖析概览

可以为 Mapplet 或映射中的对象定义列配置文件。如果要验证映射或 Mapplet 的设计而不保存配置文件结果，请对 Mapplet 或映射对象运行配置文件。还可以从配置文件生成映射。

## 对 Mapplet 或映射对象运行配置文件

在 Mapplet 或映射对象上运行配置文件时，配置文件在所有数据列上运行，并为针对数据对象暂存的数据启用向下钻取操作。可以在具有多个输出端口的 Mapplet 或映射对象上运行配置文件。

配置文件对通过映射并流向选定对象的输出端口的源数据进行跟踪。如果运行了映射，则配置文件会分析将显示在这些端口上的数据。

1. 打开 Mapplet 或映射。
2. 验证 Mapplet 或映射是否有效。
3. 右键单击数据对象或转换，然后选择**立即剖析**。

如果转换有多个输出组，将显示**选择输出组**对话框。如果转换有一个输出组，则配置文件结果将显示在配置文件的**结果**选项卡上。

4. 如果转换有多个输出组，则根据需要选择输出组。
5. 单击**确定**。

配置文件结果将显示在配置文件的**结果**选项卡上。

# 比较映射或 Mapplet 对象的配置文件

可以创建一个配置文件来分析 Mapplet 或映射中的两个对象，并且可以比较这些对象的列配置文件的结果。

与单个映射或 Mapplet 对象的列配置文件相似，配置文件比较在所有数据列上运行，并为针对数据对象暂存的数据启用向下钻取操作。将数据从源表移至目标表后，可以比较配置文件来验证数据迁移。还可以比较随时间变化的数据源上的配置文件。

与单个映射或 Mapplet 对象的配置文件相似，配置文件比较在所有数据列上运行。

1. 打开 Mapplet 或映射。
2. 验证 Mapplet 或映射是否有效。
3. 按 **CTRL** 键并单击编辑器中的两个对象。
4. 右键单击其中的一个对象，然后选择**比较配置文件**。
5. 或者，对配置文件比较进行配置，以匹配一个对象的列与其他对象的列。
6. 或者，单击一个对象中的列并将其拖到其他对象的列上以匹配列。
7. 或者，选择配置文件是分析所有列还是仅分析匹配的列。
8. 单击**确定**。

## 从配置文件生成映射

可以从配置文件创建映射对象。使用所创建的映射对象以开发有效的映射。所创建的映射具有一个基于已剖析对象的数据源，并可基于配置文件规则逻辑包含转换。创建映射后，添加对象以完成映射。

1. 在**对象浏览器**视图中，找到要在其中创建映射的配置文件。
2. 右键单击配置文件名称，然后选择**生成映射**。  
此时将显示**生成映射**对话框。
3. 输入映射名称。或者，输入映射的说明。
4. 确认映射的文件夹位置。  
默认情况下，Developer 工具在配置文件所在的项目中的**映射**文件夹内创建映射。单击**浏览**为映射选择不同的位置。
5. 确认 Developer 工具用于创建映射的配置文件定义。要使用其他配置文件，请单击**选择配置文件**。
6. 单击**完成**。  
映射将显示在**对象浏览器**中。  
将对象添加到映射中以完成映射。

## 第 13 章

# Informatica Developer 中的列配置文件结果

本章包括以下主题：

- [Informatica Developer 中的列配置文件结果, 88](#)
- [列值属性, 89](#)
- [列模式属性, 89](#)
- [列统计信息属性, 89](#)
- [列数据类型属性, 90](#)
- [Informatica Developer 中的内容管理, 90](#)
- [从 Informatica Developer 中导出配置文件结果, 91](#)

## Informatica Developer 中的列配置文件结果

列配置文件分析通过突出显示数据的值频率、模式和统计信息来提供有关数据质量的信息。

下表介绍了每一类分析的配置文件结果：

配置文件类型	配置文件结果
列配置文件	<ul style="list-style-type: none"><li>- 唯一值和空值的百分比和计数统计信息</li><li>- 推理的数据类型</li><li>- 数据源声明用于数据的数据类型</li><li>- 最大值和最小值</li><li>- 配置文件运行的最新日期和时间</li><li>- 列中每个唯一数据元素的百分比和计数统计信息</li><li>- 列中每个唯一字符模式的百分比和计数统计信息</li></ul>
主键配置文件	<ul style="list-style-type: none"><li>- 推理的主键</li><li>- 键冲突</li></ul>
功能相关性配置文件	<ul style="list-style-type: none"><li>- 推理的功能相关性</li><li>- 功能相关性冲突</li></ul>



## 列值属性

列值属性显示已剖析列中的值以及每个值在各列中出现的频率。频率以数字、百分比和条形图的形式显示。

要查看列值属性，请从**显示**列表中选择“值”。双击某一列值以向下钻取到包含该值的行。

下表介绍了列值的属性：

属性	说明
值	配置文件中列的所有值的列表。
频率	值在列中出现的次数。
百分比	值在列中出现的次数，表示为占列中所有值的百分比。
图	百分比的条形图。

## 列模式属性

列模式属性显示已剖析列中数据的模式以及模式在各列中出现的频率。模式显示为数字、百分比和条形图。

要查看模式信息，请从**显示**列表中选择“模式”。双击某一模式以向下钻取到包含该模式的行。

下表介绍了列值模式的属性：

属性	说明
模式	选定列的模式。
频率	模式在列中出现的次数。
百分比	模式在列中出现的次数，表示为占列中所有值的百分比。
图	百分比的条形图。

## 列统计信息属性

列统计信息包括各种属性，例如最大长度和最小长度，以及第一个值和最后一个值。

要查看统计信息，请从**显示**列表中选择“统计信息”。

下表介绍了列统计信息属性：

属性	说明
最大长度	列中最长值的长度。
最小长度	列中最短值的长度。
底部	列中的最后五个值。
顶部	列中的前五个值。
总和	数据类型为数值的列中所有值的总和。

**注意：**配置文件还将显示 Integer 类型的列的平均差和标准差统计信息。

## 列数据类型属性

列数据类型包括配置文件结果中每个列的所有推理的数据类型。

要查看数据类型信息，请从**显示**列表中选择**数据类型**。双击数据类型以向下钻取到包含该数据类型的行。

下表介绍了列数据类型的属性：

属性	说明
数据类型	配置文件中列的所有推理的数据类型的列表。
频率	数据类型在列中出现的次数，以数字表示。
遵从性百分比	数据类型在列中出现的百分比。
状态	<p>指示数据类型的状态。状态有“已推理”、“已批准”或“已拒绝”。</p> <p><b>已推理</b></p> <p>指示 Developer tool 推理的列数据类型。</p> <p><b>已批准</b></p> <p>指示列的已批准数据类型。批准数据类型后，该数据类型随即提交到模型存储库。</p> <p><b>已拒绝</b></p> <p>指示列的已拒绝数据类型。</p>

## Informatica Developer 中的内容管理

内容管理是验证和管理数据源的已发现元数据的过程，以便元数据适合使用和报告。在 Informatica Developer 中管理元数据时，可以批准、拒绝和重置配置文件结果中的推理的数据类型或数据域。

您可为一个列批准一个数据类型或数据域。可以对列隐藏拒绝的数据类型或数据域。批准或拒绝推理的数据类型或数据域后，您可以重置数据类型或数据域以还原已推理状态。

## 批准数据类型

配置文件结果包括数据源中每个列的推理数据类型、频率、遵从性百分比以及推理状态。您可以为每一列选择和批准单一数据类型。

1. 在**对象浏览器**视图中，选择配置文件并将其打开。
2. 验证您是否位于**结果**选项卡中。
3. 在**列剖析**视图中，选择一列以在右侧面板中查看值频率、模式、数据类型和统计信息。
4. 在**详细信息**面板下，从**显示**列表中选择**数据类型**。  
显示列的推理的数据类型。
5. 右键单击要批准的列并单击**批准**。  
数据类型的状态将更改为**已批准**。
6. 要还原数据类型的推理状态，请右键单击该数据类型，然后单击**重置**。

## 拒绝数据类型

默认情况下，Informatica Developer 会在配置文件结果中显示推理的数据类型。您可以拒绝已推理或已批准的数据类型。可以选择显示或隐藏拒绝的数据类型。

1. 在**对象浏览器**视图中，选择一个配置文件。
2. 双击该配置文件将其打开。  
配置文件将在选项卡中打开。
3. 在**列剖析**视图中，选择一行。
4. 要拒绝推理的列数据类型，请在右侧面板中选择**数据类型**视图。选择要拒绝的推理的数据类型，接着右键单击相应的行，然后选择**拒绝**。  
Informatica Developer 会在数据类型列表中灰显拒绝的数据类型。
5. 要隐藏拒绝的数据类型，请右键单击相应的行，然后选择**隐藏拒绝项**。
6. 要查看拒绝的数据类型，请右键单击其中一行，然后选择**显示拒绝项**。

# 从 Informatica Developer 中导出配置文件结果

您可以将列配置文件结果导出为 .csv 文件或 Microsoft Excel 文件。将配置文件结果导出为 Microsoft Excel 文件时，Developer tool 会将信息保存至 .xlsx 文件。

1. 在**对象浏览器**视图中，打开一个配置文件。
2. 或者，运行配置文件以更新配置文件结果。
3. 选择**结果**视图。
4. 选择一列。
5. 在**详细信息**下，选择**值**、**模式**或**数据类型**，然后单击**导出**图标。  
此时将打开**将数据导出到文件**对话框。
6. 接受或更改默认文件名。
7. 选择要导出的数据的类型。您可以选择**选定列的值**、**选定列的模式**、**选定列的数据类型**或**所有(摘要、值、模式、数据类型、统计信息、属性)**。

8. 单击**浏览**选择一个位置，并将文件本地保存在计算机上。
9. 如果不想将字段名称作为第一行导出，请清除**将字段名称作为第一行导出**复选框。
10. 单击**确定**。

## 第 14 章

# Informatica Developer 中的结果卡

本章包括以下主题：

- [Informatica Developer 中的结果卡概览, 93](#)
- [创建结果卡, 93](#)
- [为结果卡沿袭导出资源文件, 94](#)
- [查看 Informatica Developer 中的结果卡沿袭, 94](#)

## Informatica Developer 中的结果卡概览

结果卡是配置文件中对质量度量的一种图形表示形式。可以在 Developer 工具中查看结果卡。在 Developer 工具中创建结果卡后，可以连接到 Analyst 工具，以打开结果卡进行编辑。对数据对象中的当前数据或暂存在剖析仓库中的数据运行结果卡。

可以在 Analyst 工具中编辑、运行结果卡，还可以查看度量或度量组的结果卡沿袭。

## 创建结果卡

创建结果卡，然后将配置文件中的列添加到该结果卡中。必须首先运行配置文件，然后才能将列添加到结果卡中。

1. 在**对象浏览器**视图中，选择要在其中创建结果卡的项目或文件夹。
2. 单击**文件 > 新建 > 结果卡**。  
此时将显示**新建结果卡**对话框。
3. 单击**添加**。  
此时将显示**选择配置文件**对话框。选择包含要添加的列的配置文件。
4. 单击**确定**，然后单击**下一步**。
5. 选择要添加到结果卡中的列。  
默认情况下，结果卡向导将选择配置文件中定义的列和规则。无法添加未包括在配置文件中的列。
6. 单击**完成**。

Developer 工具将创建结果卡。

7. 或者，单击**使用 Informatica Analyst 打开**以连接到 Analyst 工具，然后在 Analyst 工具中打开结果卡。

## 为结果卡沿袭导出资源文件

可以导出包含结果卡和相关对象的项目，作为资源文件供 Metadata Manager 使用。在 Metadata Manager 中以 XML 格式使用已导出的资源文件，以为结果卡沿袭创建并加载资源。

1. 要打开**导出向导**，请单击**文件 > 导出**。
2. 选择 **Informatica > 资源文件(供 Metadata Manager 使用)**。
3. 单击**下一步**。
4. 单击**浏览**以选择包含需要导出的结果卡对象和沿袭的项目。
5. 单击**下一步**。
6. 选择要导出的结果卡对象。
7. 输入导出文件的名称和文件位置。
8. 要查看**导出向导**与选定对象一起导出的相关项目，请单击**下一步**。  
**导出向导**将显示相关对象。
9. 单击**完成**。

Developer 工具将对象导出到 XML 文件中。

## 查看 Informatica Developer 中的结果卡沿袭

要查看 Developer 工具中的度量或度量组的结果卡沿袭，请启动 Analyst 工具。

1. 在**对象浏览器**视图中，选择包含结果卡的项目或文件夹。
2. 双击该结果卡将其打开。  
此时结果卡将显示在选项卡中。
3. 单击**使用 Informatica Analyst 打开**。  
Analyst 工具将在浏览器窗口中打开。
4. 在 Analyst 工具的**结果卡**视图中，选择某一度量或度量组。
5. 右键单击然后选择**显示沿袭**。  
此时将在对话框中显示结果卡沿袭图表。

# 索引

## A

Analyst 工具中的列配置文件结果  
界面 [40, 50, 52](#)  
列详细信息 [41, 52](#)  
摘要 [39](#)

## B

表数据对象  
同步 [27](#)

## C

创建表达式规则  
规则 [30](#)  
创建列配置文件  
配置文件 [23](#)

## D

导出  
结果卡沿袭为 XML [94](#)

## G

规则  
应用预定义规则 [29](#)  
在 Informatica Developer 中创建 [85](#)  
表达式 [29](#)  
创建表达式规则 [30](#)  
使用规则规范创建表达式规则 [31](#)  
预定义 [28](#)  
在 Informatica Developer 中应用 [85](#)  
在 PowerCenter Express 中应用 [85](#)

## I

Informatica Analyst  
列配置文件概览 [19, 49](#)  
列配置文件结果 [38, 48](#)  
锁定和版本管理 [23](#)  
规则 [28](#)  
Informatica Developer  
规则 [84](#)

## J

结果卡  
编辑度量组 [64](#)

### 结果卡 (续)

查看 [61](#)  
创建度量组 [63](#)  
度量 [62](#)  
度量权重 [62](#)  
度量组 [63](#)  
概览 [15](#)  
固定成本 [62](#)  
Informatica Analyst [57](#)  
Informatica Analyst 进程 [58](#)  
Informatica Developer [93](#)  
可变成本 [62](#)  
配置全局通知设置 [70](#)  
配置通知 [70](#)  
趋势图表 [65](#)  
删除度量组 [64](#)  
通知 [68](#)  
无效数据的成本 [62](#)  
向下钻取 [64](#)  
移动得分 [63](#)  
运行 [60](#)  
编辑 [61](#)  
定义阈值 [62](#)  
向结果卡中添加列 [59](#)  
结果卡沿袭  
从 Informatica Developer 查看 [94](#)  
在 Informatica Analyst 中查看 [71](#)  
结果卡结果  
从 Informatica Analyst 中导出 [68](#)  
导出 [68](#)  
导出到 Excel [68](#)

## L

离群值  
检测 [44](#)  
列配置文件  
操作系统配置文件 [22](#)  
Informatica Developer [73](#)  
进程 [20](#)  
向下钻取 [53](#)  
概览 [14](#)  
选项 [15](#)  
列配置文件结果  
Informatica Developer [88](#)

## M

Mapplet 和映射剖析  
概览 [86](#)

## N

内容管理

概念 [16](#)

Informatica Analyst [54](#)

Informatica Developer [90](#)

任务 [16](#)

## P

配置文件

Avro 或 Parquets 格式 [80](#)

编辑筛选器 [36](#)

创建筛选器 [33](#)

简介 [11](#)

XML 和 JSON 格式 [78](#), [79](#)

运行 [25](#), [48](#), [49](#)

组件 [12](#)

编辑列配置文件 [24](#)

创建列配置文件 [23](#)

配置文件结果

从 Informatica Analyst 中导出 [56](#)

导出 [55](#)

Excel [55](#)

拒绝数据类型 [55](#)

列数据类型 [43](#), [90](#)

列值 [45](#)

批准数据类型 [54](#)

向下钻取 [54](#)

详细视图 [41](#)

在 Informatica Developer 中导出 [91](#)

摘要 [51](#), [53](#)

摘要视图 [39](#)

列模式 [44](#)

批准数据类型在 Informatica Developer 中 [91](#)

在 Developer tool 中拒绝数据类型 [91](#)

剖析

工具 [12](#)

剖析 (续)

进程 [12](#)

锁定和版本管理 [15](#)

平面文件数据对象

同步 [25](#)

## Q

趋势图表

查看 [66](#)

成本 [65](#)

从 Informatica Analyst 中导出 [67](#)

得分 [65](#)

## S

筛选器

概览 [33](#)

Sqoop 配置

剖析 [22](#)

数据对象配置文件

创建单个配置文件 [74](#)

创建多个配置文件 [75](#)

## Y

预定义规则

进程 [29](#)

映射对象

运行配置文件 [86](#)

运行时环境

Analyst 工具 [21](#)

Hadoop [21](#)