



Informatica®
10.4.0

新功能指南

Informatica 新功能指南

10.4.0

2019 年 12 月

© 版权所有 Informatica LLC 2009, 2019

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。

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation is subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License.

Informatica, the Informatica logo, Big Data Management, PowerCenter, and PowerExchange are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners.

Portions of this software and/or documentation are subject to copyright held by third parties. Required third party notices are included with the product.

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: 2019-12-20

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	8
Informatica Velocity	9
Informatica Marketplace	9
Informatica 全球客户支持部门	9
第 1 章：新增功能 (10.4.0)	10
CI/CD	10
数据集成服务 REST API	10
infacmd dis 命令	11
反向代理服务器	12
infacmd roh 命令	12
应用程序修补程序部署	12
连接到运行时应用程序	13
对象浏览器视图	13
标记	14
命令程序	14
infacmd isp 命令 (10.4.0 新增功能)	15
Data Engineering Integration	15
新数据类型支持	15
AWS Databricks 集成	16
通过 HDInsight 访问 ALDS Gen2 资源的群集工作流	16
Databricks Delta Lake 存储访问	16
显示映射使用的节点	16
日志汇总	16
在 Spark 引擎上解析层次结构数据	17
Spark 引擎上的配置文件和采样选项	17
Python 转换	17
Sqoop	18
Data Engineering Streaming	18
流映射中的 Confluent 架构注册表	18
流映射中的数据质量转换	18
流映射中的临时群集	19
Amazon S3 的 FileName 端口	19
Microsoft Azure Data Lake Storage Gen2	19
Azure Databricks 中的流映射	19

Data Engineering Streaming 中的动态映射	20
Enterprise Data Catalog.	20
为资源和类分配自定义属性.	20
新资源.	20
引用资源和引用资产.	21
从搜索结果导出资产.	21
沿袭和影响筛选器.	21
资产控制摘要.	21
规则 and 结果卡.	22
唯一键推理.	22
CLOB 文件类型上的数据域发现.	22
Spark 引擎上的数据发现和采样选项.	22
跟踪技术预览.	22
数据预览和置备.	23
独立扫描实用程序支持的资源类型.	23
REST API.	24
Enterprise Data Preparation.	24
Data Lake 访问管理.	24
Microsoft Azure Data Lake Storage 用作数据源.	24
将文件发布到 Data Lake.	24
将文件上传到 Data Lake.	24
Informatica 映射	25
将映射输出绑定至映射参数.	25
CLAIRE 建议和见解.	26
更新映射优化器级别.	26
Informatica 转换.	26
地址验证器转换.	26
Informatica 工作流.	26
Amazon EMR 创建群集任务高级属性.	26
Informatica 安装.	27
PostgreSQL.	27
静默模式下的预安装 (i10Pi) 系统检查工具.	27
对静默安装属性文件的密码进行加密.	27
智能结构模型.	27
其他输入类型.	27
在设计时从示例创建模型.	27
处理无法识别的数据.	28
PowerCenter.	28
连接.	28
PowerExchange 适配器.	29
Informatica PowerExchange 适配器.	29
PowerCenter PowerExchange 适配器.	31

安全.	33
配置 Web 应用程序以使用不同的 SAML 身份提供程序.	33
第 2 章：新增功能 (10.2.2 HotFix 1).	34
命令行程序.	34
infacmd ldm 命令.	34
Enterprise Data Catalog.	34
Azure Data Lake Storage Gen2.	34
链接不区分大小写.	35
Enterprise Data CatalogTableau Extension	35
新资源.	35
剖析仓库扫描程序.	35
REST API.	35
搜索增强功能.	35
搜索标签.	36
独立扫描实用程序支持的资源类型.	36
技术预览版.	36
第 3 章：新增功能 (10.2.2 Service Pack 1).	38
Big Data Management.	38
Sqoop.	38
大数据 Streaming.	39
Amazon S3 目标.	39
TIME_RANGE 函数.	39
将数据写入多个 HDFS 文件.	39
Enterprise Data Catalog.	39
独立扫描仪实用程序支持的资源类型.	39
Enterprise Data Preparation.	40
还原所有已推理数据类型.	40
准备 Avro 和 Parquet 文件.	40
Informatica PowerExchange 适配器.	40
PowerExchange for Hive.	40
第 4 章：新增功能 (10.2.2).	41
应用程序服务.	41
Mass Ingestion 服务.	41
元数据访问服务.	41
REST 操作 Hub 服务.	41
Big Data Management.	42
Azure Databricks 集成.	42
层次结构数据.	43
高可用性.	44
智能结构模型.	44

Mass Ingestion.	45
监视.	45
安全.	45
目标.	45
大数据 Streaming.	46
Azure 事件中心数据对象.	46
Amazon Kinesis 连接中的跨帐户 IAM 角色.	46
智能结构模型.	46
Big Data Streaming 数据对象的标头端口.	46
Amazon Kinesis 连接中的 AWS 凭据配置文件.	46
Spark 结构化流.	47
窗口转换.	47
命令行程序.	47
infacmd dis 命令.	47
infacmd ihs 命令.	48
infacmd ipc 命令.	48
infacmd ldm 命令.	48
infacmd mi 命令.	49
infacmd ms 命令.	50
infacmd oie 命令.	50
infacmd tools 命令.	50
infasetup 命令.	51
Enterprise Data Catalog.	51
为列自动分配业务标题.	51
资产方面的用户协作.	51
使用安装程序创建 Enterprise Data Catalog 应用程序服务.	52
自定义元数据验证实用程序.	52
更改通知.	52
Business Glossary 分配报告.	52
操作系统配置文件.	52
REST API.	52
源元数据和数据配置文件筛选器.	52
扫描程序实用程序.	53
资源类型.	53
Enterprise Data Lake.	53
应用活动规则.	53
删除重复行.	54
聚集和分类列数据.	54
基于 CLAIRE 的建议.	54
条件性汇总.	54
数据屏蔽.	54
本地化.	54

分区源和目标.	55
向脚本步骤添加注释.	55
将脚本另存为映射.	55
Amazon S3、ADLS、WASB、MapR-FS 作为数据源.	55
统计函数.	55
日期和时间函数.	56
数学函数.	56
文本函数.	56
窗口函数.	57
清除审计事件.	57
Spark 执行引擎.	57
Informatica Developer.	58
应用程序.	58
Informatica 映射.	58
数据类型.	58
映射输出.	58
映射参数.	58
优化器级别.	58
Sqoop.	59
Informatica 转换.	59
地址验证器转换.	59
更新策略转换.	62
Informatica PowerExchange 适配器.	62
PowerExchange for Amazon Redshift.	62
PowerExchange for Amazon S3.	63
PowerExchange for Google BigQuery.	63
PowerExchange for HBase.	63
PowerExchange for HDFS.	64
PowerExchange for Hive.	64
PowerExchange for MapR-DB.	65
PowerExchange for Microsoft Azure Blob 存储.	65
PowerExchange for Microsoft Azure Cosmos DB SQL API.	65
PowerExchange for Microsoft Azure Data Lake Store.	65
PowerExchange for Microsoft Azure SQL 数据仓库.	66
PowerExchange for Salesforce.	66
PowerExchange for Snowflake.	67
PowerExchange for Teradata Parallel Transporter API.	67

前言

请阅读《Informatica® 新功能和增强功能指南》以了解此版本中 Informatica 产品的新功能和增强功能。

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

新增功能 (10.4.0)

本章包括以下主题：

- [CI/CD, 10](#)
- [命令行程序, 14](#)
- [Data Engineering Integration, 15](#)
- [Data Engineering Streaming, 18](#)
- [Enterprise Data Catalog, 20](#)
- [Enterprise Data Preparation, 24](#)
- [Informatica 映射, 25](#)
- [Informatica 转换, 26](#)
- [Informatica 工作流, 26](#)
- [Informatica 安装, 27](#)
- [智能结构模型, 27](#)
- [PowerCenter, 28](#)
- [PowerExchange 适配器, 29](#)
- [安全, 33](#)

CI/CD

本节介绍版本 10.4.0 中的 CI/CD 增强功能。

CI/CD（持续集成/持续交付）可在 CI/CD 管道中自动执行集成和交付操作。在版本 10.4.0 中，您可以将增强功能并入 CI/CD 管道，以改进在生产环境中部署、测试和交付对象的方式。

数据集成服务 REST API

从版本 10.4.0 开始，您可以使用数据集成服务 REST API 自动在 CI/CD 管道中执行任务。

REST API 可以自动执行的部分任务有：

查询对象。

查询对象，包括模型存储库中的设计时对象和部署到数据集成服务的运行时对象。

您可将查询传递到其他 REST API 请求。例如，您可以将查询传递到版本控制操作，以对特定对象集执行版本控制。您还可以传递查询，以将特定的设计时对象部署到应用程序修补程序存档文件。

执行版本控制操作。

执行版本控制操作以签入、签出、撤消签出或将已签出的设计时对象重新分配给另一个开发人员。

管理标记。

管理分配给设计时对象的标记。您可以为对象分配新标记，也可以替换标记。还可以为对象取消标记。

更新应用程序。

将设计时对象部署到应用程序修补程序存档文件，并将该文件部署到数据集成服务，以更新已部署的增量式应用程序。

管理应用程序。

根据项目或组织要求启动、取消部署或停止应用程序。

运行已部署映射。

运行已部署映射以测试应用程序输出。

比较映射。

比较同一个域中的两个映射。

例如，您可以比较两个设计时映射或设计时映射与运行时映射。

要查看您可以使用的 REST API 请求和每个请求的参数，可以通过 Administrator tool 中的数据集成服务进程属性或 REST 操作 Hub 服务属性来访问 REST API 文档。

与 infacmd 命令行程序相比，REST API 没有任何设置要求，您可以在客户端计算机未安装 Informatica 服务的环境中运行 REST API。

有关 REST API 信息，请参阅《Informatica 10.4.0 应用程序服务指南》中的“数据集成服务 REST API”一章。

infacmd dis 命令

从版本 10.4.0 开始，您可以使用 infacmd dis 命令对设计时对象、应用程序中的运行时对象和已部署映射执行操作。您还可以使用 infacmd dis 命令运行数据集成服务实用程序。

下表介绍了新的 infacmd dis 命令：

命令	说明
compareMapping	比较两个查询的映射。查询映射以比较映射属性、转换属性和转换中的端口。要查询设计时映射，请指定设计时模型存储库。要查询运行时映射，请勿指定模型存储库。此查询将使用您指定的数据集成服务运行命令。
deployObjectsToFile	将设计时对象部署到应用程序修补程序存档文件。
queryDesignTimeObjects	从模型存储库服务查询设计时对象。
queryRuntimeObjects	查询部署到数据集成服务的运行时对象并返回对象列表。
replaceAllTag	将模型存储库服务中查询对象的标记替换为指定标记。
标记	为模型存储库服务中的查询对象分配标记。

命令	说明
untag	为模型存储库服务中的查询对象删除标记。
listPatchNames	列出已应用于增量式应用程序的所有修补程序。

有关详细信息，请参阅《*Informatica® 10.4.0 命令引用*》中的“infacmd dis 命令引用”一章。

反向代理服务器

从版本 10.4.0 开始，您可以使用 REST 操作 Hub 启动反向代理服务器，该服务器可针对跨数据集成服务网格中的节点的数据集成服务 API 请求执行负载平衡。

REST 操作 Hub 面向网络。

有关反向代理服务器的信息，请参阅《Informatica 10.4.0 应用程序服务指南》中的“系统服务”一章。

infacmd roh 命令

从版本 10.4.0 开始，您可以使用以下 infacmd roh 命令更新 REST 操作 Hub 服务进程。

下表介绍了新的 infacmd roh 命令：

命令	说明
listReverseProxyServerOptions	列出反向代理服务器属性。
listServiceProcessOptions	列出 REST 操作 Hub 服务进程属性。
updateServiceOptions	更新 REST 操作 Hub 服务属性。
updateReverseProxyServerOptions	更新域中的反向代理服务器属性。

从版本 10.4.0 开始，重命名了以下 infacmd roh 命令：

- listROHProperties 重命名为 listProcessProperties。
- updateROHService 重命名为 updateServiceProcessOptions。

注意：请更新所有使用了以前版本的命令名的脚本。

有关详细信息，请参阅《*Informatica 10.4.0 命令引用*》中的“infacmd roh 命令引用”一章。

应用程序修补程序部署

从版本 10.4.0 开始，应用程序修补程序部署的以下部分增加了新功能：

增量部署向导

从版本 10.4.0 开始，您可在增量部署向导中执行以下任务：

- （可选）可以输入修补程序说明。
- （可选）可以选择保留或丢弃状态信息。
状态信息是指映射属性和运行时对象的属性，例如映射输出或序列生成器转换。

有关增量部署向导的详细信息，请参阅《*Informatica 10.4.0 Developer Tool 指南*》中的“应用程序修补程序部署”一章。

有关状态信息的详细信息，请参阅《*Informatica 10.4.0 Developer tool 指南*》中的“应用程序部署”一章。

修补程序历史记录

从版本 10.4.0 开始，增量部署向导中的修补程序历史记录会显示为更新增量式应用程序而部署的修补程序的修补程序名称和修补程序说明。修补程序创建时间将附加到修补程序说明的开头。

此外，您还可以使用 Administrator tool 查看已部署的增量式应用程序的修补程序历史记录。

有关修补程序历史记录的详细信息，请参阅《*Informatica 10.4.0 Developer Tool 指南*》中的“应用程序修补程序部署”一章。

有关已部署的应用程序的详细信息，请参阅《*Informatica 10.4.0 应用程序服务指南*》中的“数据集成服务应用程序”一章。

应用程序修补程序存档文件

从版本 10.4.0 开始，应用程序修补程序存档文件还会额外存储修补程序说明并指示是否保留状态信息。

有关修补程序历史记录的详细信息，请参阅《*Informatica 10.4.0 Developer Tool 指南*》中的“应用程序修补程序部署”一章。

infacmd tools 命令

从版本 10.4.0 开始，infacmd tools patchApplication 命令包含以下新选项：

选项	参数	说明
-RetainStateInformation -rsi	True False	可选。指示保留还是丢弃状态信息。状态信息是指映射属性和运行时对象的属性，例如映射输出或序列生成器转换。

有关详细信息，请参阅《*Informatica 10.4.0 命令引用*》中的“infacmd tools 命令引用”一章。

连接到运行时应用程序

从版本 10.4.0 开始，您可以在 Developer tool 中连接到运行时应用程序。连接到运行时应用程序后，可在**对象浏览器**视图中展开该应用程序，并在**编辑器**中打开运行时对象的只读副本。

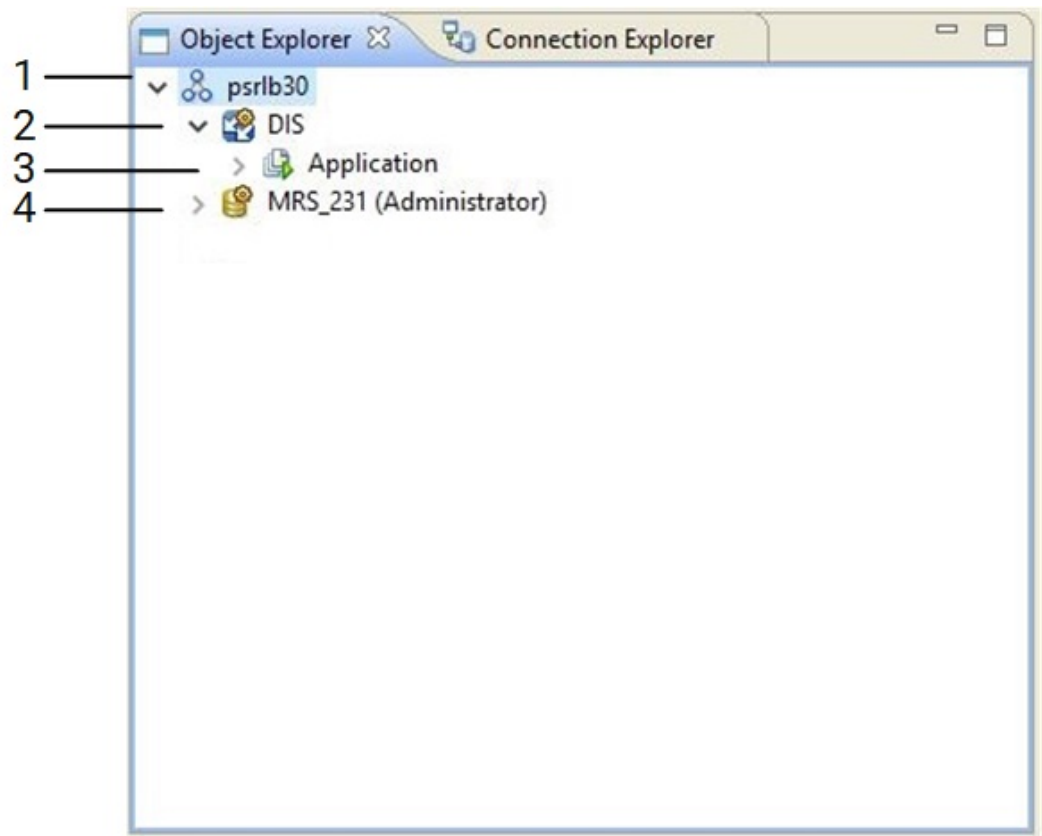
连接到运行时应用程序后，您在 Developer tool 中执行的搜索可查找应用程序中的运行时对象。

有关连接到运行时应用程序和查看运行时对象的详细信息，请参阅《*Informatica 10.4.0 Developer Tool 指南*》中的“应用程序部署”一章。

对象浏览器视图

从版本 10.4.0 开始，在您连接至模型存储库或运行时应用程序后，Developer tool 中的**对象浏览器**视图将显示域。您可以展开域，以查看模型存储库中的设计时对象或运行时应用程序中的运行时对象。

下图显示了对象浏览器视图：



1. 域
2. 数据集成服务
3. 运行时应用程序
4. 模型存储库

有关 Developer tool 中用户界面的详细信息，请参阅《*Informatica 10.4.0 Developer Tool 指南*》中的“Informatica Developer”一章。

标记

从版本 10.4.0 开始，标记包含以下功能：

- 部署与标记关联的映射时，标记会传播到数据集成服务中映射的运行时版本。
- 如果使用应用程序修补程序更新已部署的映射，修补程序名称将作为标记与映射的运行时版本关联。

有关标记的详细信息，请参阅《*Informatica 10.4.0 Developer Tool 指南*》中的 "Informatica Developer" 一章。

命令行程序

本节介绍版本 10.4.0 中的新命令。

infacmd isp 命令（10.4.0 新增功能）

下表介绍了新的 infacmd isp 命令：

命令	说明
addCustomLDAPType	添加定义 LDAP 目录服务的自定义 LDAP 类型。
listAllCustomLDAPTypes	列出指定域使用的所有自定义 LDAP 类型的配置信息。
listAllLDAPConnectivity	列出指定域使用的所有 LDAP 配置的配置信息。
listCustomLDAPType	列出自定义 LDAP 类型的配置信息。
removeCustomLDAPType	从指定域中删除指定的自定义 LDAP 类型。
removeLDAPConnectivity	从指定域中删除指定的 LDAP 配置。
updateCustomLDAPType	更新指定的自定义 LDAP 类型。
updateLDAPConnectivity	更新指定的 LDAP 配置。

有关详细信息，请参阅《*Informatica 10.4.0 命令引用*》。

Data Engineering Integration

本节介绍版本 10.4.0 中新增的 Data Engineering Integration 功能。

新数据类型支持

从版本 10.4.0 开始，您可以在复杂文件中使用以下新的数据类型：

- 在本地环境或 Hadoop 环境中运行读取或写入 Avro 和 Parquet 复杂文件对象的映射时，可以使用以下数据类型：
 - Date
 - Decimal
 - Timestamp
- 您可以使用 Time 数据类型在本地环境或 Blaze 引擎上读取和写入 Avro 或 Parquet 复杂文件对象。
- 在 Databricks Spark 引擎上运行映射时，可以使用适用的 Date、Time、Timestamp 和 Decimal 数据类型。

新的数据类型适用于以下适配器：

- PowerExchange for HDFS
- PowerExchange for Amazon S3
- PowerExchange for Google Cloud Storage
- PowerExchange for Microsoft Azure Blob 存储
- PowerExchange for Microsoft Azure Data Lake Storage Gen1

- PowerExchange for Microsoft Azure Data Lake Storage Gen2

有关数据类型的详细信息，请参阅《*Data Engineering Integration 10.4.0 用户指南*》中的“数据类型引用”。

AWS Databricks 集成

从版本 10.4.0 开始，您可以将 Informatica 域与 AWS 上运行的 Databricks 集成。

您可以使用 AWS Databricks 通过以下功能运行映射：

- 您可对 Databricks 环境中的 Amazon Simple Storage Service (S3) 和 Amazon Redshift 源与目录运行映射。
- 您可以开发群集工作流以使用 AWS 上运行的 Databricks 创建临时群集。
- 您可以将 Python 转换添加到配置为在 Databricks Spark 引擎上运行的映射中。
仅技术预览版支持 Python 转换。

AWS Databricks 支持的数据类型与 Azure Databricks 一致。

有关详细信息，请参阅以下指南：

《*Data Engineering 10.4.0 Integration 指南*》

《*Data Engineering 10.4.0 管理员指南*》

《*Data Engineering Integration 10.4.0 用户指南*》

《*Informatica 10.4.0 Developer 工作流指南*》

通过 HDInsight 访问 ALDS Gen2 资源的群集工作流

从版本 10.4.0 开始，您可以创建可在 Azure HDInsight 群集上运行以访问 ADLS Gen2 资源的群集工作流。

有关群集工作流的详细信息，请参阅《*Informatica Data Engineering Integration 10.4.0 用户指南*》。

Databricks Delta Lake 存储访问

从版本 10.4.0 开始，您可以将 Databricks Delta Lake 存储作为源和目标进行访问。

映射可以访问 AWS 和 Azure 平台上的 Delta Lake 资源。

有关配置对 Delta Lake 表的访问的信息，请参阅 *Data Engineering Integration Guide*。有关创建映射以访问 Delta Lake 表的信息，请参阅 *Data Engineering Integration User Guide*。

显示映射使用的节点

从版本 10.4.0 开始，您可以查看给定时段内某个映射使用的最大群集节点数。

您可以使用 REST 操作 Hub API ClusterStats(startTimeInmillis=[value], endTimeInmillis=[value]) 查看给定时段内某个映射使用的群集配置的最大 Hadoop 节点数。

有关 REST API 的详细信息，请参阅《*Data Engineering 10.4.0 管理员指南*》的“监视 REST API 引用”一章。

日志汇总

从版本 10.4.0 开始，可以获取在 Hadoop 环境中运行的已部署映射的汇总日志。

您可以根据 Monitoring 工具中的作业 ID 或使用 `infacmd ms fetchAggregatedClusterLogs` 命令收集映射的汇总群集日志。您可以根据作业 ID 获取映射的汇总群集日志的 .zip 或 tar.gz 文件，并将压缩的汇总日志文件写入目标目录。

有关详细信息，请参阅《*Informatica 10.4.0 管理员指南*》。

在 Spark 引擎上解析层次结构数据

从 10.4.0 开始，您可以使用复杂函数在映射中解析最多 5 MB 的数据中游。

Spark 引擎可以使用以下复杂函数解析原始字符串源数据：

- PARSE_JSON
- PARSE_XML

复杂函数可以解析源字符串中的 JSON 或 XML 数据并生成结构目标数据。

有关详细信息，请参阅《*Informatica Data Engineering Integration 10.4.0 用户指南*》中的“层次结构数据处理”。

有关复杂函数的详细信息，请参阅《*Informatica 10.4.0 Developer 转换语言参考*》中的“函数”一章。

Spark 引擎上的配置文件和采样选项

从版本 10.4.0 开始，您可以在 Spark 上运行配置文件并选择采样选项。

在 Spark 引擎上执行剖析

您可在 Informatica Developer 和 Informatica Analyst 工具中在 Spark 引擎上创建和运行配置文件。您可以在 Spark 引擎上执行数据域发现和创建结果卡。

Spark 引擎上的采样选项

您可以选择以下采样选项，以在 Spark 引擎上运行配置文件。

- **限制 n** 采样选项会基于数据对象中的行数运行配置文件。选择在 Hadoop 环境中运行配置文件时，Spark 引擎会从数据对象的多个分区中收集样本，并将样本推送到单个节点以计算样本大小。对于包含高级筛选器的配置文件，不能应用“限制 n”采样选项。
支持的平台：Oracle、SQL Server 和 DB2 数据库。
- **随机百分比** 采样选项会对数据对象中一定百分比的行运行配置文件。

有关 Spark 引擎上的配置文件和采样选项的信息，请参阅《*Informatica 10.4.0 Data Discovery 指南*》。

Python 转换

从版本 10.4.0 开始，Python 转换包含以下功能：

活动模式

您可以创建活动 Python 转换。作为活动转换，Python 转换可更改传递它的行的数量。例如，Python 转换可从单个输入行生成多个输出行，或者可从多个输入行生成单个输出行。

有关详细信息，请参阅《*Informatica Data Engineering Integration 10.4.0 用户指南*》中的“Python 转换”一章。

分区数据

您可以运行 Python 代码以根据数据的默认分区方案处理传入数据，或者您也可以在 Python 代码运行前对数据进行重新分区。要在 Python 代码运行前对数据进行重新分区，请选择一个或多个输入端口作为分区键。

有关详细信息，请参阅《*Informatica Data Engineering Integration 10.4.0 用户指南*》中的“Python 转换”一章。

Sqoop

从版本 10.4.0 开始，您可以在 JDBC 连接中配置以下 Sqoop 参数：

- --update-key
- --update-mode
- --validate
- --validation-failurehandler
- --validation-threshold
- --validator
- --mapreduce-job-name
- --bindir
- --class-name
- --jar-file
- --outdir
- --package-name

有关配置这些 Sqoop 参数的详细信息，请参阅 Sqoop 文档。

Data Engineering Streaming

本节介绍版本 10.4.0 中新增的 Data Engineering Streaming 功能。

流映射中的 Confluent 架构注册表

从版本 10.4.0 开始，您可以使用 Confluent Kafka 作为使用架构注册表的流映射中的源和目标。

您可以使用 Confluent Kafka 在流映射中存储和检索 Apache Avro 架构。架构注册表使用 Kafka 作为其基础存储机制。

有关详细信息，请参阅《*Data Engineering Streaming 10.4.0 用户指南*》。

流映射中的数据质量转换

从版本 10.4.0 开始，可以在流映射中使用数据质量转换。

您可以在流映射中使用以下数据质量转换，以对流数据应用数据质量处理：

- 地址验证器转换
- 分类器转换
- 解析器转换
- 标准创建器转换

有关详细信息，请参阅《*Data Engineering Streaming 10.4.0 用户指南*》。

流映射中的临时群集

从版本 10.4.0 开始，您可以运行工作流来创建可在云平台群集上运行映射和其他任务的临时群集。

要从群集被删除的点继续进行数据处理，可以通过指定外部存储和检查点目录在临时群集中运行流映射。

有关详细信息，请参阅《*Data Engineering Streaming 10.4.0 用户指南*》。

Amazon S3 的 FileName 端口

从版本 10.4.0 开始，在您为 Amazon S3 文件创建数据对象写入操作时，将默认显示 FileName 端口。

在运行时，数据集成服务会为 FileName 端口中的每个值创建单独的目录，并在目录中添加目标文件。

有关详细信息，请参阅《*Data Engineering Streaming 10.4.0 用户指南*》。

Microsoft Azure Data Lake Storage Gen2

从版本 10.4.0 开始，您可以在流映射中将 Microsoft Azure Data Lake Storage Gen2 作为目标。

Azure Data Lake Storage Gen2 构建在 Azure Blob Storage 的基础上。Azure Data Lake Storage Gen2 兼具 Azure Data Lake Storage Gen1 和 Azure Blob Storage 的功能。您可以使用 Azure Databricks 版本 5.4 或 Azure HDInsight 版本 4.0 访问 Azure Data Lake Storage Gen2 中存储的数据。

有关详细信息，请参阅《*Data Engineering Streaming 10.4.0 用户指南*》。

Azure Databricks 中的流映射

从版本 10.4.0 开始，您可在 Microsoft Azure 云服务的 Azure Databricks 服务中运行流映射。

源和目标

可以在 Databricks 环境中针对以下源和目标运行流映射：

Microsoft Azure 事件中心

Azure Data Lake Storage Gen2 (ADLS Gen2)

转换

您可以将以下转换添加到 Databricks 流映射：

汇总器

表达式

筛选器

联接器

规范器

等级

路由器

联合

窗口

数据类型

支持以下数据类型：

数组

长整型
日期/时间
小数
双精度型
整型
映射
结构
文本
字符串

工作流

您可以开发群集工作流以在 Databricks 环境中创建临时群集。可以使用 Azure Data Lake Storage Gen1 (ADLS Gen1) 和 Azure Data Lake Storage Gen2 (ADLS Gen2) 在 Databricks 环境中创建临时群集。

有关 Azure Databricks 中流映射的详细信息，请参阅《*Data Engineering Streaming 10.4.0 用户指南*》。

Data Engineering Streaming 中的动态映射

从版本 10.4.0 开始，可对 Data Engineering Streaming 中的动态映射支持进行技术预览。

可以在流映射中使用 Confluent Kafka 数据对象作为动态源和目标。

支持使用技术预览功能进行评估，但不对此做出保证且尚未做好生产准备。Informatica 建议您只在非生产环境中使用。Informatica 打算将预览功能包含在即将推出的版本中以供生产使用，但也可能根据不断变化的市场或技术环境而选择不这样做。有关详细信息，请联系 Informatica 全球客户支持部门。

Enterprise Data Catalog

本节介绍版本 10.4.0 中新增的 Enterprise Data Catalog 功能。

为资源和类分配自定义属性

从版本 10.4.0 开始，您可将自定义属性分配给目录中的特定资源和类，也可以分配给目录中的所有资源和类。您可在创建或修改自定义属性时执行此任务。

有关详细信息，请参阅《*Informatica 10.4.0 Catalog Administrator 指南*》。

新资源

从版本 10.4.0 开始，Enterprise Data Catalog 中添加了以下新资源：

- AWS Glue
- Microsoft Power BI
- Apache Cassandra
- Snowflake
- Google Cloud Storage

您可以从上述所有资源中提取元数据、关系和沿袭信息。有关详细信息，请参阅《*Informatica 10.4.0 Enterprise Data Catalog 扫描程序配置指南*》。

引用资源和引用资产

从版本 10.4.0 开始，您可以配置资源来提取有关数据源或该资源引用的目录中其他资源的元数据。示例包括 PowerCenter 映射中的源和目标表，以及 Tableau 报告中的源表和文件。这些被引用的数据源称为引用资源，数据源中包含的资产则被称为引用资产。您可在 Enterprise Data Catalog 应用程序中查看引用资源和引用资产。要在 Enterprise Data Catalog 中查看完整沿袭，可为引用资源执行连接分配。您可在引用资源与您在 Enterprise Data Catalog 中为引用资源数据源配置的资源之间执行连接分配。

您可以配置以下资源来提取有关数据源或该资源引用的目录中其他资源的元数据：

- PowerCenter
- AWS Glue
- Tableau 服务器
- Coudera Navigator
- Apache Atlas
- Informatica Intelligent Cloud Services
- Informatica Platform
- SQL Server Integration Service

有关详细信息，请参阅《*Informatica 10.4.0 Catalog 管理员指南*》和《*Informatica 10.4.0 Enterprise Data Catalog 用户指南*》。

从搜索结果导出资产

从版本 10.4.0 开始，您可以将搜索结果中显示的所有资产导出到逗号分隔值 (CSV) 文件并将 CSV 文件导回到 Enterprise Data Catalog。您在任意搜索选项卡上优化搜索结果后，即可导出资产。

有关详细信息，请参阅《*Informatica 10.4.0 Enterprise Data Catalog 用户指南*》中的“资产任务”一章。

沿袭和影响筛选器

从版本 10.4.0 开始，您可以在“沿袭和影响”视图中创建并应用筛选器。使用筛选器可以缩小“沿袭和影响”视图中的结果范围。可以根据“沿袭和影响”视图中可用的资产类型，使用不同的筛选器选项组合创建自定义筛选器。

有关详细信息，请参阅《*Informatica 10.4.0 Enterprise Data Catalog 用户指南*》中的“查看沿袭和影响”一章。

资产控制摘要

从版本 10.4.0 开始，您可在**资产控制摘要**选项卡中查看表和文件资产类型的控制流资产。控制流用于将约束或条件放置在数据流上。例如，SQL 查询可在 WHERE 子句中包含约束，或者映射可以包含转换。

有关详细信息，请参阅《*Informatica 10.4.0 Enterprise Data Catalog 用户指南*》中的“查看沿袭和影响”一章。

规则和结果卡

从版本 10.4.0 开始，可以从剖析仓库提取规则和结果卡结果，然后在 Enterprise Data Catalog 中查看这些结果。您可以查看规则和结果卡的列结果，以衡量数据源的数据质量。

有关详细信息，请参阅《*Informatica 10.4.0 Enterprise Data Catalog 用户指南*》中的“查看资产”一章。

唯一键推理

从版本 10.4.0 开始，您可以从源数据对象的列中生成唯一键候选项。通过标识包含重复值的列，可以查看唯一键推理结果，以了解数据质量问题。

您可以接受或拒绝推理的唯一键推理结果。接受或拒绝推理的唯一键推理后，可以重置唯一键推理以还原推理的状态。

有关详细信息，请参阅《*Informatica 10.4.0 Enterprise Data Catalog 用户指南*》中的“查看资产”一章。

CLOB 文件类型上的数据域发现

从版本 10.4.0 开始，可对 CLOB 文件类型执行数据域发现。

有关详细信息，请参阅《*Informatica 10.4.0 Enterprise Catalog 管理员指南*》中的“Enterprise Data Catalog 概念”一章。

Spark 引擎上的数据发现和采样选项

从版本 10.4.0 开始，您可以在 Spark 引擎上运行配置文件以发现数据域和选择采样选项。

Spark 引擎上的数据域发现

您可以在 Spark 引擎上执行数据域发现。

Spark 引擎上的采样选项

您可以选择以下采样选项，以在 Spark 引擎上发现数据域：

- **限制 n** 采样选项会基于数据对象中的行数运行配置文件。选择在 Hadoop 环境中发现数据域时，Spark 引擎会从数据对象的多个分区中收集样本，并将样本推送到单个节点以计算样本大小。
- **随机百分比** 采样选项会对数据对象中一定百分比的行运行配置文件。

有关详细信息，请参阅《*Informatica 10.4.0 Enterprise Catalog 管理员指南*》中的“Enterprise Data Catalog 概念”一章。

跟踪技术预览

Enterprise Data Catalog 版本 10.4.0 包含可供技术预览版使用的功能。

支持技术预览版功能，但不对此做出保证且不能用于生产。Informatica 建议您只在非生产环境中使用。Informatica 打算将预览功能包含在即将推出的 GA 版本中以供生产使用，但也可能根据不断变化的市场或技术环境而选择不这样做。有关详细信息，请联系 Informatica 全球客户支持部门。

- 从版本 10.4.0 开始，您可以选择显示“沿袭和影响”视图的精简视图。“沿袭和影响”精简视图显示资源级别的沿袭和影响汇总图表。
有关详细信息，请参阅《*Informatica 10.4.0 Enterprise Data Catalog 用户指南*》中的“查看沿袭和影响”一章。
- 从版本 10.4.0 开始，无法在运行时或脱机情况下访问元数据时，可从 SAP Business Warehouse、SAP BW/4HANA、IBM InfoSphere DataStage 和 Oracle Data Integrator 源提取元数据。
有关详细信息，请参阅《*Informatica 10.4.0 Catalog 管理员指南*》。

- 从版本 10.4.0 开始，您可以从 SAP Business Warehouse 和 SAP BW/4HANA 数据源提取元数据。有关详细信息，请参阅《*Informatica 10.4.0 Enterprise Data Catalog 扫描程序配置指南*》。

数据预览和置备

从版本 10.4.0 开始，在目录中完成数据发现后，您可以执行数据置备。数据置备可帮助您将数据移动到目标，以进行进一步分析。您可以预览关系源的数据以评估数据，然后再将数据移到目标中。

有关预览和置备数据的详细信息，请参阅《*Informatica 10.4.0 Catalog 管理员指南*》和《*Informatica 10.4.0 Enterprise Data Catalog 用户指南*》。

独立扫描实用程序支持的资源类型

从版本 10.4 开始，无法在运行时或脱机情况下访问元数据时，可以从以下外部源提取元数据：

- Amazon Redshift
- Amazon S3
- Apache Cassandra
- Axon
- Azure Data Lake Store
- Azure Microsoft SQL 数据仓库
- Azure Microsoft SQL Server
- Business Glossary
- 自定义沿袭
- 数据库脚本
- Erwin
- Glue
- Google BigQuery
- Google Cloud Storage
- Informatica Cloud Services
- IBM Cognos
- Microsoft Azure Blob 存储
- Microsoft SQL Server Integration Services
- PowerBI
- QlikView Business Intelligence
- Salesforce
- SAP HANA
- SAP PowerDesigner
- Snowflake
- Workday

有关详细信息，请参阅《*Informatica 10.4 Enterprise Data Catalog 管理员指南*》中的“从脱机和无法访问的资源提取元数据”一章。

REST API

从版本 10.4 开始，可以使用以下 Informatica Enterprise Data Catalog REST API：

- 数据置备 REST API。除了现有的 REST API，您还可以查看用户是否可使用数据置备以及列出支持数据置备的资源。
- 沿袭筛选器 REST API。您可以创建、更新、列出或删除沿袭筛选器。
- 模型信息 REST API。除了现有 REST API，您还可以列出预定义的滑块分面、滑块分面定义和沿袭筛选器定义。
- 模型修改 REST API。除了现有 REST API，您还可以创建、更新和删除滑块分面定义。
- 监视信息 REST API。您可以提交或列出作业，包括对象导出类型、对象导入类型、资源导出类型和搜索导出类型的作业。
- 对象子计数 REST API。您可以列出某个对象的子资产总数。
- 产品信息 REST API。您可以列出有关 Enterprise Data Catalog 的详细信息，包括发行版本、内部版本和构建日期。

有关 REST API 的详细信息，请参阅《*Informatica 10.4 Enterprise Data Catalog REST API 引用*》。

Enterprise Data Preparation

本节介绍版本 10.4.0 中新增的 Enterprise Data Preparation 功能。

Data Lake 访问管理

从版本 10.4.0 开始，您可以管理对 Data Lake 的访问，方法是：为 Enterprise Data Preparation 用户和用户组指派对表示 Data Lake 中的 Hive 架构和 HDFS 位置的 Enterprise Data Catalog 资源的权限。

如果授予用户或用户组对特定架构或位置的权限，当用户执行导入、发布或上传操作时，应用程序将仅显示该用户有权访问的架构和位置。

有关详细信息，请参阅《*Enterprise Data Preparation 10.4.0 管理员指南*》。

Microsoft Azure Data Lake Storage 用作数据源

从版本 10.4.0 开始，您可以将 Microsoft Azure Data Lake Storage Gen2 位置作为数据源。

将文件发布到 Data Lake

从版本 10.4.0 开始，您可将文件中准备好的数据发布到 Data Lake 中的 HDFS 位置。

发布数据时，可以选择将数据写入 Data Lake 时使用的文件类型。例如，如果选择将数据作为逗号分隔值文件发布，应用程序会将数据作为 .csv 文件写入 Data Lake。

有关详细信息，请参阅《*Enterprise Data Preparation 10.4.0 用户指南*》。

将文件上传到 Data Lake

从版本 10.4.0 开始，可将若干个格式的文件从本地驱动器上载到 Data Lake。

可在以下上传选项中进行选择：

直接将文件上传到 Data Lake。

您可以不预览数据，将逗号分隔文件、Avro 文件、JSON 文件或 UTF-8 格式的 Parquet 文件直接从本地驱动器上传到 Data Lake。如果要在不预览数据的情况下上传文件，可以选择此选项。

允许 CLAIRE 确定文件结构，然后将文件上传到 Data Lake。

可将逗号分隔文件或 Microsoft Excel 电子表格的数据上传到 Data Lake。上传文件时，Enterprise Data Preparation 会使用 CLAIRE 嵌入式发现引擎确定文件结构并显示数据预览。

使用此选项上传 Excel 电子表格时，CLAIRE 引擎会发现电子表格中的工作表和表格。可以选择要预览的工作表和表格。

注意：上传 Excel 电子表格是预览功能。

定义文件结构，然后将文件上传到 Data Lake。

可将逗号分隔文件中的数据从本地驱动器上传到 Data Lake。上传文件时，可以预览数据、指定文件结构并配置列属性以符合您的需求。如果要在上传文件前修改列属性，可以选择此选项。

有关详细信息，请参阅《Enterprise Data Preparation 10.4.0 用户指南》。

Informatica 映射

本节介绍版本 10.4.0 中的 Informatica 映射的新功能。

将映射输出绑定至映射参数

从版本 10.4.0 开始，您可以保留并绑定部署为在本地环境或 Spark 引擎上运行的映射中的映射输出。

创建映射输出。将输出绑定到映射参数，以在映射的后续运行中使用该值。运行映射时，数据集成服务会将映射输出的值传递到映射参数。要保留映射输出，必须使用 `infacmd ms runMapping` 命令的 `-RuntimeInstanceName` 选项指定运行时实例名称。

Developer tool 现在在映射属性视图中包含了**绑定列**，以将映射输出绑定到参数。

有关已部署映射中的映射输出的详细信息，请参阅《Informatica 10.4.0 Developer 映射指南》中的“映射输出”一章。

infacmd ms 命令

下表介绍了新增和更新的 `infacmd ms` 命令：

命令	说明
<code>deleteMappingPersistedOutputs</code>	新命令，用于删除已部署映射的所有已保留映射输出。使用应用程序的名称和映射的运行时实例名称指定要删除的输出。要删除指定输出，请使用选项 <code>-OutputNamesToDelete</code> 。
<code>getMappingStatus</code>	更新的命令，现在可以返回作业名称。如果在 <code>infacmd ms runMapping</code> 中定义了运行时实例名称，作业名称就是运行时实例名称。
<code>listMappingPersistedOutputs</code>	新命令，用于列出已部署映射的已保留映射输出。基于应用程序的名称和映射的运行时实例名称列出输出。

有关详细信息，请参阅《Informatica 10.4.0 命令引用》中的“`infacmd ms` 命令引用”一章。

CLAIRE 建议和见解

从版本 10.4.0 开始，在开发期间您可以使用 CLAIRE 人工智能调整和更正映射。

启用建议后，CLAIRE 会在您开发映射时自动在映射中运行，并显示映射更正或调整建议。

您还可以对项目或项目文件夹中的映射运行 CLAIRE 分析。分析一组映射时，CLAIRE 会显示有关映射之间相似性的见解。

有关建议和见解的详细信息，请参阅 *Data Engineering Integration User Guide*。

更新映射优化器级别

从版本 10.4.0 开始，可以使用 `infacmd ms UpdateOptimizationDefaultLevel` 命令将应用程序中映射的优化器级别设置为默认级别，即“自动”。

运行此命令时，必须指定应用程序的名称。UpdateOptimizationDefaultLevel 可以设置应用程序中所有映射的优化器级别。

有关详细信息，请参阅《*Informatica 10.4.0 命令引用*》和《*Informatica 10.4.0 Developer 映射指南*》。

Informatica 转换

本节介绍版本 10.4.0 中新增的 Informatica 转换功能。

地址验证器转换

本节介绍地址验证器转换的新增功能。

对于以下国家，地址验证器转换具有附加的地址功能：

美国

从版本 10.4 开始，地址验证器会将美国地址中的 MC 识别为 MSC（Mail Stop Code，区域邮编）。

有关版本 10.4 中的地址验证软件引擎的功能和操作的全面信息，请参阅《*Informatica Address Verification 5.15.0 开发人员指南*》。

Informatica 工作流

本节介绍版本 10.4.0 中新增的 Informatica 工作流功能。

Amazon EMR 创建群集任务高级属性

从版本 10.4.0 开始，在工作流中为“创建群集”任务配置 Amazon EMR 连接时，新增了以下高级属性。

在“创建群集”任务中，可以配置以下功能：

- 根设备 EBS 卷大小。EBS 根设备卷的 GB 数。
- 自定义 AMI ID。自定义 Amazon Linux Amazon Machine Image (AMI) 的 ID。

- 安全配置。群集中身份验证和加密安全配置的名称。

有关详细信息，请参阅《*Data Engineering Integration 10.4.0 用户指南*》和《*Informatica® 10.4.0 Developer 工作流指南*》。

Informatica 安装

本节介绍版本 10.4.0 中新增的安装功能。

PostgreSQL

从版本 10.4.0 开始，您可以将 PostgreSQL 数据库用于域配置存储库、模型存储库和 PowerCenter 存储库。对于 Enterprise Data Preparation，您只能在其他模型存储库服务上使用 PostgreSQL 数据库。

您还可以安装 psql 客户端应用程序版本 10.6，以在 Linux 或 Windows 中使用 PostgreSQL。

有关 PostgreSQL 的详细信息，请参阅《*Informatica 10.4.0 安装指南*》。

静默模式下的预安装 (i10Pi) 系统检查工具

从版本 10.4.0 开始，您可在静默模式下运行预安装 (i10pi) 系统检查工具。您可在静默模式下运行 i10pi，以验证计算机是否满足在没有用户互动的情况下执行安装的系统要求。

有关如何在静默模式下运行 i10Pi 的详细信息，请参阅《*Informatica 10.4.0 安装指南*》。

对静默安装属性文件的密码进行加密

从版本 10.4.0 开始，Informatica 安装程序中包括了一个可对您在属性文件（其中包含了在静默模式下安装服务所需的选项）中设置的密码进行加密的实用程序。

当您在静默模式下运行安装程序时，安装框架会对加密的密码进行解密。

有关详细信息，请参阅《*Informatica 安装和配置指南*》。

智能结构模型

本节介绍版本 10.4.0 中新增的智能结构模型功能。

其他输入类型

从版本 10.4.0 开始，Intelligent Structure Discovery 可处理 ORC、Avro 和 Parquet 输入。

有关详细信息，请参阅《*Data Engineering Integration 10.4.0 用户指南*》。

在设计时从示例创建模型

从版本 10.4.0 开始，Intelligent Structure Discovery 可以根据您在创建复杂数据对象时选择的示例文件创建模型。然后，您可以在 Informatica Intelligent Cloud Services 数据集成中优化该模型。

XML、JSON、ORC、AVRO 和 Parquet 示例文件支持此功能。

有关详细信息，请参阅《*Data Engineering Integration 10.4.0 用户指南*》。

处理无法识别的数据

从 10.4.0 开始，Intelligent Structure Discovery 将使用结构化 JSON 格式处理模型中无法识别数据的输出。

有关详细信息，请参阅《*Data Engineering Integration 10.4.0 用户指南*》。

PowerCenter

本节介绍版本 10.4.0 中新增的 PowerCenter 功能。

HTTP 转换

从版本 10.4.0 开始，HTTP 转换还包括以下用于最终 URL 构造的方法：SIMPLE PATCH、SIMPLE PUT 和 SIMPLE DELETE。

通过 SIMPLE PATCH 方法，您可以执行部分更新，输入数据不必是完整正文。您可以用它从输入端口将数据作为修补程序更新到资源。

使用 SIMPLE PUT 方法可以完全替换文档。您可以将数据作为单块数据从输入端口创建到 HTTP 服务器。如果此数据已存在，可将数据作为单块数据从输入端口更新到 HTTP 服务器。

使用 SIMPLE DELETE 方法可从 HTTP 服务器中删除数据。

您还可以参数化用于 HTTP 转换的基本 URL。

在以前的版本中，只能为以下两种方法指定最终 URL 构造：SIMPLE GET 和 SIMPLE POST。并且，您无法参数化用于 HTTP 转换的最终 URL。

有关详细信息，请参阅《*PowerCenter 10.4.0 转换指南*》中的“HTTP 转换”一章。

连接

本节介绍版本 10.4.0 版本中新增的连接功能。

Oracle 连接的用户模拟

从版本 10.4.0 开始，您可以在 PowerCenter 的 Oracle 连接中指定模拟用户名，以连接至 Oracle 源和目标。指定模拟用户时，可以代表该用户连接至 Oracle 并执行读取或写入操作。

有关详细信息，请参阅《*Informatica 10.4.0 PowerCenter workflow 基础指南*》。

支持 SAP HANA 数据库建模视图

从版本 10.4.0 开始，您可以从以下类型的 SAP HANA 数据库建模视图中读取数据：

- 分析视图
- 属性视图
- 计算视图

有关详细信息，请参阅《*Informatica 10.4.0 PowerCenter Designer 指南*》。

PowerExchange 适配器

本节介绍版本 10.4.0 中新增的 PowerExchange 适配器功能。

Informatica PowerExchange 适配器

本节介绍版本 10.4.0 中新增的 Informatica 适配器功能。

PowerExchange for Amazon Redshift

从版本 10.4.0 开始，PowerExchange for Amazon Redshift 包含以下功能：

- 可以在 AWS Databricks 环境中运行映射。
- 即使您在 JDBC URL 连接属性中指定了群集区域名称，也可以从“群集区域”连接属性中选择群集区域名称。
- 从 Amazon Redshift 读取数据时，可以保留空值。
- 在向 Amazon Redshift 写入数据时，可以指定每个批处理的暂存文件数。
- 在将 CDC 源中的数据写入 Amazon Redshift 目标时，您可以保留记录顺序。

有关详细信息，请参阅《*Informatica 10.4.0 PowerExchange for Amazon Redshift 用户指南*》。

PowerExchange for Amazon S3

从版本 10.4.0 开始，PowerExchange for Amazon S3 包含以下功能：

- 可以在 AWS Databricks 环境中运行映射。
- 可以使用临时安全凭据通过 AssumeRole 访问 AWS 资源。
- 您可在运行时在读取和写入操作属性中参数化数据格式类型和架构。

有关详细信息，请参阅《*Informatica 10.4.0 PowerExchange for Amazon S3 用户指南*》。

PowerExchange for Google BigQuery

从版本 10.4.0 开始，PowerExchange for Google BigQuery 包含以下功能：

- 可以使用 Google Dataproc 群集在 Spark 引擎上运行映射。
- 可以通过在优化的 Spark 模式下运行映射来提高映射的性能。使用优化的 Spark 模式读取数据时，可以指定要使用的分区数。您可在高级读取和写入操作属性中指定是在**常规**还是**优化**模式下运行映射。优化的 Spark 模式可提高映射性能。
- 还可以配置 SQL 替代，以替代用于从 Google BigQuery 源提取数据的默认 SQL 查询。
- 您可以在 Google BigQuery 中读取或写入 NUMERIC 数据类型的数据。NUMERIC 数据类型是精确的数值，有 38 位精度位数和 9 个小数位数。读取或写入 NUMERIC 数据类型时，数据集成服务会将 NUMERIC 数据类型映射到 Decimal 转换数据类型，允许的精度位数最多 38 位、小数位数最多 9 位。

有关详细信息，请参阅《*Informatica 10.4.0 PowerExchange for Google BigQuery 用户指南*》。

PowerExchange for Google Cloud Storage

从版本 10.4.0 开始，PowerExchange for Google Cloud Storage 包含以下功能：

- 可以使用 Google Dataproc 群集在 Spark 引擎上运行映射。

- 在从 Google Cloud Storage 源读取数据时，可以配置以下 Google Cloud Storage 数据对象读取操作高级属性：

Google Cloud Storage 路径

覆盖您在 Google Cloud Storage 数据对象中选择的文件的 Google Cloud Storage 路径。
使用以下格式：

gs://<bucket name> 或 gs://<bucket name>/<folder name>

源文件名

覆盖在 Google Cloud Storage 数据对象中指定的 Google Cloud Storage 源文件名。

是目录

读取在 **Google Cloud Storage 路径** 数据对象读取操作高级属性中指定的文件夹中的所有可用文件。

有关详细信息，请参阅《*Informatica 10.4.0 PowerExchange for Google Cloud Storage 用户指南*》。

PowerExchange for Microsoft Azure Blob 存储

从版本 10.4.0 开始，PowerExchange for Microsoft Azure Blob 存储包含以下功能：

- 您可在运行时在读取和写入操作属性中参数化数据格式类型和架构。
- 在创建 Microsoft Azure Blob 存储连接时，可以使用共享访问签名身份验证。

有关详细信息，请参阅《*Informatica 10.4.0 PowerExchange for Microsoft Azure Blob 存储用户指南*》。

PowerExchange for Microsoft Azure SQL 数据仓库

从版本 10.4.0 开始，您可以在位于虚拟网络 (VNet) 内的 Microsoft Azure SQL Data Warehouse 端点上读取或写入数据。

有关详细信息，请参阅《*Informatica 10.4.0 PowerExchange for Microsoft Azure SQL Data Warehouse 用户指南*》。

PowerExchange for Salesforce

从版本 10.4.0 开始，PowerExchange for Salesforce 包括以下功能：

- 您可以使用版本 45.0、46.0 和 47.0 的 Salesforce API 创建 Salesforce 连接以及访问 Salesforce 对象。
- 对于代表父对象上共享条目的共享对象上的查询，可以启用主键分块。只有在父对象受支持的情况下，共享对象才支持主键分块。例如，如果要对 CaseHistory 进行查询，则父对象 Case 必须支持主键分块。
- 您可以创建分配规则，以在使用标准 API 插入、更新或更新插入 Lead 和 Case 目标对象的记录时重新分配记录中的属性。

有关详细信息，请参阅《*Informatica 10.4.0 PowerExchange for Salesforce 用户指南*》。

PowerExchange for SAP NetWeaver

从版本 10.4.0 开始，PowerExchange for SAP NetWeaver 包含以下功能：

- 您可以配置用于 SAP 表读取器映射的 HTTPS 流传输。
- 如果您的 SAP NetWeaver 系统版本为 7.50 或更高版本，可以使用 SAP 表读取器从 ABAP CDS 视图读取数据。
- 您可以从具有以下数据类型的字段的 SAP 表中读取数据：
 - DF16_DEC

- DF32_DEC
- DF16_RAW
- DF34_RAW
- INT8
- RAWSTRING
- SSTRING
- STRING

有关详细信息，请参阅《*Informatica PowerExchange for SAP NetWeaver 10.4.0 用户指南*》。

PowerExchange for Snowflake

从版本 10.4.0 开始，PowerExchange for Snowflake 包含以下功能：

- 您可以在 Databricks 环境中运行 Snowflake 映射。
- 可以使用 Snowflake 对象作为映射中的动态源和目标。
- 您可以使用**创建目标**选项来创建 Snowflake 目标。
- 可为映射中的 Snowflake 目标配置目标架构策略。您可从可用选项中选择是保留现有目标架构还是创建目标（如果不存在）。您还可以指定目标架构策略选项作为参数值。
- 您可以在 Snowflake 高级目标属性中指定拒绝项文件名和路径，数据集成服务将用其写入在写入目标时被拒绝的记录。
- 当 Snowflake ODBC 连接中的 ODBC 提供程序类型为 Snowflake 时，您可以配置下推优化以将转换逻辑推送到 Snowflake 数据库。
- 通过在数据集成服务属性中配置 EnableSDKDecimal38 自定义标志，可以读取或写入精度和小数位数为 38 位的 Decimal 数据类型的数据。

有关详细信息，请参阅《*Informatica 10.4.0 PowerExchange for Snowflake 用户指南*》。

PowerExchange for HDFS

从版本 10.4.0 开始，PowerExchange for HDFS 包含以下功能：

- 您可在运行时在读取和写入操作属性中参数化数据格式类型和架构。
- 您可设置读取或写入操作的复杂文件类型对象的架构格式。

有关详细信息，请参阅《*Informatica 10.4.0 PowerExchange for HDFS 用户指南*》。

PowerCenter PowerExchange 适配器

本节介绍版本 10.4.0 中新增的 PowerCenter 适配器功能。

PowerExchange for Google BigQuery

从版本 10.4.0 开始，您可以在 Google BigQuery 中读取或写入 NUMERIC 数据类型的数据。NUMERIC 数据类型是精确的数值，有 38 位精度位数和 9 个小数位数。读取或写入 NUMERIC 数据类型时，PowerCenter Integration Service 会将 NUMERIC 数据类型映射到 Decimal 转换数据类型，允许的精度位数最多 28 位、小数位数最多 9 位。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica 10.4.0 PowerExchange for Google BigQuery 用户指南*》。

PowerExchange for Google Cloud Storage

从版本 10.4.0 开始，在从 Google Cloud Storage 源读取数据时，可以配置以下 Google Cloud Storage 数据对象读取操作高级属性：

Google Cloud Storage 路径

覆盖您在 Google Cloud Storage 数据对象中选择的文件的 Google Cloud Storage 路径。

使用以下格式：

gs://<bucket name> 或 gs://<bucket name>/<folder name>

源文件名

覆盖在 Google Cloud Storage 数据对象中指定的 Google Cloud Storage 源文件名。

是目录

读取在 **Google Cloud Storage 路径** 数据对象读取操作高级属性中指定的文件夹中的所有可用文件。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica 10.4.0 PowerExchange for Google Cloud Storage 用户指南》。

PowerExchange for Greenplum

从版本 10.4.0 开始，您可以使用 PowerExchange for Greenplum 从 Greenplum 读取数据。您可以为 Greenplum 源配置特定的会话属性，以确定如何从 Greenplum 提取数据。

运行 Greenplum 会话以读取数据时，PowerCenter 集成服务会调用 Greenplum 数据库并行文件服务器 gpfdist（Greenplum 的文件分发程序）来读取数据。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica 10.4.0 PowerExchange for Greenplum 用户指南》。

PowerExchange for JD Edwards EnterpriseOne

从版本 10.4.0 开始，可以使用版本 9.2 的 JD Edwards EnterpriseOne API 来创建 JD Edwards EnterpriseOne 连接以及访问 JD Edwards EnterpriseOne 对象。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica 10.4.0 PowerExchange for JD Edwards EnterpriseOne 用户指南》。

PowerExchange for Kafka

从版本 10.4.0 开始，您可以配置以下 SSL 属性以启用到 Kafka Broker 的安全连接：

- SSL 模式
- SSL TrustStore 文件路径
- SSL TrustStore 密码
- SSL 密钥库文件路径
- SSL 密钥库密码

您可以配置 Kafka 消息传递代理以使用 Kafka Broker 版本 0.10.1.1 及更高版本。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Kafka 10.4.0 用户指南》。

PowerExchange for Salesforce

从版本 10.4.0 开始，可以使用版本 46.0 和 47.0 的 Salesforce API 来创建 Salesforce 连接以及访问 Salesforce 对象。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica 10.4.0 PowerExchange for Salesforce 用户指南*》。

PowerExchange for SAP NetWeaver

从版本 10.4.0 开始，您可以使用 PowerExchange for SAP 动态 ABAP 表提取程序通过 HTTP/HTTPS 流传输从 SAP 表和 ABAP Core Data Services (CDS) 视图读取数据。如果您的 SAP NetWeaver 系统版本为 7.50 或更高版本，您可以使用 PowerExchange for SAP 动态 ABAP 表提取程序从 ABAP CDS 视图读取数据。

有关详细信息，请参阅《*Informatica PowerExchange for SAP NetWeaver 10.4.0 用户指南*》。

安全

本节介绍版本 10.4.0 中的新安全功能。

配置 Web 应用程序以使用不同的 SAML 身份提供程序

从版本 10.4.0 开始，您可以将域中运行的 Informatica Web 应用程序配置为使用不同的安全断言标记语言 (SAML) 身份提供程序。例如，您可以将 Informatica Administrator 配置为以 Active Directory Federation Services 作为身份提供程序，并将 Informatica Analyst 配置为以 PingFederate 作为身份提供程序。

有关详细信息，请参阅《*Informatica 10.4.0 安全指南*》。

第 2 章

新增功能 (10.2.2 HotFix 1)

本章包括以下主题：

- [命令行程序, 34](#)
- [Enterprise Data Catalog, 34](#)

命令行程序

本节介绍版本 10.2.2 HotFix 1 中新增的命令。

infacmd ldm 命令

下表介绍了新的 infacmd ldm 命令选项：

新选项	说明
-Force -fr	可选。如果要在备份模式脱机时强制备份。强制执行备份并覆盖现有备份。
-Force -fr	可选。如果您想清除 HDFS 和 Apache Zookeeper 中的现有内容。强制还原备份数据。

有关详细信息，请参阅《*Informatica 10.2.2 HotFix 1 命令引用*》中的“infacmd ldm 命令引用”一章。

Enterprise Data Catalog

本节介绍版本 10.2.2 HotFix 1 中新增的 Enterprise Data Catalog 功能。

Azure Data Lake Storage Gen2

从版本 10.2.2 HotFix 1 开始，您可以从 Azure Data Lake Storage Gen2 源中提取元数据。您可以访问 Azure Data Lake Storage Gen2 存储帐户中文件和文件夹的元数据并将其提取到 Enterprise Data Catalog。

有关详细信息，请参阅《*Informatica 10.2.2 HotFix1 Enterprise Data Catalog 扫描程序配置指南*》中的“Azure Data Lake Store”一章。

链接不区分大小写

从版本 10.2.2 HotFix 1 开始，您可以将资源配置为不区分大小写。借助此选项，您可以查看源和目标之间的链接沿袭，而不区分资源名称的大小写。

有关链接不区分大小写的信息，请参阅《*Informatica 10.2.2 HotFix 1 Catalog 管理员指南*》中的“管理资源”一章。

Enterprise Data Catalog Tableau Extension

从版本 10.2.2 HotFix 1 开始，Enterprise Data Catalog Tableau Extension 是 Tableau 仪表板的本机扩展，可在 Tableau 中使用。

可以在 Tableau Desktop、Tableau Server 以及 Tableau 支持的所有 Web 浏览器中使用 Enterprise Data Catalog Tableau Extension。从 Enterprise Data Catalog 应用程序下载此扩展，然后将其添加到 Tableau 的仪表板中。

有关此扩展的详细信息，请参阅《*适用于 Tableau 的 Informatica 10.2.2 HotFix1 Enterprise Data Catalog Extension*》指南。

新资源

从版本 10.2.2 HotFix 1 开始，Enterprise Data Catalog 中添加了以下新资源：

- SAP PowerDesigner。可以从 SAP PowerDesigner 数据源提取元数据、关系和沿袭信息。
- SAP HANA。可以从 SAP HANA 数据库提取对象和沿袭元数据。

有关详细信息，请参阅《*Informatica 10.2.2 HotFix1 扫描程序配置指南*》。

剖析仓库扫描程序

从版本 10.2.2 HotFix 1 开始，您可以创建一个或多个剖析仓库资源。可以关联剖析仓库并在资源中选择架构。运行资源时，扫描程序将从关联的剖析仓库中提取所选架构的剖析结果并将其迁移到目录中。

有关详细信息，请参阅《*Informatica 10.2.2 HotFix1 Enterprise Data Catalog 扫描程序配置指南*》中的“配置 Informatica Platform 扫描程序”一章。

REST API

从版本 10.2.2 HotFix 1 开始，您可以使用以下 Informatica Enterprise Data Catalog REST API：

- 数据置备 REST API。可以还原、更新或删除连接和资源。
- 目录模型 REST API。除了现有 REST API 之外，您还可以访问、更新或删除字段面、查询面和搜索选项卡。
- 对象 API。除了现有 REST API 之外，您还可以列出目录搜索和建议。

有关 REST API 的详细信息，请参阅《*Informatica 10.2.2 HotFix 1 Enterprise Data Catalog REST API 参考*》。

搜索增强功能

从版本 10.2.2 HotFix 1 开始，您可以利用以下增强功能，在 Enterprise Data Catalog 中执行更强大的搜索：

双引号搜索

可以使用双引号 (") 搜索资产，以查找与双引号内的资产名称完全匹配的资产，但不能查找目录中资产名称的变体。

搜索运算符

可以使用更新的搜索运算符获得更准确的搜索结果。这些搜索运算符包括 AND、OR、NOT、title 和 description。

搜索等级

Enterprise Data Catalog 使用等级算法对搜索结果页面中的数据资产进行等级排序。搜索等级是指与特定搜索结果中的其他资产相比，资产的优先级。

相关搜索

可以启用**搜索结果**页面中的**显示相关搜索**选项来查看相关资产。

有关搜索增强功能的详细信息，请参阅《*Informatica 10.2.2 HotFix 1 Enterprise Data Catalog 用户指南*》中的“搜索资产”一章。

搜索标签

从版本 10.2.2 HotFix 1 开始，您可以使用搜索标签搜索资产，而不必在执行资产搜索时重复设置相同的搜索条件。搜索标签是目录中的预定义筛选器。

有关搜索标签的详细信息，请参阅《*Informatica 10.2.2 HotFix 1 Enterprise Data Catalog 用户指南*》中的“自定义搜索”一章。

独立扫描实用程序支持的资源类型

从版本 10.2.2 HotFix 1 开始，无法在运行时或脱机情况下访问元数据时，可以从以下外部源提取元数据：

- Apache Atlas
- Cloudera Navigator
- 文件系统
- HDFS
- Hive
- Informatica Platform
- MicroStrategy
- OneDrive
- Oracle Business Intelligence
- SharePoint
- Sybase
- Tableau

有关详细信息，请参阅《*Informatica 10.2.2 HotFix1 Enterprise Data Catalog 管理员指南*》中的“从脱机和无法访问的资源提取元数据”一章。

技术预览版

Enterprise Data Catalog 版本 10.2.2 HotFix 1 包含可供技术预览版使用的功能。

支持技术预览版功能，但不对此做出保证且不能用于生产。Informatica 建议您只在非生产环境中使用。

Informatica 打算将预览功能包含在即将推出的 GA 版本中以供生产使用，但也可能根据不断变化的市场或技术环境而选择不这样做。有关详细信息，请联系 Informatica 全球客户支持部门。

从版本 10.2.2 HotFix 1 开始，以下功能可供技术预览版使用：

- 从版本 10.2.2 HotFix 1 开始，您可以在包括来自 Oracle 数据集成器数据源的转换逻辑的列级别提取数据沿袭的元数据。
- 从版本 10.2.2 HotFix 1 开始，您可以在包括来自 IBM InfoSphere DataStage 数据源的转换逻辑的列级别提取数据沿袭的元数据。
- 从版本 10.2.2 HotFix 1 开始，您可以在用于 Oracle 和 SQL Server 中存储过程的列级别提取数据沿袭。
- 从版本 10.2.2 HotFix 1 开始，在目录中完成数据发现后，您可以执行数据置备。数据置备可帮助您将数据移动到目标，以进行进一步分析。
有关预览数据的详细信息，请参阅《*Informatica 10.2.2 HotFix 1 Catalog 管理员指南*》和《*Informatica 10.2.2 Hotfix 1 Enterprise Data Catalog 用户指南*》。
- 从版本 10.2.2 HotFix 1 开始，将数据移动到目标之前，您可以预览数据以对其进行评估。只能对 Oracle 和 Microsoft SQL Server 资源中的表格式资产进行数据预览。
有关预览数据的详细信息，请参阅《*Informatica 10.2.2 HotFix 1 Catalog 管理员指南*》和《*Informatica 10.2.2 Hotfix 1 Enterprise Data Catalog 用户指南*》。

第 3 章

新增功能 (10.2.2 Service Pack 1)

本章包括以下主题：

- [Big Data Management, 38](#)
- [大数据 Streaming, 39](#)
- [Enterprise Data Catalog, 39](#)
- [Enterprise Data Preparation, 40](#)
- [Informatica PowerExchange 适配器, 40](#)

Big Data Management

本节介绍版本 10.2.2 Service Pack 1 中新增的 Big Data Management 功能。

Sqoop

从版本 10.2.2 Service Pack 1 开始，可以使用以下新的 Sqoop 功能：

连接到启用了 SSL 的 Oracle 和 Microsoft SQL Server 数据库

可以将 Sqoop 配置为连接到安全关系数据库，例如 Oracle 和 Microsoft SQL Server。

在 JDBC 连接中使用适当的 JDBC 连接字符串和连接参数连接到启用了 SSL 的 Oracle 或 Microsoft SQL Server 数据库。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.2 Service Pack 1 用户指南*》。

支持 connection-param-file Sqoop 参数

可以定义 connection-param-file 参数以连接到启用了 SSL 的 Oracle 数据库。connection-param-file 是一个属性文件，可指定额外的 JDBC 参数，Sqoop 必须使用这些参数才能连接到启用了 SSL 的数据库。

创建连接时，此文件的内容会解析为标准 Java 属性并传递到驱动程序中。

可以在 JDBC 连接的 **Sqoop 参数** 字段中指定 connection-param-file 参数。

请使用以下语法：

```
--connection-param-file <参数文件名>
```

有关详细信息，请参阅《*Informatica Big Data Management 10.2.2 Service Pack 1 用户指南*》。

大数据 Streaming

本节介绍版本 10.2.2 Service Pack 1 中新增的 Big Data Streaming 功能。

Amazon S3 目标

从版本 10.2.2 Service Pack 1 开始，可以创建一个流映射来将数据写入 Amazon S3。

创建一个 Amazon S3 数据对象，以将数据写入 Amazon S3。可以创建一个 Amazon S3 连接，使用 Amazon S3 作为目标。可以通过 Developer tool 或 infacmd 创建和管理 Amazon S3 连接。

有关详细信息，请参阅《*Informatica Big Data Streaming 10.2.2 Service Pack 1 用户指南*》。

TIME_RANGE 函数

从版本 10.2.2 Service Pack 1 开始，可以在连接器转换中使用 TIME_RANGE 函数，该函数可确定要联接的各个流事件的时间范围。

TIME_RANGE 函数只适用于流映射中的连接器转换。

语法

```
TIME_RANGE(EventTime1,EventTime2,Format,Interval)
```

有关 TIME_RANGE 函数的详细信息，请参阅《*Informatica 10.2.2 Service Pack 1 转换语言参考*》指南。

将数据写入多个 HDFS 文件

从版本 10.2.2 Service Pack 1 开始，可以使用一个复杂文件数据对象将源数据写入多个 HDFS 文件。根据复杂文件数据对象的 FileName 标头端口值，数据集成服务会在运行时在目标位置创建多个 HDFS 文件。

有关详细信息，请参阅《*Informatica Big Data Streaming 10.2.2 Service Pack 1 用户指南*》。

Enterprise Data Catalog

本节介绍版本 10.2.2 Service Pack 1 中新增的 Enterprise Data Catalog 功能。

独立扫描仪实用程序支持的资源类型

从版本 10.2.2 Service Pack 1 开始，无法在运行时或脱机获得元数据时，可以从以下外部源提取元数据：

- IBM DB2
- IBM DB2 for z/OS
- IBM Netezza
- JDBC
- PowerCenter
- SQL Server Integration Services

有关详细信息，请参阅《*Informatica 10.2.2 Service Pack 1 Enterprise Data Catalog 管理员指南*》中的“从脱机和无法访问的资源提取元数据”一章。

Enterprise Data Preparation

本节介绍版本 10.2.2 Service Pack 1 中新增的 Enterprise Data Preparation 功能。

还原所有已推理数据类型

从版本 10.2.2 Service Pack 1 开始中，可以将在一个工作表中将应用于源列的所有已推理类型和数据域还原为其原始类型。

有关详细信息，请参阅《*Informatica 10.2.2 Service Pack 1 Enterprise Data Preparation 用户指南*》中的“准备数据”一章。

准备 Avro 和 Parquet 文件

从版本 10.2.2 Service Pack 1 开始，在数据准备过程的第一步将 Avro 和 Parquet 文件添加到项目中后，可以对文件中的层次结构数据进行采样。Enterprise Data Preparation 会将 Avro 或 Parquet 文件结构转换为平面结构并在工作表中提供数据，您可以使用该工作表来准备数据。

有关详细信息，请参阅《*Informatica 10.2.2 Service Pack 1 Enterprise Data Preparation 用户指南*》中的“准备数据”一章。

Informatica PowerExchange 适配器

本节介绍版本 10.2.2 Service Pack 1 中新增的 Informatica 适配器功能。

PowerExchange for Hive

从版本 10.2.2 Service Pack 1 开始中，导入 Hive 表时，可以使用**仅显示默认架构**选项清除搜索结果，只显示使用默认架构的表，或搜索使用非默认架构的表。

有关详细信息，请参阅《*Informatica PowerExchange for Hive 10.2.2 Service Pack 1 用户指南*》。

第 4 章

新增功能 (10.2.2)

本章包括以下主题：

- [应用程序服务, 41](#)
- [Big Data Management, 42](#)
- [大数据 Streaming, 46](#)
- [命令行程序, 47](#)
- [Enterprise Data Catalog, 51](#)
- [Enterprise Data Lake, 53](#)
- [Informatica Developer, 58](#)
- [Informatica 映射, 58](#)
- [Informatica 转换, 59](#)
- [Informatica PowerExchange 适配器, 62](#)

应用程序服务

本节介绍版本 10.2.2 中新增的应用程序服务功能。

Mass Ingestion 服务

从版本 10.2.2 开始，可以在 Mass Ingestion 服务进程和外部组件之间启用安全通信。您可以为 HTTPS 协议指定 HTTPS 端口号，为 SSL 证书指定密钥库文件。

有关详细信息，请参阅《*Informatica 10.2.2 应用程序服务指南*》中的“Mass Ingestion 服务”一章。

元数据访问服务

从版本 10.2.2 开始，可以将元数据访问服务配置为使用操作系统配置文件。从 Hadoop 群集导入和预览元数据时，使用操作系统配置文件既可增加安全性，又可隔离设计时用户环境。

有关详细信息，请参阅《*Informatica 10.2.2 安全指南*》中的“用户和组”一章。

REST 操作 Hub 服务

从版本 10.2.2 开始，可以为 REST 应用程序配置 REST 操作 Hub 服务。REST 操作 Hub 服务是 Informatica 域中的 REST 系统服务，可通过 REST API 向外部客户端公开 Informatica 产品功能。

您可以通过 Administrator 工具或通过 infacmd 来配置 REST 操作 Hub 服务。可以使用 REST 操作 Hub 服务查看应用程序中部署的映射作业的映射执行统计信息。

可以使用 REST 操作 Hub 服务获取在数据集成服务上或 Hadoop 环境中运行的大数据映射的映射执行统计信息。

有关 REST API 的详细信息，请参阅《*Big Data Management 10.2.2 管理员指南*》。

Big Data Management

本节介绍版本 10.2.2 中新增的 Big Data Management 功能。

Azure Databricks 集成

从版本 10.2.2 开始，可以将 Informatica 域与 Azure Databricks 环境相集成。

Azure Databricks 是针对 Microsoft Azure 云服务优化的分析云平台，其中集成了开源 Apache Spark 群集技术和功能。

可以在 Azure VM 上或本地安装 Informatica 域。该集成过程类似于与 Hadoop 环境的集成。您执行集成任务，包括从 Databricks 环境导入群集配置。Informatica 域使用标志身份验证来访问 Databricks 环境。Databricks 标志 ID 存储在 Databricks 连接中。

源和目标

可以在 Databricks 环境中针对以下源和目标运行映射：

- Microsoft Azure Data Lake Store
- Microsoft Azure Blob 存储
- Microsoft Azure SQL 数据仓库
- Microsoft Azure Cosmos DB

转换

您可以将以下转换添加到 Databricks 映射：

- 汇总器
- 表达式
- 筛选器
- 联接器
- 查找
- 规范器
- 等级
- 路由器
- 排序器
- 联合

Databricks Spark 引擎对转换的处理方式与 Spark 引擎在 Hadoop 环境中的处理方式大致相同。

数据类型

支持以下数据类型：

数组
长整型
日期/时间
小数
双精度型
整型
映射
结构
文本
字符串

映射

配置映射时，可以选择在 Databricks 环境中验证并运行映射。运行映射时，数据集成服务会生成 Scala 代码并将其传递给 Databricks Spark 引擎。

工作流

您可以开发群集工作流以在 Databricks 环境中创建临时群集。

有关详细信息，请参阅以下指南：

《*Big Data Management 10.2.2 集成指南*》
《*Big Data Management 10.2.2 管理员指南*》
《*Big Data Management 10.2.2 用户指南*》

层次结构数据

本节介绍版本 10.2.2 中的层次结构数据的新功能。

动态复杂端口

从版本 10.2.2 开始，您可以将动态复杂端口添加到在 Spark 引擎上运行的动态映射中。使用动态复杂端口可管理复杂文件中层次结构数据的常见架构更改。

动态复杂端口根据运行时的架构更改接收复杂端口的新元素或已更改元素。输入规则决定了动态复杂端口的元素。根据输入规则，动态复杂端口接收来自上游转换的复杂端口的一个或多个元素。您可以在 Spark 引擎上的某些转换中使用动态复杂端口，如动态数组、动态映射和动态结构。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.2 用户指南*》中的“处理更改了架构的层次结构数据”一章。

层次结构数据预览

从版本 10.2.2 开始，您可以在 Developer tool 中预览映射中的层次结构数据。预览层次结构数据有助于设计和调试大数据映射。

可以选择源和转换作为映射中的预览点，其中包含以下层次结构类型：

- 数组
- 结构
- 映射

层次结构数据预览可用于技术预览版。支持技术预览版功能，但不对此做出保证且尚未做好生产准备。Informatica 建议您只在非生产环境中使用这些功能。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.2 用户指南*》。

高可用性

本节介绍版本 10.2.2 中新增的高可用性功能。

大数据作业恢复

从版本 10.2.2 开始，当数据集成服务节点意外停止时，数据集成服务可以恢复配置为在 Spark 引擎上运行的大数据作业。如果数据集成服务节点在作业完成之前发生故障，数据集成服务则将作业发送至另一节点，该节点将从发生节点故障的时间点继续处理作业任务。

要恢复大数据映射，必须在数据集成服务属性中启用大数据作业恢复，然后从 infacmd 运行作业。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.2 管理员指南*》中的“数据集成服务处理”一章。

分布式数据集成服务队列

从版本 10.2.2 开始，当为部署的大数据作业启用大数据恢复时，数据集成服务使用分布式队列来存储作业信息。分布式队列存储在模型存储库中，当资源可用时，任何可用的数据集成服务均可运行队列中的作业。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.2 管理员指南*》中的“数据集成服务处理”一章。

智能结构模型

本节介绍版本 10.2.2 中新增的智能结构模型功能。

XML 文件中的别名

从版本 10.2.2 开始，Intelligent Structure Discovery 可以处理使用不同别名来标识同一命名空间的 XML 文件，用法与在创建智能结构模型时使用的 XML 文件中一样。

数据类型

从版本 10.2.2 开始，自冬季 2019 年 3 月版的 Informatica 智能云服务起，当复杂文件读取器使用智能结构模型时，Intelligent Structure Discovery 会将数据类型传递至输出数据端口。

例如，当 Intelligent Structure Discovery 检测到字段包含日期时，它会将数据作为日期而非字符串传递至输出数据端口。

字段名称

从版本 10.2.2 开始，自冬季 2019 年 3 月版的 Informatica 智能云服务起，从智能结构模型导入的复杂文件数据对象中的字段名称能够以数字和预留字开头，且可包含以下特殊字符：`|.[]{}()*+-?.^$|`

字段以数字或预留字开头时，Big Data Management 映射会在字段名称开头添加下划线 (`_`)。例如，如果智能结构模型中的某个字段以 `OR` 开头，映射则将字段导入为 `_OR`。当字段名称包含某个特殊字符时，映射会将该字符转换为下划线。

处理大型 XML 文件

从版本 10.2.2 开始，Intelligent Structure Discovery 可以对 XML 文件进行流式传输，并处理块中重复元素的数据。这可提高处理大型 XML 文件的效率。

数据偏移

从版本 10.2.2 开始，自冬季 2019 年 3 月版的 Informatica 智能云服务起，Intelligent Structure Discovery 增强了对数据偏移的处理。

在 Intelligent Structure Discovery 中，当输入数据含有示例文件中没有的字段时，便会出现数据偏移。这种情况下，Intelligent Structure Discovery 会将未定义数据传递到目标上的未分配数据端口，而不是丢弃这些数据。

Mass Ingestion

从版本 10.2.2 开始，可以运行增量加载以获取增量数据。运行增量加载时，Spark 引擎根据时间戳或 ID 列提取增量数据，并将增量数据加载到 Hive 或 HDFS 目标。如果将数据获取到 Hive 目标，Spark 引擎还会传播已对源表所进行的架构更改。

如果获取增量数据，Mass Ingestion 服务将使用 Sqoop 的增量导入模式。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.2 Mass Ingestion 指南*》。

监视

本节介绍版本 10.2.2 中与 Big Data Management 监视相关的新功能。

Spark 监视

从版本 10.2.2 开始，可以在“摘要统计信息”窗格中查看作业前和作业后任务，以执行 Spark 监视。

有关作业前和作业后任务的详细信息，请参阅《*Informatica Big Data Management 10.2.2 用户指南*》。

安全

本节介绍版本 10.2.2 中与 Big Data Management 安全相关的新功能。

Enterprise 安全包

从版本 10.2.2 开始，Informatica 支持具有 Enterprise 安全包的 Azure HDInsight 群集。

Enterprise 安全包使用 Kerberos 进行身份验证，使用 Apache Ranger 进行授权。

有关 Enterprise 安全包的详细信息，请参阅《*Informatica Big Data Management 10.2.2 管理员指南*》。

目标

本节介绍版本 10.2.2 中对目标新增的功能。

HDFS 平面文件目标

从版本 10.2.2 开始，您可以将输出数据附加到 HDFS 目标文件和拒绝文件。要附加输出数据，请选择在存在 HDFS 目标时附加数据。

为了便于您管理包含已附加数据的文件，数据集成服务会将映射执行 ID 附加到目标文件和拒绝文件的名称中。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.2 用户指南*》中的“目标”一章。

大数据 Streaming

本节介绍版本 10.2.2 中新增的 Big Data Streaming 功能。

Azure 事件中心数据对象

从版本 10.2.2 开始，可以在以下发行版中部署以事件中心作为源的流映射：

- Amazon EMR
- Azure HDInsight with ADLS storage
- Cloudera CDH
- Hortonworks HDP

Amazon Kinesis 连接中的跨帐户 IAM 角色

从版本 10.2.2 开始，可以使用跨帐户 IAM 角色对 Amazon Kinesis 源进行身份验证。

使用跨帐户 IAM 角色，可以将一个 AWS 帐户中的资源共享给不同 AWS 帐户中的用户，而无需在每个帐户中创建用户。

有关详细信息，请参阅《*Informatica Big Data Streaming 10.2.2 用户指南*》。

智能结构模型

从版本 10.2.2 开始，可以在 Big Data Streaming 中使用智能结构模型。

您可以将智能结构模型纳入 Kafka、Kinesis 或 Azure 事件中心数据对象中。向映射中添加数据对象时，可以处理模型能够解析的任何输入类型。

数据对象可以接受输入并解析 PDF 表单、JSON、Microsoft Excel、Microsoft Word 表、CSV、文本或 XML 输入文件，具体取决于您用来创建该模型的文件。

有关详细信息，请参阅《*Informatica Big Data Streaming 10.2.2 用户指南*》。

Big Data Streaming 数据对象的标头端口

从版本 10.2.2 开始，某些数据对象包含提供与事件关联的元数据的默认标头端口。例如，时间戳端口包含生成事件的时间。您可以使用标头端口对数据进行分组和处理。

有关标头端口的详细信息，请参阅《*Informatica Big Data Streaming 10.2.2 用户指南*》。

Amazon Kinesis 连接中的 AWS 凭据配置文件

从版本 10.2.2 开始，可以在 Amazon Kinesis 连接中使用基于 AWS 凭据配置文件的身份验证。

创建 Amazon Kinesis 连接时，可以输入 AWS 凭据配置文件名称。在运行时，映射通过 AWS 凭据文件中列出的配置文件名称访问 AWS 凭据。

有关详细信息，请参阅《*Informatica Big Data Streaming 10.2.2 用户指南*》。

Spark 结构化流

从版本 10.2.2 开始，Big Data Streaming 使用 Spark 结构化流来处理流数据。

Spark 结构化流是基于 Spark 引擎构建的可扩展容错开源流处理引擎，可处理延迟到达的流事件并基于源时间戳处理流数据。

Spark 引擎持续运行流映射。它会读取数据，将数据分为微批、处理微批、发布结果，然后写入至目标。

有关详细信息，请参阅《*Informatica Big Data Streaming 10.2.2 用户指南*》。

窗口转换

从版本 10.2.2 开始，您可以在创建窗口转换时使用以下功能：

水印延迟

水印延迟定义要累计到数据组中的延迟事件阈值时间。

水印延迟是一个阈值，使用它时，您可以指定可对数据进行分组和处理的数据延迟到达时间。如果事件数据在阈值时间内到达，则会处理数据，并将数据累计到相应的数据组中。

窗口端口

窗口端口指定包含时间戳值（可依据其对事件分组）的列。累计的数据包含时间戳值。使用“窗口端口”列可对延迟到达的事件时间数据分组。

有关详细信息，请参阅《*Informatica Big Data Streaming 10.2.2 用户指南*》。

命令程序

本节介绍版本 10.2.2 中的新命令。

infacmd dis 命令

本主题介绍新增的 infamcd dis 命令选项和执行选项。

更新服务选项

更新服务选项命令可更新数据集成服务属性。

下表介绍了新的 infacmd dis updateServiceOptions 命令选项：

新选项	说明
-RecycleMode -rm	可选。“再次应用”模式会重新启动服务并应用最新服务和进程属性。选择“Abort”或“Complete”。 <ul style="list-style-type: none">- Complete。停止所有应用程序并取消每个应用程序中的所有作业。等待所有作业取消，然后再禁用服务。- Abort。停止所有应用程序并尝试取消所有作业，然后再中止作业并禁用服务。 默认值：Complete

下表介绍了新的 infacmd dis updateServiceOptions 命令执行选项：

新选项	说明
ExecutionOptions.BigDataJobRecovery	启用大数据作业恢复。设置为“true”或“false”。默认值：false。

有关详细信息，请参阅《Informatica 10.2.2 命令引用》中的“infacmd dis 命令引用”一章。

infacmd ihs 命令

下表介绍了新的 infacmd ihs 命令选项：

新选项	说明
-PrimaryNode -nm	可选。运行服务的主节点。
-BackupNodes -bn	可选。主节点不可用时可以运行服务的节点。如果具有高可用性，则可以配置备份节点。

下表介绍了新的 infacmd ihs 命令：

命令	说明
cleanCluster	清理 Informatica 群集服务。

有关详细信息，请参阅《Informatica 10.2.2 命令引用》中的“infacmd ihs 命令引用”一章。

infacmd ipc 命令

下表介绍了已过时的 infacmd ipc 命令：

命令	说明
ExportToPC	从模型存储库或某个导出文件导出对象，并将它们转换为 PowerCenter 对象。

infacmd ldm 命令

下表介绍了新的 infacmd ldm 命令选项：

新选项	说明
-PrimaryNode -nm	可选。如果要为 Enterprise Data Catalog 配置高可用性，请指定主节点名称。
-BackupNodes -bn	可选。如果要为 Enterprise Data Catalog 配置高可用性，请指定备份节点名称的逗号分隔列表。

新选项	说明
-isNotifyChangeEmailEnabled -cne	可选。如果要启用资产更改通知，则指定 True。默认值为 False。
-ExtraJarsPath -ejp	可选。安装了 Informatica 域的计算机上的目录路径。该目录必须包含使用 WANDisco Fusion 在现有群集上部署 Enterprise Data Catalog 时所需的 JAR 文件。
-ExtraJarsPath -ejp	可选。安装了 Informatica 域的计算机上的目录路径。该目录必须包含使用 WANDisco Fusion 在现有群集上部署 Enterprise Data Catalog 时所需的 JAR 文件。

下表介绍了新的 infacmd ldm 命令：

命令	说明
collectAppLogs	收集为启用目录服务而运行的 YARN 应用程序的日志文件。
publishArchive	在脱机模式下创建资源，并运行扫描。

有关详细信息，请参阅《Informatica 10.2.2 命令引用》中的“infacmd ldm 命令引用”一章。

infacmd mi 命令

下表介绍了对 infacmd mi 命令进行的更改：

命令	更改说明
createService	从版本 10.2.2 开始，可以使用 -HttpsPort、-KeystoreFile 和 -KeystorePassword 选项来指定 Mass Ingestion 服务进程是否使用安全连接与外部组件通信。
extendedRunStats	从版本 10.2.2 开始，必须使用 -RunID 选项来指定 Mass Ingestion 规范的 RunID，使用 -SourceName 选项来指定源表的名称才能查看源表的扩展运行统计信息。如果源表是使用增量加载获取的，运行统计信息则显示增量键和起始值。 在以前的版本中，您指定的是获取源表的获取映射作业的 JobID。 如果升级到 10.2.2，则必须更新运行 infacmd mi extendedRunStats 的脚本以使用新选项。
listSpecRuns	从版本 10.2.2 开始，该命令会额外返回 Spark 引擎用于运行 Mass Ingestion 规范的加载类型。
runSpec	从版本 10.2.2 开始，可以使用 -LoadType 选项来指定运行 Mass Ingestion 规范的加载类型。加载类型可以是完全加载，也可以是增量加载。

有关详细信息，请参阅《Informatica 10.2.2 命令引用》中的“infacmd mi 命令引用”一章。

infacmd ms 命令

下表介绍了新的 infacmd ms 命令：

命令	说明
abortAllJobs	中止配置为在 Spark 引擎上运行的所有已部署映射作业。您可以选择中止排队的作业和/或正在运行的作业。
createConfigurationWithParams	通过在命令行中指定的群集参数创建群集配置。
listMappingOptions	列出应用程序中的映射选项。
purgeDatabaseWorkTables	为数据集成服务启用大数据恢复时，从队列中清除所有作业信息。
updateMappingOptions	更新应用程序中的映射选项。
updateOptimizationLevel	更新应用程序中多个映射的优化级别。

有关详细信息，请参阅《*Informatica 命令引用*》中的“infacmd ms 命令引用”一章。

infacmd oie 命令

从版本 10.2.2 开始，oie 插件已弃用，在将来的版本中将停止对该插件的支持。infacmd oie 命令已迁移到 tools 插件。有关详细信息，请参阅“[infacmd tools 命令](#)”[页面上 50](#)。

infacmd tools 命令

从版本 10.2.2 开始，tools infacmd 插件替代 oie 插件。tools 插件执行 oie 插件所执行的对象导入和导出操作，您可以使用它来执行某些其他操作。

下表列出已迁移至 tools 插件的 infacmd oie 命令：

先前命令	当前命令
infacmd oie deployApplication	infacmd tools deployApplication
infacmd oie exportObjects	infacmd tools exportObjects
infacmd oie exportResources	infacmd tools exportResources
infacmd oie importObjects	infacmd tools importObjects

下表介绍了新的 infacmd tools 命令：

命令	说明
patchApplication	使用 .piar 文件将应用程序修补程序部署到数据集成服务。

有关详细信息，请参阅《*Informatica 10.2.2 命令引用*》中的“infacmd tools 命令引用”一章。

infasetup 命令

下表介绍了已更改的 infasetup 命令：

命令	说明
DefineDomain	从版本 10.2.2 开始，DefineDomain 命令添加了 -spid 选项。
updateDomainSamlConfig	从版本 10.2.2 开始，updateDomainSamlConfig 命令添加了 -spid 选项。

有关详细信息，请参阅《Informatica 10.2.2 命令引用》中的“infasetup 命令引用”一章。

Enterprise Data Catalog

本节介绍版本 10.2.2 中新增的 Enterprise Data Catalog 功能。

为列自动分配业务标题

从版本 10.2.2 开始，Enterprise Data Catalog 将业务术语推理为列资产的业务标题。置信度得分为 80% 或更高的业务术语被推理为列的业务标题。置信度得分通过一个内部算法计算得出，该算法基于数据域上接受的业务术语、列名相似性以及列和业务术语之间的名称相似性。

有关详细信息，请参阅《Informatica 10.2.2 Enterprise Data Catalog 用户指南》中的“执行资产任务”一章。

资产方面的用户协作

从版本 10.2.2 开始，您可以就资产与其他 Enterprise Data Catalog 用户展开协作。利用资产协作，您可以与其他用户互动、分享有关资产的观点、请求与资产相关的查询、对所有资产更改进行后续跟踪以及认证资产。

可通过 ([a href="#">以下方式进行资产协作：

关注资产

您可以关注资产以监视目录中的资产更改。关注资产可了解其他用户对资产所做的更改，这样您便可以监视资产并进行必要的操作。

评定和审查资产

您可以根据目录中的五星级别对 ([a href="#">资产进行评定和审查。评定和审查资产可根据资产的不同方面（如资产的质量、适用性、易用性和可用性）提供资产的相关反馈。

资产查询

如果要更好地了解目录中的资产，可以提出有关资产的问题。请提出具有描述性、探索性、预测性或自然因果 ([a href="#">关系的问题。

认证资产

您可以对资产进行认证使其合法化，这样其他用户便可优先使用已认证的资产作为可靠资产。

有关详细信息，请参阅《Informatica 10.2.2 Enterprise Data Catalog 指南》中的“资产方面的用户协作”一章。

使用安装程序创建 Enterprise Data Catalog 应用程序服务

从版本 10.2.2 开始，可以在安装 Enterprise Data Catalog 后使用安装程序创建 Enterprise Data Catalog 应用程序服务。如果已安装 Enterprise Data Catalog 而未创建应用程序服务，则可以使用安装程序。

有关使用安装程序创建应用程序服务的详细信息，请参阅《*Informatica Enterprise Data Catalog 10.2.2 安装和配置指南*》。

自定义元数据验证实用程序

从版本 10.2.2 开始，可以使用基于 Java 的独立验证命令行实用程序来验证要在目录中引入的自定义元数据的语法和语义。自定义元数据表示要从 Enterprise Data Catalog 未为其提供资源的自定义数据源引入的元数据。

有关使用该实用程序的详细信息，请参阅知识库文章《*操作：在目录中引入自定义元数据之前验证自定义元数据*》。有关下载该实用程序的说明，请联系 Informatica 全球客户支持。

更改通知

从版本 10.2.2 开始，当您关注的资产发生更改时，Enterprise Data Catalog 会显示通知。通知类型包括应用程序通知、更改电子邮件通知和摘要电子邮件通知。

有关详细信息，请参阅《*Informatica 10.2.2 Enterprise Data Catalog 指南*》中的“资产方面的用户协作”一章。

Business Glossary 分配报告

从版本 10.2.2 开始，可以接受或拒绝为资源建议的多个业务术语。

有关详细信息，请参阅《*Informatica 10.2.2 Enterprise Data Catalog 指南*》中的“执行资产任务”一章。

操作系统配置文件

从版本 10.2.2 开始，如果没有默认操作系统配置文件，可以选择一个操作系统配置文件。数据集成服务使用操作系统配置文件用户凭据来执行数据发现。数据发现包括列配置文件和数据域发现配置文件。

有关在 Enterprise Data Catalog 中使用操作系统配置文件的详细信息，请参阅《*Informatica 10.2.2 Catalog Administrator 指南*》中的“Enterprise Data Catalog 概念”一章。

REST API

从版本 10.2.2 开始，可以使用以下 Informatica Enterprise Data Catalog REST API：

- 业务术语 REST API。可以返回、更新或删除已接受、已推理或已拒绝的业务术语。
- 目录事件 REST API。可以访问、更新或删除用户配置、电子邮件配置和用户订阅。
- 对象证书 API。可以列出、更新和删除对象的证书属性。
- 对象注释 API。可以列出、创建、更新和删除数据对象的注释、回复和投票。
- 对象审阅 API。可以列出、创建、更新和删除审阅的评论、评级和投票。

有关 REST API 的详细信息，请参阅《*Informatica 10.2.2 Enterprise Data Catalog REST API 参考*》。

源元数据和数据配置文件筛选器

从版本 10.2.2 开始，可以在资源运行中使用源元数据筛选器和数据配置文件筛选器指定源表和视图。使用这些筛选器时，Enterprise Data Catalog 将从特定表和视图中提取源元数据和配置文件元数据。

有关源元数据和数据配置文件筛选器的详细信息，请参阅《*Informatica 10.2.2 Catalog Administrator 指南*》中的“管理资源”一章。

扫描程序实用程序

从版本 10.2.2 开始，Informatica 提供独立的扫描程序实用程序，可用于从脱机和不可访问的资源提取元数据。该实用程序包含需要与关联的命令一起按顺序运行的脚本。

有关该独立的扫描程序实用程序的详细信息，请参阅《*Informatica 10.2.2 Catalog Administrator 指南*》中的附录“从脱机和不可访问的资源提取元数据”。

资源类型

从版本 10.2.2 开始，您可以为以下数据源类型创建资源：

Google BigQuery

可以从 Google BigQuery 数据源的以下资产中提取元数据、关系和沿袭信息：

- 项目
- 数据集
- 表
- 视图

有关配置 Google BigQuery 数据源的详细信息，请参阅《*Informatica 10.2.2 Catalog Administrator 指南*》。

Workday

可以从 Workday 数据源的以下资产中提取元数据、关系和沿袭信息：

- 服务
- 实体
- 报表
- 操作
- 数据源
- 属性
- 业务对象

有关配置 Workday 数据源的详细信息，请参阅《*Informatica 10.2.2 Catalog Administrator 指南*》。

Enterprise Data Lake

本节介绍版本 10.2.2 中新增的 Enterprise Data Lake 功能。

应用活动规则

从版本 10.2.2 开始，可以在项目中使用活动规则：

活动规则是使用 Developer tool 开发的 Maplet。可以使用活动规则将复杂转换（如汇总器和数据质量转换）应用于工作表以进行匹配和合并。

活动规则使用数据集内的所有行作为输入。您可以选择多个工作表作为规则的输入。应用程序将向项目中添加包含规则输出的工作表。

有关详细信息，请参阅《*Informatica 10.2.2 Enterprise Data Lake 用户指南*》中的“准备数据”一章。

删除重复行

从版本 10.2.2 开始，您可以从工作表中删除包含重复值的行。

有关详细信息，请参阅《*Informatica 10.2.2 Enterprise Data Lake 用户指南*》中的“准备数据”一章。

聚集和分类列数据

从版本 10.2.2 开始，您可以聚集列中的类似值，然后根据 Enterprise Data Lake 中的建议对这些值分类。该应用程序使用语音算法聚集相似值，然后建议您使用最常出现的值替换不常出现的值。

有关详细信息，请参阅《*Informatica 10.2.2 Enterprise Data Lake 用户指南*》中的“准备数据”一章。

基于 CLAIRE 的建议

从版本 10.2.2 开始，应用程序使用嵌入式 CLAIRE 机器学习发现引擎在您准备数据时提供建议。

当您查看“项目”页面时，应用程序根据数据沿袭以及记录的主-外键关系显示派生自上游数据源的备用和其他建议。

在数据准备期间从工作表中选择列时，应用程序根据“列概览”面板中的列数据类型显示数据改善的相关建议。

对两个工作表执行联接操作时，应用程序使用主-外键关系指示所需键对重叠较少时的不兼容采样。

有关详细信息，请参阅《*Informatica 10.2.2 Enterprise Data Lake 用户指南*》中的“准备数据”一章。

条件性汇总

从 10.2.2 开始，可以使用 AND 和 OR 逻辑针对在项目中创建汇总工作表时使用的 IF 计算应用多个条件。

- AND 与所有运算符配合使用可在条件中包含多个列。
- OR 与 IS、IS NOT 和 IS BETWEEN 运算符配合使用可在条件中的列内包含多个值。

有关详细信息，请参阅《*Informatica 10.2.2 Enterprise Data Lake 用户指南*》中的“准备数据”一章。

数据屏蔽

从版本 10.2.2 开始，Enterprise Data Lake 与一款数据安全产品 Informatica Dynamic Data Masking 相集成，以屏蔽数据资产中的敏感数据。

要在 Enterprise Data Lake 中启用数据屏蔽，需配置 Dynamic Data Masking 服务器以将屏蔽规则应用于数据湖中的数据资产。此外，还需配置 Informatica 域，以便 Enterprise Data Lake 可以连接到 Dynamic Data Masking 服务器。

Dynamic Data Masking 会拦截从 Enterprise Data Lake 发送至数据湖的请求，并将屏蔽规则应用于所请求资产中的各个列。当 Enterprise Data Lake 用户查看包含被屏蔽数据的列或对这些列执行操作时，系统会根据所应用的屏蔽规则对全部或部分实际数据进行模糊处理。

有关详细信息，请参阅《*Informatica 10.2.2 Enterprise Data 管理员指南*》中的“屏蔽敏感数据”一章。

本地化

从版本 10.2.2 开始，用户界面支持日语。您还可以在项目名称和说明中使用非拉丁字符。

分区源和目标

从版本 10.2.2 开始，Enterprise Data Lake 可以在导入、发布或复制操作期间读取分区源中的数据。应用程序还可以在导入、发布、复制或上载操作期间将数据附加到数据湖中的分区目标。

向脚本步骤添加注释

从版本 10.2.2 开始，可以向脚本步骤中添加注释。使用注释改善协作，并提供详细信息以满足审计要求。

有关详细信息，请参阅《*Informatica 10.2.2 Enterprise Data Lake 用户指南*》中的“准备数据”一章。

将脚本另存为映射

从版本 10.2.2 开始，可以将脚本另存为映射，而非发布脚本并创建新输出表。

可以将映射保存到与 Enterprise Data Lake 服务关联的模型存储库中，也可以将映射保存到 .xml 文件。开发人员可以使用 Developer tool 查看和修改映射，然后根据系统资源可用性适时执行映射。

有关详细信息，请参阅《*Informatica 10.2.2 Enterprise Data Lake 用户指南*》中的“准备数据”一章。

Amazon S3、ADLS、WASB、MapR-FS 作为数据源

从版本 10.2.2 开始，您可以在下列数据源中所存储的文件内准备数据：

- Amazon S3
- MapR-FS
- Microsoft Azure Data Lake Storage
- Windows Azure 存储 Blob

必须在 Enterprise Data Catalog 中为包含待准备数据的每个数据源创建一个资源。资源为表示外部数据源或元数据存储库的存储库对象。附加至资源的扫描程序从资源提取元数据，并将元数据存储在企业数据目录中。

有关在 Enterprise Data Catalog 中创建资源的详细信息，请参阅《*Informatica 10.2.2 Catalog Administrator 指南*》中的“管理资源”一章。

统计函数

从版本 10.2.2 开始，您可以在准备数据时对工作表中的列应用以下统计函数：

- AVG
- AVGIF
- COUNT
- COUNTIF
- COUNTDISTINCT
- COUNTDISTINCTIF
- MAX
- MAXIF
- MIN
- MINIF

- STDDEV
- STDDEVIF
- SUM
- SUMIF
- VARIANCE
- VARIANCEIF

有关详细信息，请参阅《*Informatica 10.2.2 Enterprise Data Lake 用户指南*》中的“准备数据”一章。

日期和时间函数

从版本 10.2.2 开始，您可以在准备数据时对工作表中的列应用以下日期和时间函数：

- ADD_TO_DATE
- CURRENT_DATETIME
- DATETIME
- DATE_DIFF
- DATE_TO_UNIXTIME
- EXTRACT_MONTH_NAME
- UNIXTIME_TO_DATE
- 转换日期为文本
- 转换文本为日期

有关详细信息，请参阅《*Informatica 10.2.2 Enterprise Data Lake 用户指南*》中的“准备数据”一章。

数学函数

从版本 10.2.2 开始，您可以在准备数据时对列应用以下数学函数：

- EXP
- LN
- LOG
- PI
- POWER
- SQRT

有关详细信息，请参阅《*Informatica 10.2.2 Enterprise Data Lake 用户指南*》中的“准备数据”一章。

文本函数

从版本 10.2.2 开始，您可以在准备数据时对列应用以下文本函数：

- ENDSWITH
- ENDSWITH_IGNORE_CASE
- FIND_IGNORE_CASE
- FIND_REGEX

- FIRST_CHARACTER_TO_NUMBER
- NUMBER_TO_CHARACTER
- PROPER_CASE
- REMOVE_NON_ALPHANUMERIC_CHARACTERS
- STARTSWITH
- STARTSWITH_IGNORE_CASE
- SUBSTITUTE_REGEX
- TRIM_ALL
- 转换日期为文本
- 转换数字为文本
- 转换文本为日期
- 转换文本为数字

有关详细信息，请参阅《*Informatica 10.2.2 Enterprise Data Lake 用户指南*》中的“准备数据”一章。

窗口函数

从版本 10.2.2 开始，可以使用窗口函数对工作表中的行组执行操作。在其上运行函数的行组称为窗口，可使用分区键、按键排序和可选的偏移量定义窗口。窗口函数计算窗口上下文内每个输入行的返回值。

您可以使用窗口函数执行以下任务：

- 从以前或后续行中检索数据。
- 根据行组计算累计总和或累计平均值。
- 将顺序行编号分配给行组中的每个行。
- 在行组内将行中的空值替换为前面的非空值。
- 根据特定时间段（例如，日志文件中记录的网站访问次数）生成可用于归组行的会话标识符。

您可以对工作表应用多个窗口函数。例如，可以应用一个函数来计算窗口内当前行后面的每个行的值总和，然后应用另一个函数来计算相同值的平均值。

对于应用于数据表的每个函数，Enterprise Data Lake 都会添加一个包含相应结果的列。

有关详细信息，请参阅《*Informatica 10.2.2 Enterprise Data Lake 用户指南*》中的“准备数据”一章。

清除审计事件

从版本 10.2.2 开始，可以运行 `infacmd edl purgeevents` 命令以从审计历史记录数据库中删除用户活动。您可以有选择性地运行此命令以从数据库中删除项目历史记录事件。

Spark 执行引擎

从版本 10.2.2 开始，Enterprise Data Lake 使用 Spark 引擎执行占用资源较高的活动（例如资产发布），并运行使用 Python 转换的活动规则 Maplet。使用 Spark 引擎执行占用资源较高的活动可以提高性能，并可在 Amazon Elastic MapReduce (EMR) 上部署 Enterprise Data Lake 以利用自动调整功能。

Informatica Developer

本节介绍版本 10.2.2 中新增的 Developer tool 功能。

应用程序

从版本 10.2.2 开始，可以创建增量应用程序。增量应用程序是可通过部署应用程序补丁来更新一部分应用程序对象的应用程序。数据集成服务会更新补丁中的对象，同时其他应用程序对象继续运行。

如果升级到版本 10.2.2，现有应用程序将标记为“完整应用程序”。您可以继续在版本 10.2.2 中创建完整应用程序，但无法将完整的应用程序转换为增量应用程序。

有关详细信息，请参阅《*Informatica 10.2.2 Developer Tool 指南*》中的“应用程序部署”和“应用程序补丁部署”两章。

Informatica 映射

本节介绍版本 10.2.2 中新增的 Informatica 映射功能。

数据类型

从版本 10.2.2 开始，您可以在运行于 Spark 引擎上的批量映射中启用高精度模式。Spark 引擎可以处理最高精度为 38 位数的小数值。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.2 用户指南*》。

映射输出

从版本 10.2.2 开始，可以在作为工作流中的映射任务在 Spark 引擎上运行的批量映射中使用映射输出。您可以将映射输出保留在模型存储库中，也可以将映射输出绑定到工作流变量。

有关详细信息，请参阅《*Informatica 10.2.2 Developer 映射指南*》中的“映射输出”一章和《*Informatica 10.2.2 Developer 工作流指南*》中的“映射任务”一章。

映射参数

从版本 10.2.2 开始，可以将表达式参数分配给在本地和非本地环境下运行的汇总器转换、表达式转换和等级转换中的端口表达式。

有关详细信息，请参阅《*Informatica 10.2.2 Developer 映射指南*》中的“在何处分配参数”和“动态映射”这两章。

优化器级别

从版本 10.2.2 开始，可以为映射和映射任务配置“自动”优化器级别。在“自动”优化级别下，数据集成服务可根据执行模式和映射内容来应用优化。

新映射的默认优化器级别为“自动”。

升级到版本 10.2.2 时，在映射中配置的优化器级别保持不变。要对升级后的映射使用“自动”优化器级别，必须手动更改优化器级别。

有关详细信息，请参阅《*Informatica 10.2.2 Developer 映射指南*》中的“优化器级别”一章。

Sqoop

从版本 10.2.2 开始，可以使用以下新的 Sqoop 功能：

增量数据提取支持

可以将 Sqoop 映射配置为根据 ID 或时间戳来执行增量数据提取。使用增量数据提取时，Sqoop 仅提取自上次数据提取后发生更改的数据。增量数据提取可提高映射性能。

Vertica 连接支持

您可以将 Sqoop 配置为从 Vertica 源读取数据，或将数据写入到 Vertica 目标。

Sqoop 传递映射的 Spark 引擎优化

在 Spark 引擎上使用 Sqoop 源运行传递映射时，数据集成服务在下列情况下会优化映射性能：

- 将数据写入到使用自定义 DDL 查询创建的 Hive 目标。
- 将数据写入到使用自定义 DDL 查询分区或使用自定义 DDL 查询分区并带桶的现有 Hive 目标。
- 将数据写入到分区并带桶的现有 Hive 目标。

--infaownername 参数支持

您可以配置 --infaownername 参数，以指示 Sqoop 是否必须优先处理数据对象的所有者名称。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.2 用户指南*》。

Informatica 转换

本节介绍版本 10.2.2 中新增的 Informatica 转换功能。

地址验证器转换

本节介绍地址验证器转换的新增功能。

对于以下国家/地区，地址验证器转换具有附加的地址功能：

所有国家/地区

从版本 10.2.2 开始，地址验证器转换在 Informatica 提供引用地址数据的每一个国家/地区均支持单行地址验证。

在早期版本中，该转换为 26 个国家/地区支持单行地址验证。

要验证单行地址，请在完整地址端口中输入地址。如果地址确定默认首选语种不是拉丁或西语语种的国家/地区，请在具有该地址的转换上使用默认“首选语种”属性。

澳大利亚

从版本 10.2.2 开始，可以将地址验证器转换配置为向澳大利亚地址中添加地址扩充内容。可以使用扩充内容来发现澳大利亚统计局为其分配地址的地理区域和地区。这些地理区域和地区包括人口调查收集区、网格块和统计区域。

该转换使用以下端口提供扩充内容：

- 人口调查收集区代码 2006

- 地理编码国家地址文件标识符
- 大首都城市统计区域 5 位
- 大首都城市统计区域名称
- 一级统计区域 11 位
- 一级统计区域 7 位
- 二级统计区域 9 位
- 二级统计区域 5 位
- 二级统计区域名称
- 三级统计区域 5 位
- 三级统计区域名称
- 四级统计区域 3 位
- 四级统计区域名称
- 网格块 11 位 2011
- 网格块 11 位 2016
- 国家或地区代码
- 国家或地区名称
- 补充澳大利亚状态

在“澳大利亚补充”端口组中查找此端口。

玻利维亚

从版本 10.2.2 开始，地址验证器转换改进了玻利维亚地址的解析和验证。此外，Informatica 更新了玻利维亚的引用数据。

转换还包括针对玻利维亚的以下改进：

- 到街道级别的地址验证。
- 将大城市地址的地理坐标定位在街道中点级别。

加拿大

Informatica 对加拿大引入了以下功能和增强功能：

在加拿大地址中支持全局首选描述符属性

从版本 10.2.2 开始，可以将地址验证器转换配置为返回短或长形式的元素描述符。

该转换可以返回短或长形式的以下描述符：

- 街道描述符
- 方向值
- 建筑物描述符
- 子建筑物描述符

要指定描述符的输出格式，请配置该转换的“全局首选描述符”属性。该属性适用于英语和法语描述符。默认情况下，转换会以引用数据指定的格式返回描述符。如果您在该属性上选择 PRESERVE INPUT 选项，那么“首选语言”属性将优先于“全局首选描述符”属性。

支持 CH 和 CHAMBER 作为子建筑物描述符

从版本 10.2.2 开始，地址验证器转换在加拿大地址中将 CH 和 CHAMBER 识别为子建筑物描述符。

哥伦比亚

从版本 10.2.2 开始，地址验证器转换改进了哥伦比亚地址中的街道数据处理。此外，Informatica 更新了哥伦比亚的引用数据。

法国

从版本 10.2.2 开始，Informatica 对法国地址引入了以下改进：

- Informatica 改进了法国的补充引用数据。
- 地址验证器转换可以更精确地将地址分配给法国的 IRIS 单位。转换使用地址中的门牌号来验证地址所属的 IRIS 单位。当地址位于不同单位间的边界附近时，使用门牌号可以提高分配的精度。

以色列

从版本 10.2.2 开始，Informatica 引入了以下适用于以色列地址的功能和增强功能：

以色列地址的多语言支持

可以将地址验证器转换配置为以英语或希伯来语返回以色列地址。

使用“首选语言”属性可选择转换返回的地址的首选语言。

以色列地址的默认语言为希伯来语。要以希伯来语返回地址信息，请将“首选语言”属性设置为 DATABASE 或 ALTERNATIVE_1。要以英语返回地址信息，请将该属性设置为 ENGLISH 或 ALTERNATIVE_2。

支持将多个字符集用于以色列地址

地址验证器转换可以在希伯来语和拉丁字符集中读取和写入以色列地址。

使用“首选脚本”属性选择地址数据的首选字符集。

以色列地址的默认字符集为希伯来语。将“首选脚本”属性设置为 Latin 或 Latin-1 时，该转换会将希伯来语地址数据音译为拉丁字符。

秘鲁

从版本 10.2.2 开始，地址验证器转换会将秘鲁地址验证到门牌号级别。此外，Informatica 更新了秘鲁的引用数据。

瑞典

从版本 10.2.2 开始，地址验证器转换改进了瑞典地址中街道名称的验证。

转换通过以下方式改进了街道名称的验证：

- 转换可以将以字符 G 结尾的街道名称识别为最后字符为 GATAN 的相同名称的别名。
- 转换可以将以字符 V 结尾的街道名称识别为最后字符为 VÄGEN 的相同名称的别名。
- 在使用长形式或短形式的描述符时，地址验证器转换可以识别并更正描述符不正确的街道名称。

例如，转换可以将以下地址中的 RUNIUSV 或 RUNIUSVÄGEN 更正为 RUNIUSGATAN：

RUNIUSGATAN 7

SE-112 55 STOCKHOLM

美国

从版本 10.2 HotFix2 开始，可以将地址验证器转换配置为识别在一周内的一天或多天不接收邮件的美国地址。

要识别这些地址，请使用 *非交付天数* 端口。该端口包含一个七位数字字符串，用于表示星期几（从星期日到星期六）。字符串中的每个位置代表不同的一天。

如果地址在某天未收到邮件，地址验证器转换则会返回端口上对应位置的星期几的第一个字母。该转换在一周内的其他天的对应位置返回虚线符号。

例如，“非交付天数”端口上的 S---FS 值表示地址在星期日、星期五和星期六不接收邮件。

在“基本”模型的“特定于美国”端口组中查找“非交付天数”端口。要在“非交付天数”端口上接收数据，请在认证模式下运行地址验证器转换。该转换即从 USA5C129.MD 和 USA5C130.MD 数据库文件中读取端口值。

有关版本 10.2.2 中的地址验证软件引擎的功能和操作的全面信息，请参阅《*Informatica Address Verification 5.14.0 开发人员指南*》。

更新策略转换

从版本 10.2.2 开始，可以在运行于 Spark 引擎上的映射中使用更新策略转换来更新关系目标。

以前，在运行于 Spark 引擎上的映射中使用更新策略转换只能更新 Hive 目标。

有关详细信息，请参阅《*Developer 转换指南*》中的“更新策略转换”一章。

Informatica PowerExchange 适配器

本节介绍版本 10.2.2 中新增的 Informatica 适配器功能。

PowerExchange for Amazon Redshift

从版本 10.2.2 开始，PowerExchange for Amazon Redshift 包含以下功能：

- 可以在以下区域读取或写入数据：
 - 中国（宁夏）
 - 欧盟（巴黎）
- 可以在映射中使用 Amazon Redshift 对象作为动态源和目标。
- 可以使用可打印和不可打印 ASCII 字符的八进制值作为 DELIMITER 或 QUOTE。
- 可以输入 pre-SQL 和 post-SQL 命令以便为映射中的源对象和目标对象运行查询。
- 可为映射中的读取数据对象定义 SQL 查询，以替代默认查询。可以输入受 Amazon Redshift 数据库支持的 SQL 语句。
- 分多个部分下载大型 Amazon S3 对象时，可以指定 Amazon S3 对象的最大大小（以字节为单位）。
- 从 Amazon Redshift 源读取数据时，可以读取唯一值。
- 将对象上载到 Amazon S3 时，可以指定对象的大小下限以及用于将对象作为一组独立的多个部分并行上载的线程数。
- 可以选择保留现有目标表，在运行时替换目标表，或在目标中不存在表时创建新目标表。
- 可以在本地环境中为 Amazon Redshift 目标配置更新策略转换。
- 将数据写入 Amazon Redshift 时，可以在运行时替代 Amazon Redshift 目标表架构和表名称。
- 如果连接类型为 ODBC，数据集成服务可以使用源端和完整下推优化将转换逻辑推送至 Amazon Redshift 源和目标。
- 可以在 Amazon EMR 版本 5.16 和 Cloudera CDH 版本 5.15 及 5.16 上配合使用服务器端加密和 AWS KMS（AWS 密钥管理服务）。
- PowerExchange for Amazon Redshift 支持适用于 Java 的 AWS 开发工具包版本 1.11.354。

有关详细信息，请参阅《*Informatica PowerExchange for Amazon Redshift 10.2.2 用户指南*》。

PowerExchange for Amazon S3

从版本 10.2.2 开始，PowerExchange for Amazon S3 包含以下功能：

- 可以在以下区域读取或写入数据：
 - 中国（宁夏）
 - 欧盟（巴黎）
 - AWS GovCloud（美国）
- 可以在映射中使用 Amazon S3 对象作为动态源和目标。
- 在本地环境中或在 Spark 引擎上运行映射以从 Avro、平面、JSON、ORC 或 Parquet 文件读取数据时，可以使用通配符字符来指定源目录名称或源文件名称。
- 可以向存储在 Amazon S3 存储桶中的对象添加单个或多个标记，以便对这些对象进行分类。每个标记都包含一个键值对。可以输入键值对，或者指定包含键值对的绝对文件路径。
- 可以指定最大阈值大小，以便分多个部分下载 Amazon S3 对象。
- 将对象上载到 Amazon S3 时，可以指定对象的大小下限以及用于将对象作为一组独立的多个部分并行上载的线程数。
- 创建数据对象读取或写入操作时，可以读取包含端点名称和文件源路径的“文件名”端口中的数据。
- 在创建数据对象读取或写入操作时，可以直接在“端口”选项卡中添加新列或修改列。
- 可以从“端口”选项卡复制源转换、目标转换或任何其他转换的列，并在创建映射来读取或写入 Avro、JSON、ORC 或 Parquet 文件时，直接在对象读取或写入操作中粘贴列。
- 可以更新 Amazon S3 文件格式，而不丢失列投影属性的“架构”字段中的列元数据，甚至在为另一 Amazon S3 文件格式配置列投影属性后也是如此。
- 可以在 Amazon EMR 版本 5.16 和 Cloudera CDH 版本 5.15 及 5.16 上配合使用服务器端加密和 AWS KMS（AWS 密钥管理服务）。
- PowerExchange for Amazon S3 支持适用于 Java 的 AWS 开发工具包版本 1.11.354。

有关详细信息，请参阅《*Informatica PowerExchange for Amazon S3 10.2.2 用户指南*》。

PowerExchange for Google BigQuery

从版本 10.2.2 开始，可以使用**创建目标**鼠标右键菜单选项来创建 Google BigQuery 目标。

有关详细信息，请参阅《*Informatica PowerExchange for Google BigQuery 10.2.2 用户指南*》。

PowerExchange for HBase

从版本 10.2.2 开始，PowerExchange for HBase 包含以下新功能：

- 创建 HBase 数据对象时，如果从 Hadoop 群集导入并预览元数据，则可以选择操作系统配置文件以提高安全性，并隔离设计时用户环境。
注意：如果元数据访问服务配置为使用操作系统配置文件，可以选择操作系统配置文件。元数据访问服务将使用分配给用户的默认操作系统配置文件导入元数据。可以从可用操作系统配置文件的列表中更改操作系统配置文件。
- 可以在映射中使用 HBase 对象作为动态源和目标。
- 可以在 Spark 引擎上运行映射，以在 HBase 资源中查找数据。

有关详细信息，请参阅《*Informatica PowerExchange for HBase 10.2.2 用户指南*》。

PowerExchange for HDFS

从版本 10.2.2 开始，PowerExchange for HDFS 包含以下新功能：

- 创建复杂文件数据对象时，如果从 Hadoop 群集导入并预览元数据，则可以选择操作系统配置文件以提高安全性，并隔离设计时用户环境。

注意：如果元数据访问服务配置为使用操作系统配置文件，可以选择操作系统配置文件。元数据访问服务将使用分配给用户的默认操作系统配置文件导入元数据。可以从可用操作系统配置文件的列表中更改操作系统配置文件。

- 在本地环境中或 Spark 引擎上运行映射以从复杂文件数据对象读取数据时，可以使用通配符指定源目录名称或源文件名称。

您可以使用以下通配符：

? (问号)

问号字符 (?) 允许任意字符出现 1 次。

* (星号)

星号字符 (*) 允许任意字符出现 0 次或超过 1 次。

- 可以在映射中使用复杂文件对象作为动态源和目标。
- 可以使用复杂文件对象从复杂文件系统读取数据或将数据写入复杂文件系统。
- 在本地机环境中或 Spark 引擎上运行映射以将数据写入复杂文件数据对象时，可以覆盖目标数据，数据集成服务将在写入新数据之前删除目标数据。
- 创建数据对象读取或写入操作时，可以读取包含端点名称和文件源路径的 FileName 端口中存在的数据库。
- 现在可以在创建数据对象读取或写入操作之后立即查看数据对象操作。
- 创建数据对象读取或写入操作时，可以添加新列或修改列。
- 可以复制源转换、目标转换或任何其他转换的列，并在读取或写入到 Avro、JSON、ORC 或 Parquet 文件时，直接在数据对象读取或写入操作中粘贴列。

有关详细信息，请参阅《*Informatica PowerExchange for HDFS 10.2.2 用户指南*》。

PowerExchange for Hive

从版本 10.2.2 开始，PowerExchange for Hive 包含以下新功能：

- 可以为 Hive 目标配置以下目标架构策略选项：
 - RETAIN - 保留现有目标架构
 - CREATE - 在运行时创建或替换表
 - APPLYNEWCOLUMNS - 仅更改表并应用新列
 - APPLYNEWSHEMA - 更改表并应用新架构
 - FAIL - 如果目标架构不同，映射将失败
 - 分配参数
- 可以在加载数据之前截断内部或外部分区 Hive 目标。在 Hadoop 环境中运行映射时，此选项适用。
- 可以在本地模式下为 Hive 创建读取或写入转换，以从 Hive 源读取数据或将数据写入 Hive 目标。
- 将数据写入 Hive 目标时，可以在 Hive 连接中配置以下属性：
 - HDFS 上的 Hive 暂存目录。表示 Hive 暂存表的 HDFS 目录。在本地环境中将数据写入 Hive 目标时，此选项适用且为必填选项。

- Hive 暂存数据库名称。表示 Hive 暂存表的命名空间。在本地环境中运行映射以将数据写入 Hive 目标时，此选项适用。如果在 Blaze 或 Spark 引擎上运行映射，则无需在 Hive 连接中配置 Hive 暂存数据库名称。数据集成服务使用在 Hadoop 连接中配置的值。

有关详细信息，请参阅《*Informatica PowerExchange for Hive 10.2.2 用户指南*》。

PowerExchange for MapR-DB

从版本 10.2.2 开始，在为 MapR-DB 创建 HBase 数据对象时，如果从 Hadoop 群集导入并预览元数据，则可以选择操作系统配置文件以提高安全性，并隔离设计时用户环境。

注意：如果元数据访问服务配置为使用操作系统配置文件，可以选择操作系统配置文件。元数据访问服务将使用分配给用户的默认操作系统配置文件导入元数据。可以从可用操作系统配置文件的列表中更改操作系统配置文件。

有关详细信息，请参阅《*Informatica PowerExchange for MapR-DB 10.2.2 用户指南*》。

PowerExchange for Microsoft Azure Blob 存储

从版本 10.2.2 开始，PowerExchange for Microsoft Azure Blob 存储包含以下功能：

- 可以在 Azure Databricks 环境中运行映射。
- 可以配置美国政府 Microsoft Azure 端点。
- 从 Microsoft Azure Blob 存储读取数据或向其中写入数据时，可以将数据压缩为以下格式：
 - 无
 - Deflate
 - Gzip
 - Bzip2
 - Lzo
 - Snappy
- 可以在映射中使用 Microsoft Azure Blob 存储对象作为动态源和目标。
- 对于数据集成服务在本地环境中在运行时从中读取数据的文件，可以读取其文件名。
- 可以在高级源和目标属性的 **Blob 容器替代**中配置相对路径。

有关详细信息，请参阅《*Informatica PowerExchange for Microsoft Azure Blob 存储 10.2.2 用户指南*》。

PowerExchange for Microsoft Azure Cosmos DB SQL API

从版本 10.2.2 开始，PowerExchange for Microsoft Azure Cosmos DB SQL API 包含以下功能：

- 可以在 Azure Databricks 环境中运行映射。Databricks 对 PowerExchange for Microsoft Azure Cosmos DB SQL API 的支持可用于技术预览版。支持技术预览版功能，但不对此做出保证且尚未做好生产准备。Informatica 建议您只在非生产环境中使用这些功能。

有关详细信息，请参阅《*Informatica PowerExchange for Microsoft Azure Cosmos DB SQL API 10.2.2 用户指南*》。

PowerExchange for Microsoft Azure Data Lake Store

从版本 10.2.2 开始，PowerExchange for Microsoft Azure Data Lake Store 包含以下功能：

- 可以在 Azure Databricks 环境中运行映射。

- 可以在运行于 Hadoop 环境下的映射中使用复杂数据类型，如数组、结构和映射。使用复杂数据类型时，各个引擎可以直接在 Avro、JSON 和 Parquet 复杂文件中读取、处理和写入层次结构数据。对于智能结构源，可以仅配置读取操作。
- 可以在本地环境中将映射创建为读取和写入仅包含原始数据类型的 Avro 和 Parquet 文件。
- 可以选择一个目录作为映射中的源，以便从该目录读取多个文件。
- 可以在映射中使用 Microsoft Azure Data Lake Store 对象作为动态源和目标。
- 可以使用“创建目标”选项创建 Microsoft Azure Data Lake Store 目标。

有关详细信息，请参阅《*Informatica PowerExchange for Microsoft Azure Data Lake Store 10.2.2 用户指南*》。

PowerExchange for Microsoft Azure SQL 数据仓库

从版本 10.2.2 开始，PowerExchange for Microsoft Azure SQL 数据仓库包含以下功能：

- 可以在 Azure Databricks 环境中运行映射。
- 可以在运行于本地环境中或 Spark 引擎上的映射中配置美国政府 Microsoft Azure 端点。
- 可以在 Microsoft Azure Blob 存储容器中生成错误文件。错误文件包含已拒绝的行以及拒绝行的原因。
- 可以在本地环境的高级目标属性中定义批处理大小。
- 可以配置完整下推优化，以将转换逻辑推送至源数据库和目标数据库。通过使用数据库资源，利用下推优化改善任务性能。
- 可以在映射中使用 Microsoft Azure SQL 数据仓库对象作为动态源和目标。

PowerExchange for Microsoft Azure SQL 数据仓库的完整下推优化和动态映射功能可用于技术预览版。支持技术预览版功能，但不对此做出保证且尚未做好生产准备。Informatica 建议您只在非生产环境中使用这些功能。

有关详细信息，请参阅《*Informatica PowerExchange for Microsoft Azure SQL 数据仓库 10.2.2 用户指南*》。

PowerExchange for Salesforce

从版本 10.2.2 开始，PowerExchange for Salesforce 包含以下新功能：

- 您可以使用版本 43.0 和 44.0 的 Salesforce API 创建 Salesforce 连接以及访问 Salesforce 对象。
- 您可以为 Salesforce 连接配置 OAuth。
- 可以为源数据对象操作配置本地表达式筛选器。
- 可以参数化 Salesforce 数据对象的以下读取操作属性：
 - SOQL 筛选条件
 - CDC 开始时间戳
 - CDC 结束时间戳
 - PK 块大小
 - PK 块起始行 ID

可以参数化 Salesforce 数据对象的以下写入操作属性：

- 为 BULK 成功和错误文件设置前缀
- SFDC 成功文件目录
- 设置 BULK 错误文件的位置

有关详细信息，请参阅《*Informatica PowerExchange for Salesforce 10.2.2 用户指南*》。

PowerExchange for Snowflake

从版本 10.2.2 开始，PowerExchange for Snowflake 包含以下新功能：

- 您可以通过在 Snowflake 连接的 JDBC URL 参数中指定身份验证详细信息来配置 Okta SSO 身份验证。
- 可以配置 SQL 替代，以替代用于从 Snowflake 源提取数据的默认 SQL 查询。在 Snowflake 数据对象读取操作属性中指定 SQL 替代。
- 可以选择在将文件写入到 Snowflake 表之前压缩文件，并优化写入性能。在高级属性中。可以在 Snowflake 数据对象写入操作高级属性的**其他写入运行时参数**字段中将压缩参数设置为“打开”或“关闭”。
- 数据集成服务使用 Snowflake Spark Connector API 在 Spark 引擎上运行 Snowflake 映射。
- 您可以从 Azure 或 Amazon 中的暂存数据启用的 Snowflake 中读取数据，或者将数据写入其中。

有关详细信息，请参阅《*Informatica PowerExchange for Snowflake 10.2.2 用户指南*》。

PowerExchange for Teradata Parallel Transporter API

从版本 10.2.2 开始，PowerExchange for Teradata Parallel Transporter API 在高级目标属性中包含以下功能：

- 在将数据加载到目标时，可以指定替换字符以替代 Teradata 数据库中不支持的 Teradata unicode 字符。
- 如果在将数据加载到 Teradata 目标时指定用于替代不支持字符的字符，则可以为目标 Teradata 数据库指定版本 8.x - 13.x 或 14.x 和更高版本。此属性与“替换字符”属性结合使用。如果在将数据加载到 Teradata 目标时未指定替换字符，数据集成服务将忽略此属性。
- 将数据写入 Teradata 时，可以在运行时替代 Teradata 目标表架构和表名称。

有关详细信息，请参阅《*Informatica PowerExchange for Teradata Parallel Transporter API 10.2.2 用户指南*》。