



Informatica®
10.4.1.3

发行指南(10.4.1.3)

Informatica 发行指南(10.4.1.3)

10.4.1.3

2021 年 1 月

© 版权所有 Informatica LLC 2003, 2021

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。

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其他权利。

Informatica、Informatica 标志、PowerCenter、PowerExchange、Big Data Management 和 Live Data Map 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。

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

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

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

发布日期: 2021-01-29

目录

前言	28
Informatica 资源	28
Informatica Network	28
Informatica 知识库	28
Informatica 文档	28
Informatica 产品可用性矩阵	29
Informatica Velocity	29
Informatica Marketplace	29
Informatica 全球客户支持部门	29
第 I 部分： 版本 10.4.1.....	30
第 1 章： 通知和新增功能 10.4.1.3.....	31
支持更改(10.4.1.3)	31
Data Engineering	31
新增功能 (10.4.1.3)	31
Cloudera CDP	31
第 2 章： 通知、新增功能和更改 (10.4.1.2).....	32
通知(10.4.1.2)	32
已启动技术预览	32
新增功能 (10.4.1.2)	32
Data Engineering Integration	32
Data Privacy Management	33
Enterprise Data Catalog	33
安全	34
更改 (10.4.1.2)	35
Enterprise Data Catalog	35
第 3 章： 通知、新增功能和更改 (10.4.1.1).....	38
通知(10.4.1.1)	38
支持更改	38
新增功能 (10.4.1.1)	39
Data Engineering Integration	39
Data Privacy Management	39
Enterprise Data Catalog	40
更改 (10.4.1.1)	41
Data Privacy Management	41
Enterprise Data Catalog	42

第 4 章：通知 (10.4.1).....	43
新产品(10.4.1).	43
PowerExchange for MongoDB JDBC.	43
支持更改.	43
技术预览支持.	43
延期.	45
IPv6 支持.	45
发行版支持.	45
Enterprise Data Catalog.	46
安装程序更改.	46
发布任务.	46
Informatica PowerExchange 适配器.	46
PowerCenter PowerExchange 适配器.	47
第 5 章：新增功能 (10.4.1).....	48
连接(10.4.1).	48
DataDirect ODBC 和 JDBC 驱动程序.	48
Oracle Connection Manager.	48
Data Engineering Integration.	48
Hive 源和目标上的层次结构数据.	49
Mass Ingestion 的 SAML 发送验证.	49
Data Engineering Streaming.	49
ADLS Gen2 的 FileName 端口.	49
从多个 Kafka 主题引入 CDC 数据.	49
JDBC V2 查找转换.	49
复杂目标的 Parquet 数据格式.	49
Amazon S3 目标和 ADLS Gen2 目标中的滚存参数.	50
源和目标源和目标.	50
AWS Databricks 中的流映射.	50
Data Privacy Management.	51
数据存储.	51
作业.	51
隐私仪表板.	51
主题注册表数据加密.	52
结构化数据存储.	52
Enterprise Data Catalog.	52
业务术语关联.	52
为数据规则和列名称规则配置冲突解决方案.	52
上下文沿袭信息.	52
Data Asset Analytics.	53
通过导入和导入进行数据管理.	53
数据发现.	53

Enterprise Data Catalog 演练.	53
提取转换端点.	53
平面文件的字段级别沿袭.	54
云存储的文件沿袭.	54
Hive 资源.	54
Informatica MDM 资源.	55
Microsoft SQL Server 资源.	55
MicroStrategy 资源.	55
分区文件发现.	55
引用资源和引用资产.	55
资源级别权限.	55
提取数据库名称.	56
REST API.	56
SAP BW 和 SAP BW/4HANA 资源.	56
SSIS 资源.	56
Informatica 转换.	56
宏转换.	57
智能结构模型.	57
更改数组结构中的根节点.	57
Metadata Manager.	57
为业务对象资源配置 Java 位置和堆大小.	57
PowerExchange 适配器.	57
Informatica PowerExchange 适配器.	57
PowerCenter PowerExchange 适配器.	59
第 6 章：更改 (10.4.1).	62
命令行程序.	62
Data Engineering Integration.	63
Intelligent Structure Discovery.	63
Data Privacy Management.	63
Data Privacy Management 安装程序.	63
概览工作区.	63
扫描作业报告.	64
敏感文件页面视图.	64
文档.	64
Data Privacy Management.	64
Test Data Management.	64
域安全性.	65
启用对启用 Kerberos 的域的访问.	65
Enterprise Data Catalog.	65
更改摘要页面.	65
将资产数据导出到 Tableau 数据提取文件.	65
资产的导出路径.	65

显示活动业务术语.	65
对列和字段资产进行排序.	66
Catalog Administrator 中的开始选项卡名称.	66
Informatica Developer	66
关系对象搜索.	66
查看执行计划.	66
Metadata Manager.	66
以 HTML 5 格式查看数据沿袭.	66
Informatica PowerExchange 适配器.	67
PowerExchange for Amazon S3.	67
PowerExchange for Microsoft Azure Blob 存储.	67
PowerExchange for Microsoft Azure Data Lake Storage Gen1 and Gen2.	67
Test Data Management.	67
Test Data Management 安装程序.	67
 第 II 部分： 版本 10.4.0.	 68
 第 7 章： 新增功能 (10.4.0.2).	 69
更新群集发行版本.	69
 第 8 章： 通知、新增功能和更改 (10.4.0.1).	 70
通知 (10.4.0.1).	70
已提升技术预览.	70
新增功能 (10.4.0.1).	70
Data Engineering Integration.	70
Enterprise Data Catalog.	71
Informatica PowerExchange 适配器.	71
更改 (10.4.0.1).	71
Enterprise Data Catalog (10.4.0.1 更改)	71
Data Engineering Integration.	72
Informatica PowerExchange 适配器.	72
 第 9 章： 通知 (10.4.0).	 73
产品名称更改.	73
升级到版本 10.4.0.	73
支持更改.	74
技术预览支持.	74
延期.	75
删除的支持.	75
PowerCenter.	75
PowerCenter PowerExchange 适配器.	76
Informatica PowerExchange 适配器.	76
发布任务.	77

Data Engineering Integration.	77
第 10 章：新产品(10.4.0).	78
Informatica PowerExchange 适配器.	78
PowerExchange for JDBC V2.	78
PowerExchange for Microsoft Azure Data Lake Storage Gen2.	78
PowerExchange for Salesforce Marketing Cloud.	78
PowerCenter PowerExchange 适配器.	79
PowerExchange for Db2 Warehouse.	79
PowerExchange for Microsoft Dynamics 365 for Sales.	79
PowerExchange for PostgreSQL.	79
第 11 章：新增功能 (10.4.0).	80
CI/CD.	80
数据集成服务 REST API.	80
infacmd dis 命令.	81
反向代理服务器.	82
infacmd roh 命令.	82
应用程序修补程序部署.	82
连接到运行时应用程序.	83
对象浏览器视图.	83
标记.	84
命令行程序.	84
infacmd isp 命令 (10.4.0 新增功能)	85
Data Engineering Integration.	85
新数据类型支持.	85
AWS Databricks 集成.	86
通过 HDInsight 访问 ALDS Gen2 资源的群集工作流.	86
Databricks Delta Lake 存储访问.	86
显示映射使用的节点.	86
日志汇总.	86
在 Spark 引擎上解析层次结构数据.	87
Spark 引擎上的配置文件和采样选项.	87
Python 转换.	87
Sqoop.	88
Data Engineering Streaming.	88
流映射中的 Confluent 架构注册表.	88
流映射中的数据质量转换.	88
流映射中的临时群集.	89
Amazon S3 的 FileName 端口.	89
Microsoft Azure Data Lake Storage Gen2.	89
Azure Databricks 中的流映射.	89
Data Engineering Streaming 中的动态映射.	90

Enterprise Data Catalog.	90
为资源和类分配自定义属性.	90
新资源.	90
引用资源和引用资产.	91
从搜索结果导出资产.	91
沿袭和影响筛选器.	91
资产控制摘要.	91
规则和结果卡.	92
唯一键推理.	92
CLOB 文件类型上的数据域发现.	92
Spark 引擎上的数据发现和采样选项.	92
跟踪技术预览.	92
数据预览和置备.	93
独立扫描实用程序支持的资源类型.	93
REST API.	94
Enterprise Data Preparation.	94
Data Lake 访问管理.	94
Microsoft Azure Data Lake Storage 用作数据源.	94
将文件发布到 Data Lake.	94
将文件上传到 Data Lake.	94
Informatica 映射.	95
将映射输出绑定至映射参数.	95
CLAIRE 建议和见解.	96
更新映射优化器级别.	96
Informatica 转换.	96
地址验证器转换.	96
Informatica 工作流.	96
Amazon EMR 创建群集任务高级属性.	96
Informatica 安装.	97
PostgreSQL.	97
静默模式下的预安装 (i10Pi) 系统检查工具.	97
对静默安装属性文件的密码进行加密.	97
智能结构模型.	97
其他输入类型.	97
在设计时从示例创建模型.	97
处理无法识别的数据.	98
PowerCenter.	98
连接.	98
PowerExchange 适配器.	99
Informatica PowerExchange 适配器.	99
PowerCenter PowerExchange 适配器.	101
安全.	103

配置 Web 应用程序以使用不同的 SAML 身份提供程序.	103
第 12 章：更改 (10.4.0).	104
Data Engineering Integration.	104
数据预览.	104
Databricks 中的日期/时间格式.	105
目标中的空值.	105
Python 转换.	105
Enterprise Data Preparation.	106
空值处理.	106
Solr 端口属性.	106
Enterprise Data Catalog.	106
SAP HANA 资源的沿袭增强功能.	106
剖析和数据域发现.	106
Informatica Data Engineering Streaming.	107
资源名称更改.	107
搜索建议.	107
配置自定义属性.	107
Informatica Developer.	107
导入关系数据对象.	107
Informatica 转换.	107
地址验证器转换.	107
PowerCenter.	108
刷新 Designer 和 Workflow Manager 中的元数据.	108
导入和导出.	108
Informatica PowerExchange 适配器.	109
PowerExchange for Amazon Redshift.	109
PowerExchange for Amazon S3.	109
PowerExchange for Microsoft Azure Blob 存储.	110
PowerExchange for Microsoft Azure Data Lake Storage Gen1.	110
安全.	110
infacmd isp 命令.	110
LDAP 目录服务.	110
LDAP 配置.	111
SAML 身份验证.	111
第 III 部分：版本 10.2.2.	112
第 13 章：通知、新增功能和更改 (10.2.2 HotFix 1).	113
通知 (10.2.2 HotFix 1).	113
支持更改.	113
新增功能 (10.2.2 HotFix 1).	114
命令行程序.	114

Big Data Management.	114
Enterprise Data Catalog.	114
Informatica PowerExchange 适配器.	117
更改 (10.2.2 HotFix 1).	118
Data Transformation.	118
Enterprise Data Preparation.	118
Enterprise Data Catalog.	118
Informatica PowerExchange 适配器.	119
第 14 章：通知、新增功能和更改 (10.2.2 Service Pack 1).	121
通知 (10.2.2 Service Pack 1).	121
支持更改.	121
产品和服务名称更改.	122
发布任务.	122
新增功能 (10.2.2 Service Pack 1).	122
Big Data Management.	122
大数据 Streaming.	123
Enterprise Data Catalog.	123
Enterprise Data Preparation.	124
Informatica PowerExchange 适配器.	124
更改 (10.2.2 Service Pack 1).	125
Big Data Management.	125
大数据 Streaming.	125
Informatica Analyst.	126
Informatica PowerExchange 适配器.	126
第 15 章：通知 (10.2.2).	128
安装和配置.	128
OpenJDK.	128
安装程序代码签名.	128
恢复安装程序.	129
Informatica Docker 实用程序.	129
安装程序.	129
支持更改.	129
Hive 引擎.	129
发行版支持.	130
Big Data Management 的支持更改.	130
Big Data Streaming 的支持更改.	130
Enterprise Data Catalog 中的通用连接框架.	131
发布任务.	131
Big Data Management.	131
大数据 Streaming.	132
Informatica PowerExchange 适配器.	133

第 16 章：新产品(10.2.2).....	134
PowerExchange 适配器.....	134
Informatica PowerExchange 适配器.....	134
第 17 章：新增功能 (10.2.2).....	135
应用程序服务.....	135
Mass Ingestion 服务.....	135
元数据访问服务.....	135
REST 操作 Hub 服务.....	135
Big Data Management.....	136
Azure Databricks 集成.....	136
Spark 引擎上的数据预览.....	137
层次结构数据.....	137
高可用性.....	138
智能结构模型.....	138
Mass Ingestion.....	139
监视.....	139
安全.....	139
目标.....	139
大数据 Streaming.....	140
Azure 事件中心数据对象.....	140
Amazon Kinesis 连接中的跨帐户 IAM 角色.....	140
智能结构模型.....	140
Big Data Streaming 数据对象的标头端口.....	140
Amazon Kinesis 连接中的 AWS 凭据配置文件.....	140
Spark 结构化流.....	141
窗口转换.....	141
命令行程序.....	141
infacmd dis 命令.....	141
infacmd ihs 命令.....	142
infacmd ipc 命令.....	142
infacmd ldm 命令.....	142
infacmd mi 命令.....	143
infacmd ms 命令.....	144
infacmd oie 命令.....	144
infacmd tools 命令.....	144
infasetup 命令.....	145
Enterprise Data Catalog.....	145
为列自动分配业务标题.....	145
资产方面的用户协作.....	145
使用安装程序创建 Enterprise Data Catalog 应用程序服务.....	146
自定义元数据验证实用程序.....	146

更改通知.	146
Business Glossary 分配报告.	146
操作系统配置文件.	146
REST API.	146
源元数据和数据配置文件筛选器.	146
扫描程序实用程序.	147
资源类型.	147
Enterprise Data Lake.	147
应用活动规则.	147
删除重复行.	148
聚集和分类列数据.	148
基于 CLAIRE 的建议.	148
条件性汇总.	148
数据屏蔽.	148
本地化.	148
分区源和目标.	149
向脚本步骤添加注释.	149
将脚本另存为映射.	149
Amazon S3、ADLS、WASB、MapR-FS 作为数据源.	149
统计函数.	149
日期和时间函数.	150
数学函数.	150
文本函数.	150
窗口函数.	151
清除审计事件.	151
Spark 执行引擎.	151
Informatica Developer.	152
应用程序.	152
Informatica 映射.	152
数据类型.	152
映射输出.	152
映射参数.	152
优化器级别.	152
Sqoop.	153
Informatica 转换.	153
地址验证器转换.	153
更新策略转换.	156
Informatica PowerExchange 适配器.	156
PowerExchange for Amazon Redshift.	156
PowerExchange for Amazon S3.	157
PowerExchange for Google BigQuery.	157
PowerExchange for HBase.	157

PowerExchange for HDFS.	158
PowerExchange for Hive.	158
PowerExchange for MapR-DB.	159
PowerExchange for Microsoft Azure Blob 存储.	159
PowerExchange for Microsoft Azure Cosmos DB SQL API.	159
PowerExchange for Microsoft Azure Data Lake Store.	159
PowerExchange for Microsoft Azure SQL 数据仓库.	160
PowerExchange for Salesforce.	160
PowerExchange for Snowflake.	161
PowerExchange for Teradata Parallel Transporter API.	161
第 18 章：更改 (10.2.2).	162
应用程序服务.	162
元数据访问服务.	162
Big Data Management.	162
Hive 连接.	162
Mass Ingestion.	163
Spark 监视.	163
Sqoop.	163
Hadoop 环境中的转换.	164
大数据 Streaming.	164
Big Data Streaming 与 Big Data Management 集成.	164
Kafka 连接.	165
转换.	165
Enterprise Data Catalog.	165
Java Development Kit 更改.	165
Enterprise Data Lake.	165
MAX 和 MIN 函数.	166
Informatica Developer.	166
Informatica Developer 名称更改.	166
Informatica 转换.	166
地址验证器转换.	166
写入转换.	167
Informatica PowerExchange 适配器.	167
PowerExchange for Amazon Redshift.	168
PowerExchange for Amazon S3.	168
PowerExchange for Google Analytics.	168
PowerExchange for Google Cloud Storage.	169
PowerExchange for HBase.	169
PowerExchange for HDFS.	169
PowerExchange for Hive.	169
PowerExchange for Microsoft Azure Blob 存储.	169

第 IV 部分： 版本 10.2.1.....	170
第 19 章： 新增功能 (10.2.1).....	171
应用程序服务.....	171
内容管理服务.....	171
数据集成服务.....	172
Mass Ingestion 服务.....	172
元数据访问服务.....	172
模型存储库服务.....	172
Big Data Management.....	173
Blaze 引擎资源节约.....	173
群集工作流.....	173
云置备配置.....	173
高可用性.....	174
Hadoop 环境中的 Hive 功能.....	174
从 PowerCenter 导入.....	174
智能结构模型.....	175
Mass Ingestion.....	175
监视.....	175
在 Spark 引擎上处理层次结构数据.....	177
Spark 引擎上的规则规范支持.....	177
安全.....	177
Sqoop.....	178
Hadoop 环境中的转换支持.....	178
Big DataStreaming.....	180
源和目标.....	180
流映射中的有状态计算.....	181
转换支持.....	181
截断已分区的 Hive 目标表.....	181
命令程序.....	181
infacmd autotune 命令.....	181
infacmd ccps 命令.....	181
infacmd cluster 命令.....	182
infacmd cms 命令.....	182
infacmd dis 命令.....	183
infacmd ihs 命令.....	183
infacmd isp 命令.....	183
infacmd ldm 命令.....	183
infacmd mi 命令.....	184
infacmd mrs 命令.....	184
infacmd wfs 命令.....	184
infasetup 命令.....	185

Enterprise Data Catalog.	185
向资产添加业务标题.	185
安装程序中的群集验证实用程序.	185
数据域发现类型.	185
筛选器设置.	186
缺少的链接报表.	186
新的资源类型.	186
REST API.	186
Enterprise Data Catalog 应用程序的 SAML 身份验证.	186
SAP 资源.	187
从 ServiceNow 导入.	187
类似列.	187
指定目录服务的加载类型.	187
数据发现支持的资源类型.	187
Enterprise Data Lake.	188
列数据.	188
管理 Data Lake 资源.	188
数据准备操作.	188
准备 JSON 文件.	188
脚本步骤.	189
计划导出、导入和发布活动.	189
安全断言标记语言身份验证.	189
查看项目流程和项目历史记录.	189
Informatica Developer.	189
默认布局.	189
编辑器搜索.	190
从 PowerCenter 导入会话属性.	190
视图.	191
Informatica 映射.	191
动态映射.	191
映射参数.	192
运行映射.	193
截断已分区的 Hive 目标表.	193
Informatica 转换语言.	194
适用于 map 数据类型的复杂函数.	194
适用于 map 数据类型的复杂运算符.	194
Informatica 转换.	194
地址验证器转换.	194
Informatica 工作流.	196
从 PowerCenter 导入命令任务.	197
Informatica PowerExchange 适配器.	197
PowerExchange for Amazon Redshift.	197

PowerExchange for Amazon S3.	197
PowerExchange for Cassandra.	197
PowerExchange for HBase.	198
PowerExchange for HDFS.	198
PowerExchange for Hive.	198
PowerExchange for Microsoft Azure Blob 存储.	198
PowerExchange for Microsoft Azure SQL 数据库.	199
PowerExchange for Salesforce.	199
PowerExchange for SAP NetWeaver.	199
PowerExchange for Snowflake.	199
安全.	200
密码复杂度.	200
 第 20 章：更改 (10.2.1).	 201
支持更改.	201
升级支持更改.	201
Big Data Hadoop 发行版支持.	202
Hive 运行时引擎.	203
安装程序更改.	204
产品名称更改.	205
应用程序服务.	205
模型存储库服务.	206
Big Data Management.	206
Azure 存储访问.	206
配置 Hadoop 发行版.	206
Developer tool 配置.	207
Hadoop 连接更改.	207
Hive 连接属性.	208
监视.	209
Hive 引擎上的精度和小数位.	209
Sqoop.	209
Hive 引擎上的转换支持.	210
Big DataStreaming.	210
配置 Hadoop 发行版.	210
Developer tool 配置.	210
Kafka 连接属性.	211
命令程序.	211
内容安装程序.	212
Enterprise Data Catalog.	212
常规选项卡中的附加属性部分.	212
连接分配.	212
列相似性.	212
创建目录服务.	212

HDFS 资源类型增强功能.	212
Hive 资源.	213
Informatica 平台扫描程序.	213
概览选项卡.	213
产品名称更改.	213
邻近数据域.	213
搜索结果.	214
通用连接框架.	214
Informatica Analyst.	214
结果卡.	214
Informatica Developer.	214
从 PowerCenter 导入对象以及将对象导出到其中.	214
Informatica 转换.	215
地址验证器转换.	215
数据转换.	215
序列生成器转换.	215
排序器转换.	215
Informatica PowerExchange 适配器.	216
PowerExchange for Amazon Redshift.	216
PowerExchange for Cassandra.	216
PowerExchange for Snowflake.	217
第 21 章：发布任务 (10.2.1).	218
Informatica PowerExchange 适配器.	218
适用于 Amazon S3 的 PowerExchange 适配器.	218
第 V 部分：版本 10.2.	219
第 22 章：新功能、更改和发布任务 (10.2 HotFix 2).	220
支持更改(10.2 HotFix 2).	220
验证 Hadoop 发行版支持.	220
OpenJDK.	221
DataDirect SQL Server 旧版 ODBC 驱动程序.	221
PowerExchange for SAP NetWeaver.	221
新产品(10.2 HotFix 2).	221
PowerCenter PowerExchange 适配器.	221
新增功能 (10.2 HotFix 2).	221
命令行程序.	221
Informatica Analyst.	222
Informatica 转换.	222
Metadata Manager.	224
PowerCenter.	224
PowerCenter PowerExchange 适配器.	225

安全.	227
更改 (10.2 HotFix 2).	227
Analyst 工具.	227
命令程序.	227
Informatica 转换.	228
Metadata Manager.	228
PowerCenter.	228
Informatica PowerExchange 适配器.	228
PowerCenter PowerExchange 适配器.	229
发布任务(10.2 HotFix 2).	230
Informatica PowerExchange 适配器.	230
PowerCenter PowerExchange 适配器.	230
 第 23 章：新功能、更改和发布任务 (10.2 HotFix 1).	 231
新增功能 (10.2 HotFix 1).	231
应用程序服务.	231
Business Glossary	231
命令程序.	232
连接.	232
数据类型.	232
安装程序.	233
Informatica 转换.	233
Metadata Manager.	236
PowerCenter.	236
Informatica PowerExchange 适配器.	237
PowerCenter PowerExchange 适配器.	238
安全.	241
更改 (10.2 HotFix 1).	241
支持更改.	241
应用程序服务.	242
Big Data Management.	242
Business Glossary.	243
文档.	243
Informatica Development Platform.	243
Informatica 转换.	243
PowerCenter.	244
PowerCenter PowerExchange 适配器.	244
引用数据.	246
发布任务 (10.2 HotFix 1).	247
Informatica PowerExchange 适配器.	247
PowerCenter PowerExchange 适配器.	247

第 24 章：新产品 (10.2).....	249
PowerExchange 适配器.....	249
Informatica PowerExchange 适配器.....	249
第 25 章：新增功能 (10.2).....	250
应用程序服务.....	250
模型存储库服务.....	250
大数据.....	251
Big Data Management 安装.....	251
群集配置.....	251
处理层次结构数据.....	251
Spark 引擎上的有状态计算.....	252
数据集成服务队列.....	252
Blaze 作业监视.....	252
适用于 Hadoop 集成的数据集成服务属性.....	253
Sqoop.....	253
在 Amazon EMR 群集中自动扩展.....	253
Blaze 引擎上的转换支持.....	253
Blaze 引擎的 Hive 功能.....	254
Spark 引擎上的转换支持.....	254
Spark 引擎的 Hive 功能.....	254
命令行程序.....	254
infacmd cluster 命令.....	254
infacmd dis 选项.....	255
infacmd ipc 命令.....	256
infacmd isp 命令.....	256
infacmd mrs 命令.....	260
infacmd ms 命令.....	260
infacmd wfs 命令.....	260
infasetup 命令.....	261
pmrep 命令.....	261
数据类型.....	262
Informatica 数据类型.....	262
文档.....	262
Enterprise Information Catalog.....	263
新数据源.....	264
自定义扫描器框架.....	264
REST API.....	264
复合数据域.....	264
数据域.....	265
导出和导入自定义属性.....	265
将富文本作为自定义属性值.....	265

转换逻辑	266
非结构化文件类型.	266
值频率.	266
对 Azure HDInsight 部署的支持.	266
Informatica Analyst.	266
配置文件.	267
Intelligent Data Lake.	267
使用可视化及 Apache Zeppelin 验证和评估数据.	267
在数据预览期间使用筛选器评估数据.	267
增强的脚本面板布局.	267
应用数据质量规则.	267
在数据预览和工作表视图中查看数据资产的业务术语.	268
为带分隔符的文件准备数据.	268
在联接工作表中编辑联接.	268
编辑数据准备的采样设置.	268
数据湖中支持多个 Enterprise Information Catalog 资源.	268
为数据准备服务存储库使用 Oracle.	268
提高了数据准备服务的可扩展性.	268
Informatica Developer.	268
非关系数据对象.	269
配置文件.	269
Informatica 安装.	269
Informatica Upgrade Advisor.	269
Intelligent Streaming.	269
CSV 格式.	269
数据类型.	269
连接.	270
传递映射.	270
源和目标.	270
转换支持.	270
Metadata Manager.	270
Cloudera 导航器.	270
PowerCenter.	271
PowerExchange 适配器.	271
Informatica PowerExchange 适配器.	271
PowerCenter PowerExchange 适配器.	274
规则规范.	275
安全.	275
用户活动日志.	275
转换语言.	276
Informatica 转换语言.	276
转换.	277

Informatica 转换.	277
PowerCenter 转换.	280
工作流.	280
Informatica 工作流.	280
第 26 章：更改 (10.2).	282
支持更改.	282
Big Data Hadoop 发行版支持.	283
Metadata Manager.	285
应用程序服务.	285
内容管理服务.	285
数据集成服务.	285
大数据.	286
Hadoop 连接.	286
HBase 连接属性.	288
Hive 连接属性.	289
适用于 MapR-DB 的 HBase 连接属性.	289
映射运行时属性.	290
监视.	290
S3 访问和密钥属性.	290
Sqoop.	290
命令程序.	291
Enterprise Information Catalog.	291
产品名称更改.	292
Informatica Analyst.	292
参数.	292
Intelligent Streaming.	292
Kafka 数据对象更改.	292
PowerExchange 适配器.	292
Informatica PowerExchange 适配器.	293
PowerCenter PowerExchange 适配器.	293
安全.	294
SAML 身份验证.	295
转换.	295
Informatica 转换.	295
工作流.	296
Informatica 工作流.	296
第 27 章：发布任务 (10.2).	297
PowerExchange 适配器.	297
PowerCenter PowerExchange 适配器.	297

第 VI 部分： 版本 10.1.1.....	299
第 28 章： 新功能、更改和发布任务 (10.1.1 HotFix 1).....	300
新产品 (10.1.1 HotFix 1).	300
PowerExchange for Cloud Applications.	300
新功能 (10.1.1 HotFix 1).	300
命令行程序.	300
Informatica Analyst.	301
PowerCenter.	302
PowerExchange 适配器.	302
更改 (10.1.1 HotFix 1).	303
支持更改.	304
第 29 章： 新功能、更改和发布任务 (10.1.1 Update 2).....	305
新产品 (10.1.1 Update 2).	305
PowerExchange for MapR-DB.	305
新功能 (10.1.1 Update 2).	305
Big Data Management.	305
Enterprise Information Catalog.	307
Intelligent Data Lake.	307
Informatica PowerExchange 适配器.	308
更改 (10.1.1 Update 2).	308
支持更改.	308
Big Data Management.	310
Enterprise Information Catalog.	310
Informatica PowerExchange 适配器.	311
第 30 章： 新功能、更改和发布任务 (10.1.1 Update 1).....	312
新功能 (10.1.1 Update 1).	312
Big Data Management.	312
更改 (10.1.1 Update 1).	312
Informatica PowerExchange 适配器.	313
发布任务 (10.1.1 Update 1).	313
Informatica PowerExchange 适配器.	313
第 31 章： 新产品 (10.1.1).....	314
智能流.	314
第 32 章： 新功能 (10.1.1).....	316
应用程序服务.	316
分析服务.	316
大数据.	317

Blaze 引擎.	317
安装和配置.	318
Spark 引擎.	319
安全.	320
Sqoop.	320
Business Glossary	320
将富文本导出为纯文本.	321
包含冲突资产的富文本内容.	321
命令行程序.	321
infacmd as 命令.	321
infacmd dis 命令.	322
infacmd mrs 命令.	322
pmrep 命令.	322
Enterprise Information Catalog.	322
Business Glossary 集成.	323
列相似性剖析.	323
数据域和数据域组.	324
沿袭和影响分析.	324
用户和用户组的权限.	324
新的资源类型.	324
同义词定义文件.	325
通用连接框架.	325
Informatica Analyst.	325
配置文件.	325
Informatica 安装.	325
Informatica Upgrade Advisor.	325
Intelligent Data Lake.	326
外部源中的表的数据预览.	326
从外部源的表中导入数据.	326
将数据导出到外部目标.	326
为数据准备配置采样标准.	326
对工作表执行查找.	326
下载为 TDE 文件.	326
Sentry 和 Ranger 支持.	327
映射.	327
Informatica 映射.	327
Metadata Manager.	327
Cloudera Navigator 资源的数据集提取.	327
Informatica Platform 资源的映射提取.	327
PowerExchange 适配器.	328
Informatica PowerExchange® 适配器.	328
PowerCenter® PowerExchange 适配器.	328

安全.	329
自定义 Kerberos 库.	329
已启用 Kerberos 的域中的计划程序服务支持.	330
Informatica Web 应用程序的单点登录.	330
转换.	330
Informatica 转换.	330
Web 服务.	333
Informatica Web 服务.	333
工作流.	333
Informatica 工作流.	333
第 33 章：更改 (10.1.1).	335
支持更改.	335
Big Data Management Hive 引擎.	335
支持更改 - Big Data Management Hadoop 发行版.	336
Big Data Management Spark 支持.	336
Data Analyzer.	336
操作系统.	337
PowerExchange for SAP NetWeaver.	337
报告和仪表板服务.	337
报告服务.	337
大数据.	337
Hadoop 环境中支持的函数.	337
Hadoop 配置管理器.	338
Business Glossary.	338
导出文件限制.	338
数据集成服务.	339
数据类型.	339
Informatica 数据类型.	339
Informatica Analyst.	340
配置文件.	340
Informatica Developer.	340
配置文件.	340
映射.	340
Informatica 映射.	340
Enterprise Information Catalog.	341
HDFS 扫描器增强.	341
关系视图.	341
Metadata Manager.	341
Cloudera 导航器资源.	341
Netezza 资源.	342
PowerExchange 适配器.	342
Informatica PowerExchange 适配器.	342

PowerCenter PowerExchange 适配器	343
转换.	343
Informatica 转换.	343
工作流.	344
Informatica 工作流.	344
文档.	344
Metadata Manager 文档.	344
PowerExchange for SAP NetWeaver 文档.	344
 第 34 章：发布任务 (10.1.1).	 345
Metadata Manager.	345
业务情报资源.	345
Cloudera 导航器资源.	345
Tableau 资源.	346
 第 VII 部分：版本 10.1.	 347
 第 35 章：新产品 (10.1).	 348
Intelligent Data Lake.	348
PowerExchange 适配器.	350
Informatica PowerExchange 适配器.	350
 第 36 章：新增功能 (10.1).	 352
应用程序服务.	352
系统服务.	353
大数据.	353
Hadoop 生态系统.	353
Hadoop 安全系统.	353
Spark 运行时引擎.	353
针对关系源和目标的 Sqoop 连接.	354
Blaze 引擎上的转换支持.	354
Business Glossary.	355
将词汇表内容管理者继承给所有资产.	355
双向自定义关系.	355
关系视图图表中的自定义颜色.	355
连接.	355
IBM DB2 连接中的架构名称.	355
命令行程序.	356
文档.	360
异常管理.	360
Informatica Administrator.	361
域视图.	361
监视.	361

Informatica Analyst.	362
配置文件.	362
Informatica Developer.	363
生成源文件名.	363
从 PowerCenter 导入.	363
在 Excel 和 Developer Tool 之间复制文本.	363
逻辑数据对象读取和写入映射编辑.	363
DDL 查询.	363
配置文件.	364
Informatica Development Platform.	364
Live Data Map.	365
电子邮件通知.	366
关键字搜索.	366
剖析.	366
扫描器.	366
映射.	366
Informatica 映射.	366
Metadata Manager.	367
通用资源.	367
适用于 Oracle 资源和 Teradata 资源的增量加载.	367
隐藏摘要视图中的资源.	367
从多个包文件创建 SQL Server Integration Services 资源.	368
Metadata Manager 命令程序.	368
应用程序属性.	368
将 Business Glossary 审计跟踪的历史记录和链接迁移到技术元数据.	368
PowerCenter.	369
PowerExchange 适配器.	369
Informatica PowerExchange 适配器.	369
PowerCenter PowerExchange 适配器.	369
安全性.	370
转换.	370
Informatica 转换.	370
工作流.	372
PowerCenter 工作流.	372
第 37 章：更改 (10.1).	373
支持更改.	373
应用程序服务.	374
系统服务.	374
大数据.	374
Business Glossary.	375
自定义关系.	375
双向默认关系.	375

控制者关系.	375
词汇表工作区.	375
Business Glossary Desktop.	375
Business Glossary 命令程序的 Kerberos 身份验证.	375
命令行程序.	376
异常管理.	376
Informatica Developer.	377
Live Data Map.	377
Enterprise Information Catalog.	377
“Live Data Map 管理员” 主页.	377
Metadata Manager.	378
Microsoft SQL Server Integration Services 资源.	378
命令行程序的证书验证.	378
PowerCenter.	378
安全.	379
转换.	379
Informatica 转换.	379
工作流.	380
Informatica 工作流.	380
 第 38 章：发布任务 (10.1).	 382
Metadata Manager.	382
Informatica Platform 资源.	382
验证命令程序的信任库文件.	382
安全性.	383
权限.	383

前言

请参阅《*Informatica® 发行指南*》了解当前和最新产品版本中的新功能与增强功能。了解各版本之间的行为更改，以及从以前的版本升级后需要执行的任务。《发行指南》包括 Data Engineering 产品和传统产品的内容。

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 I 部分： 版本 10.4.1

本部分包含以下章节：

- [通知和新增功能 10.4.1.3, 31](#)
- [通知、新增功能和更改 \(10.4.1.2\), 32](#)
- [通知、新增功能和更改 \(10.4.1.1\), 38](#)
- [通知 \(10.4.1\), 43](#)
- [新增功能 \(10.4.1\), 48](#)
- [更改 \(10.4.1\), 62](#)

第 1 章

通知和新增功能 10.4.1.3

本章包括以下主题：

- [支持更改\(10.4.1.3\), 31](#)
- [新增功能 \(10.4.1.3\), 31](#)

支持更改(10.4.1.3)

本节介绍版本 10.4.1.3 中的支持更改。

Data Engineering

从版本 10.4.1.3 开始，Informatica 支持以下功能和集成：

已提升技术预览

支持 Amazon EMR 6.x

您可以将 Informatica 与 Amazon EMR 6.x 集成。以前，此支持仅作为技术预览提供。有关详细信息，请参阅《*Data Engineering 10.4.1 集成指南*》。

新增功能 (10.4.1.3)

本节介绍版本 10.4.1.3 中新增的功能。

Cloudera CDP

从版本 10.4.1.3 开始，对于 Data Engineering，您可以使用 Blaze 引擎在 Cloudera CDP Private Cloud 和 Cloudera CDP Public Cloud 集群上运行作业。

当 Blaze 引擎在 CDP Private Cloud 群集上运行作业时，您可以使用 Kerberos 身份验证。当 Blaze 引擎在 CDP Public Cloud 群集上运行作业时，您可以使用 Apache Knox 或 Kerberos 身份验证。您可以对 Blaze 引擎在 CDP 群集上运行的作业使用 Apache Ranger、HDFS 透明加密或 SSL/TLS 授权。

有关详细信息，请参阅《*Data Engineering 10.4.1 集成指南*》。

第 2 章

通知、新增功能和更改 (10.4.1.2)

本章包括以下主题：

- [通知\(10.4.1.2\), 32](#)
- [新增功能 \(10.4.1.2\), 32](#)
- [更改 \(10.4.1.2\), 35](#)

通知(10.4.1.2)

已启动技术预览

从版本 10.4.1.2 开始，Informatica 包含以下技术预览功能：

Amazon EMR 6.0

您可以将 Informatica Data Engineering 产品与 Amazon EMR 6.0 群集一起使用。

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。

新增功能 (10.4.1.2)

Data Engineering Integration

本节介绍版本 10.4.1.2 中新增的 Data Engineering Integration 功能。

Cloudera CDP Public Cloud

从版本 10.4.1.2 开始，您可以在 AWS 或 Azure 上的 Cloudera CDP Public Cloud 群集上运行数据工程作业。

在 Spark 引擎或 Blaze 引擎上运行数据工程作业时，可以使用 Cloudera CDP Public Cloud。在 Hadoop 环境中运行映射时，可以对 CDP Public Cloud 使用与 CDP Public Cloud 相同的连接。

Data Privacy Management

本节介绍版本 10.4.1.2 中新增的 Data Privacy Management 功能。

非结构化源上的增量主题注册表扫描

从版本 10.4.1.2 开始，您可以为非结构化源配置增量主题注册表扫描。

增量扫描将更新上次扫描运行后更改的记录，而不是完全扫描所有记录。

有关如何配置增量主题注册表扫描的信息，请参阅《《Informatica Data Privacy Management 10.4.1 用户指南》》和《《Informatica Data Privacy Management 10.4.1 管理员指南》》。

Enterprise Data Catalog

本节介绍版本 10.4.1.2 中新增的 Enterprise Data Catalog 功能。

Amazon S3 资源

从版本 10.4.1.2 开始，Amazon S3 资源包括以下增强功能：

元数据提取

您可以将 Amazon S3 资源配置为从兼容 Amazon S3 的存储（如 Scality RING）中提取元数据。

支持临时会话标志

您可以将 Amazon S3 资源配置为使用临时会话标志连接 Amazon S3 数据源并从中提取元数据。

有关详细信息，请参阅《Informatica 10.4.1 Enterprise Data Catalog 扫描程序配置指南》。

Informatica Axon 资源

从版本 10.4.1.2 开始，Informatica Axon 资源包括以下增强功能：

元数据提取

Informatica Axon 资源从以下资产中提取元数据：

- 属性
- 策略
- 系统
- 数据集

您还可以读取与目录中的资产关联的本地用户、LDAP 用户和 SSO 用户的详细信息。

查看关系

您可以在目录中查看以下关系：

- Axon 项目与提取对象之间的父子关系
- 词汇表资产之间的语义关系

支持协调域用户

您可以通过在 Catalog Administrator 中为电子邮件系统属性选择**允许筛选**选项来协调域用户。

有关详细信息，请参阅《Informatica 10.4.1 Enterprise Data Catalog 扫描程序配置指南》。

通过导入和导出批量管理业务术语

从版本 10.4.1.2 开始，您可以在导出的 .csv 文件中同时接受或拒绝多个推理的业务术语。

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 用户指南*》。

批量管理数据域

从版本 10.4.1.2 开始，您可以接受数据域，并拒绝列或字段资产的所有其他已分配、已推理或已接受的数据域。

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 用户指南*》。

Data Asset Analytics

从版本 10.4.1.2 开始，Data Asset Analytics 包括以下功能：

资产使用情况报告

您可以使用**资产使用情况**报告查看 Enterprise Data Catalog 中使用的资产列表。您可以根据资产类型、事件日期、资源类型和资源名称来筛选报告。除了资产使用情况详细信息外，该报告还会列出您在访问资产时所使用的 Enterprise Data Catalog 功能。除了与资产关联的其他功能外，列出的功能还包括**概览**、**沿袭**和**影响**以及**关系**选项卡。

“最常查看的资产”图表

用户采用选项卡页上的**最常查看的资产**图表显示了目录中查看次数最多的资产的列表。

“功能使用情况值”图表

数据值选项卡页上的**功能使用情况值**图表根据 Enterprise Data Catalog 功能在一段时间内的使用情况显示其值。

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 用户指南*》。

SAP S/4HANA 资源

从版本 10.4.1.2 开始，您可以使用 SAP S/4HANA 资源查看功能模块和数据源资产之间的详细沿袭。

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 扫描程序配置指南*》。

安全

本节介绍版本 10.4.1.2 中的新安全功能。

HSTS 响应标头

从版本 10.4.1.2 开始，您可以选择在域中启用 HTTP Strict Transport Security HSTS 响应标头，以停止将 HTTP 重定向到 HTTPS。这可以防止中间人 (MITM) 安全威胁。

有关详细信息，请参阅《*Informatica 安全指南*》。

更改 (10.4.1.2)

Enterprise Data Catalog

本节介绍版本 10.4.1.2 中对 Enterprise Data Catalog 进行的更改。

Data Asset Analytics

从版本 10.4.1.2 开始，Data Asset Analytics 包括以下更改：

资源值计算

在**数据值**选项卡页上，通过添加以下值来计算**资源值**：

- 将指定的资源类型值乘以基于指定的资产关键字选择的资产总数。
- 将为所有其他资源类型指定的值乘以基于指定的资产关键字选择的资产总数。

注意：可以在**应用程序配置**的**资源值计算的资产关键字**文本框中指定资产关键字。

应用版本 10.4.1.2 后，您将丢失先前配置的所有与资源相关的配置和值。回滚不会保留您先前配置的值。资源类型的值和配置将保留，并且您可能会看到根据更新后的计算方法计算的新值有所增加。

报告名称中的特殊字符

对于新报告，可以在报告名称中使用以下特殊字符：-（连字符）、@（at 符号）和 _（下划线）。在早期版本中创建的包含不受支持的特殊字符的报告仍然存在，也仍会有现有的问题。如果您复制包含不支持的特殊字符的现有报告来创建新报告，则 Data Asset Analytics 将显示验证错误。

资源值计算的资产关键字

在**应用程序配置**中，可以指定的资产关键字，使用这些关键字可以从目录中选择资产以计算资源值。

货币

在**应用程序配置**中，可以选择在**数据值**选项卡页上查看资产值所用的货币。

“自动扩充”列

资产扩充、**数据域关联**和**业务术语关联**报告中的**自动扩充**列会根据资产的扩充方式列出以下详细信息：

- 已推理。根据配置文件结果推荐的资产扩充。
- 已推理-已接受。用户已接受推理出的扩充。
- 已推理-已自动接受。已根据指定的遵从性分数自动接受推理出的扩充。
- 已推理-已拒绝。用户已拒绝推理出的扩充。
- 手动。资产的扩充是手动完成的。

注意：以前，报告会在**自动扩充**列中列出 Y、否或 I。

“自动扩充”筛选器

资产扩充报告的**自动扩充**筛选器会列出可用于筛选报告的以下选项：

- 已推理
- 已推理-已接受
- 已推理-已自动接受
- 已推理-已拒绝
- 手动

注意: 以前, **自动扩充** 筛选器会列出是或否作为选项。

刷新

监视 页面包含一个 **刷新** 图标, 您可以单击该图标来刷新页面并查看报告运行的当前状态。

时间范围

分析 主页上 **时间范围** 下拉列表中的默认选项从 **从开始** 更改为 **1 年**。

“资产组类型” 筛选器和列

资产组类型 筛选器和报告列已从下列报告中删除:

报告	已从筛选器、报告列或这两者中删除
资产扩充	两者
资产计数	报告列
具有沿袭和影响的资产	报告列
业务术语关联	报告列
数据域关联	两者

注意: 对于使用早期版本创建的报告, 下载报告时将保留 **资产组类型** 报告列。

资源值计算器

资源类型 和 **资源** 选项已从 **资源值计算器** 的 **将值分配给** 部分中删除。

有关详细信息, 请参阅 《*Informatica 10.4.1 Enterprise Data Catalog 用户指南*》。

部分目录备份

从版本 10.4.1.2 开始, 您可以根据需要对部分目录进行备份来解决特定问题。

您可以在 backupContents 命令中使用 -st 选项来备份目录中的以下一个或多个数据存储:

- 暂存详细信息
- 图形数据库
- 搜索详细信息
- 相似性详细信息
- 更改通知详细信息
- 扫描程序暂存详细信息

您还可以使用 backup REST API 来备份目录中的以下一个或多个数据存储:

- 暂存详细信息
- 图形数据库
- 搜索详细信息

有关详细信息, 请参阅 《*Informatica 10.4.1 Enterprise Data Catalog REST API 引用*》 和 《*Informatica 10.4.1 命令引用*》。

SAP S/4HANA、SAP BW 和 SAP BW/4HANA 资源

从版本 10.4.1.2 开始，如果资产类型不包括沿袭、层次结构、类别和关系信息，则无法在目录中查看 SAP BW、SAP BW/4HANA 和 SAP S/4HANA 资源的以下选项卡：

- 沿袭和影响
- 关系
- 包含

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 扫描程序配置指南*》。

配置冲突解决方案以自动接受数据域

从版本 10.4.1.2 开始，您可以选择**匹配数据或列名称规则**选项来根据数据或列名规则自动接受数据域。如果选择**匹配数据或列名称规则**冲突解决方案选项，则当数据或列名规则结果匹配时，Enterprise Data Catalog 将自动接受数据域。

有关详细信息，请参阅《*Informatica 10.4.1 Catalog Administrator 指南*》中的“Enterprise Data Catalog 概念”一章。

第 3 章

通知、新增功能和更改 (10.4.1.1)

本章包括以下主题：

- [通知\(10.4.1.1\), 38](#)
- [新增功能 \(10.4.1.1\), 39](#)
- [更改 \(10.4.1.1\), 41](#)

通知(10.4.1.1)

支持更改

本节介绍版本 10.4.1.1 中的支持更改。

已提升技术预览

从版本 10.4.1.1 开始，提升了技术预览中的以下功能：

Cloudera Data Platform 集成

对于 Data Engineering Integration，您可以使用 Cloudera Data Platform (CDP) 作为计算群集，在 Hadoop 环境中执行数据工程作业。在 Spark 引擎上运行数据工程作业时，可以使用 Cloudera CDP。在 Blaze 引擎上不支持 Cloudera CDP。

Cloudera CDP 使用基本群集和工作负载群集来执行数据工程作业。此体系结构用于通过使用共享目录、统一安全、统一管理和数据生命周期管理来部署工作负载并在组件之间共享数据。

在具有以下连接的 Hadoop 环境中运行映射时，可以使用 Cloudera CDP：

- PowerExchange for Amazon Redshift
- PowerExchange for Amazon S3
- PowerExchange for HDFS
- PowerExchange for Microsoft Azure Blob Storage
- PowerExchange for Microsoft Azure CosmosDB SQL API
- PowerExchange for Microsoft Azure Data Lake Storage Gen1
- PowerExchange for Microsoft Azure Data Lake Storage Gen2
- PowerExchange for Microsoft Azure SQL Data Warehouse

- Sqoop

在本地环境运行 PowerExchange for HDFS 映射时，也可以使用 Cloudera CDP。

有关详细信息，请参阅《*Informatica® Data Engineering 10.4.1 集成指南*》。

发行版支持

在每个版本中，Informatica 都可以添加、延迟和删除对非本地发行版和各发行版版本的支持。Informatica 可能会在将来的版本中恢复对延期版本的支持。要查看支持的版本的最新列表，请参阅 Informatica 客户门户上的“产品可用性列表”，网址为：

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

新增功能 (10.4.1.1)

Data Engineering Integration

从版本 10.4.1.1 开始，可以访问 Amazon Glue 作为 Amazon EMR 5.29 群集的 Hive 元存储。

Amazon Glue 是 Amazon 的 ETL 服务。Amazon EMR 5.29 群集可以使用 Glue 作为 Hive 元存储。

要与具有 Glue 的 EMR 群集集成，请参阅《*Informatica 10.4.1.1 发行说明*》。

Data Privacy Management

本节介绍版本 10.4.1.1 中新增的 Data Privacy Management 功能。

增加了对 Apache Cassandra NoSQL 数据库的支持

从版本 10.4.1.1 开始，可以对 Apache Cassandra 数据库执行域发现。

您可以从类别类型中选择 NoSQL 类别，创建一个与 Apache Cassandra 数据库连接的数据存储。可以在扫描中使用此数据存储来执行域发现。

有关如何创建数据存储或运行扫描的信息，请参阅《*Informatica Data Privacy Management 10.4.1 用户指南*》。

在主题注册表搜索、索引和匹配配置中使用“完全匹配”

从版本 10.4.1.1 开始，可以将搜索、索引和匹配规则类型配置为使用“完全匹配”。当您在搜索或查找匹配过程中对索引使用“完全匹配”时，完成相关任务所需的时间更少。

在配置匹配时，可以对 *SearchType*、*IndexType* 和 *MatchType* 字段使用 Exact 值。

如果为索引配置“完全匹配”，则配置 IgnoreNullEmptyValues 属性，以确定在“链接主题”扫描期间是否忽略 null 或空值。

有关如何配置完全匹配的信息，请参阅《*Informatica Data Privacy Management 10.4.1 管理员指南*》。

增量主题注册表扫描

从版本 10.4.1.1 开始，对于重复的一次性扫描和反复执行的扫描，可以将“主题注册表”扫描配置为增量扫描。增量扫描考虑的是上次运行扫描后有更改的记录，并更新结果，而不是完全扫描所有记录。

有关如何对重复或反复出现的扫描将“主题注册表”扫描配置为增量扫描的信息，请参阅《*Informatica Data Privacy Management 10.4.1 用户指南*》和《*Informatica Data Privacy Management 10.4.1 管理员指南*》。

隐私仪表板数据更新的新作业类型

从版本 10.4.1.1 开始，“主题注册表”主扫描作业会在完成时触发一项新的“隐私仪表板”作业。

“隐私仪表板”作业将确认主题注册表 HBase 数据库中的更改，并在您刷新仪表板时更新您查看的隐私仪表板数据。

单独的作业确保了扫描任务和隐私仪表板上数据刷新的分离。

您不能手动触发、暂停或停止“隐私仪表板”作业。

有关隐私仪表板作业的信息，请参阅《*Informatica Data Privacy Management 10.4.1 用户指南*》。

主题注册表搜索和扫描字段

从版本 10.4.1.1 开始，可以配置在主题注册表扫描、搜索或两者中使用的字段。

您可能希望在扫描中加入一些字段，但不希望这些字段出现在主题注册表搜索和详细信息页面上。

您可以在实体的 MatchConfiguration 中加入 UsedIn 属性，以确定字段是否出现在主题注册表搜索、扫描或两者中。

有关如何配置该属性的信息，请参阅《*Informatica Data Privacy Management 10.4.1 管理员指南*》。

在 Windows 上支持 Informatica Discovery Agent

从版本 10.4.1.1 开始，可以使用 Informatica Discovery Agent 在 Windows 计算机上创建代理。

使用 Informatica Discovery Agent 对非结构化源和主题注册表扫描运行域发现。

有关如何在扫描中安装和使用代理的信息，请参阅《*Informatica Data Privacy Management 10.4.1 用户指南*》。

Enterprise Data Catalog

本节介绍版本 10.4.1.1 中新增的 Enterprise Data Catalog 功能。

选择因子来传播智能数据域

从版本 10.4.1.1 开始，可以选择一个或多个因子，将智能数据域传播到其他类似列。这些因子包括列名称、列数据模式、唯一值和值频率因子。

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 管理员指南*》中的“Enterprise Data Catalog 概念”一章。

收集数据集成日志进行数据剖析

从版本 10.4.1.1 开始，可以使用日志收集实用程序在“收集资源日志”页面中收集“数据集成服务”日志进行数据剖析。

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 管理员指南*》中的“目录备份和日志收集实用程序”一章。

SAP S/4HANA 资源

从版本 10.4.1.1 开始，SAP S/4HANA 资源从“附加视图”资产中提取元数据。您可以在 Catalog 中查看“附加视图”资产的沿袭和关系信息。有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 扫描程序配置指南*》。

独立扫描实用程序支持的资源类型

从版本 10.4.1.1 开始，在运行时或脱机时无法访问 SAP S/4HANA 数据源时，您可以从该数据源中提取元数据。

有关详细信息，请参阅《*Informatica 10.4.1 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。

更改 (10.4.1.1)

Data Privacy Management

本节介绍版本 10.4.1.1 中对 Data Privacy Management 进行的更改。

接口更改

从版本 10.4.1.1 开始，Data Privacy Management 界面中有以下更改：

“扫描创建”页面上的“主题扫描选项”字段

该选项可指定主题扫描选项。包括**完整主题扫描**选项，选择该选项可以对重复和反复执行的扫描配置增量扫描。

NoSQL 类别和 Apache Cassandra 数据存储类型

新建数据存储和**新建扫描**页面的“类别”列表包含一个新增的 **NoSQL** 选项。**数据存储类型**列表在 NoSQL 类别中加入了 Apache Cassandra。

有关数据存储和扫描创建选项的信息，请参阅《*Informatica Data Privacy Management 10.4.1 用户指南*》。

主题注册表实体文件更改

从版本 10.4.1.1 开始，主题注册表实体文件新增了一些属性来为实体配置选项。

主题注册表实体文件新增了以下属性：

UsedIn

指定在主题注册表扫描、搜索或两者同时执行时是否使用匹配配置。

以前，不能为搜索和扫描配置单独的字段。

IsIncrementalScan

添加此标志并将值设置为 true，可将重复和反复出现的主题注册表扫描配置为增量扫描。将该标志添加到一个包含时间值的字段。扫描操作使用该字段中的值来确定增量扫描的最后一次更新时间。适用于结构化源。

以前，您不能配置增量主题注册表扫描。

IgnoreNullOrEmptyValuesForScan

确定在遇到 null 或空值时，是忽略空值并继续，还是在使作业失败。适用于对匹配配置中包括 *完全匹配* 类型、索引类型或搜索类型的实体的主题注册表扫描。

以前，您不能配置完全匹配类型、搜索类型或索引类型。

有关实体文件的信息，请参阅《*Informatica Data Privacy Management 10.4.1 管理员指南*》。

Enterprise Data Catalog

本节介绍版本 10.4.1.1 中对 Enterprise Data Catalog 进行的更改。

清除缓存间隔

从版本 10.4.1.1 开始，**清除到期间隔** 标签重命名为 **清除缓存间隔**。

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 用户指南*》中的“配置分析应用程序”一章。

使用列相似性进行模式检测

从版本 10.4.1.1 开始，列相似性不会推断模式匹配的数值数据类型。也不会对所有数据类型推断 X、XX 和 XXX 模式。

以前，列相似性使用模式匹配和其他因素来推断数值数据，而且还对所有数据类型推断 X、XX 和 XXX 模式。

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 管理员指南*》中的“Enterprise Data Catalog 概念”一章。

资源扫描历史记录报告

从版本 10.4.1.1 开始，**扫描类型** 列显示 **手动**、**脱机** 和 **已计划** 作为资源扫描类型来指示资源扫描。

以前，**扫描类型** 列显示 **联机** 来指示所有扫描类型。**扫描类型** 列显示 **手动**，代表之前标记为 **联机** 的所有扫描类型。

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 用户指南*》中的“分析报告”一章。

第 4 章

通知 (10.4.1)

本章包括以下主题：

- [新产品\(10.4.1\), 43](#)
- [支持更改, 43](#)
- [安装程序更改, 46](#)
- [发布任务, 46](#)

新产品(10.4.1)

PowerExchange for MongoDB JDBC

从版本 10.4.1 开始，您可以使用 PowerCenter 客户端从 MongoDB JDBC 读取数据以及将数据写入到其中。您可以将 MongoDB 对象作为源和目标导入以创建映射，并运行会话以在 MongoDB 中读取或写入数据。在映射中使用 MongoDB 对象时，可以配置特定于 MongoDB 的属性。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for MongoDB JDBC 10.4.1 用户指南》。

支持更改

本节介绍版本 10.4.1 中的支持更改。

技术预览支持

已启动技术预览

从版本 10.4.1 开始，Informatica 包含以下技术预览功能：
Cloudera Data Platform 集成

对于 Data Engineering Integration，您可以使用 Cloudera Data Platform (CDP) 作为计算群集，在 Hadoop 环境中执行数据工程作业。在 Spark 引擎上运行数据工程作业时，可以使用 Cloudera CDP。在 Blaze 引擎上不支持 Cloudera CDP。

Cloudera CDP 使用基本群集和工作负载群集来执行数据工程作业。此体系结构用于通过使用共享目录、统一安全、统一管理和数据生命周期管理来部署工作负载并在组件之间共享数据。

在具有以下连接的 Hadoop 环境中运行映射时，可以使用 Cloudera CDP 群集：

- PowerExchange for Amazon Redshift
- PowerExchange for Amazon S3
- PowerExchange for Microsoft Azure Blob Storage
- PowerExchange for Microsoft Azure CosmosDB SQL API
- PowerExchange for Microsoft Azure Data Lake Storage Gen1
- PowerExchange for Microsoft Azure Data Lake Storage Gen2
- PowerExchange for Microsoft Azure SQL Data Warehouse
- PowerExchange for Google Cloud Storage
- PowerExchange for Google BigQuery
- PowerExchange for Google Analytics
- PowerExchange for Google Cloud Spanner
- Sqoop

当您运行使用 PowerExchange for HDFS 连接的映射时，还可以使用 Cloudera CDP 在本地环境或 Hadoop 环境中从源读取。

有关详细信息，请参阅《*Informatica® Data Engineering 10.4.1 集成指南*》。

Amazon S3 和 ADLS Gen2 目标的滚存参数

对于 Data Engineering Streaming，您可以为 Amazon S3 和 ADLS Gen2 目标使用不同的滚存参数来决定每个目标的滚存时间或大小。

SAP S/4HANA 资源

对于 Enterprise Data Catalog，可以从 SAP S/4HANA 数据源中提取元数据。有关详细信息，请参阅《*Informatica® 10.4.1 Enterprise Data Catalog 扫描程序配置指南*》。

服务和资源日志收集

从版本 10.4.1 开始，可以基于资源名称收集资源的日志。还可以收集服务（例如 HBase、Solr、Ingestion 和目录服务）的日志。可以使用 infacmd CLI、REST API 或 Catalog Administrator 收集日志。有关详细信息，请参阅《*Informatica® 10.4.1 Catalog Administrator 指南*》中的“Enterprise Data Catalog 实用程序”章节。

无中断目录备份

从版本 10.4.1 开始，可以在不禁用目录服务的情况下备份目录。可以使用 infacmd CLI、REST API 或 Catalog Administrator 备份目录。有关详细信息，请参阅《*Informatica® 10.4.1 Catalog Administrator 指南*》中的“Enterprise Data Catalog 实用程序”章节。

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。

已提升技术预览

从版本 10.4.1 开始，提升了技术预览中的以下功能：

Databricks 上的 Python 转换

对于 Data Engineering Integration，您可以在配置为在 Databricks Spark 引擎上运行的映射中包含 Python 转换。

SAP BW、SAP BW/4HANA 和 IBM InfoSphere DataStage 资源

对于 Enterprise Data Catalog，可以从 SAP BW、SAP BW/4HANA 和 IBM InfoSphere DataStage 源中提取元数据。有关详细信息，请参阅《*Informatica® 10.4.1 Enterprise Data Catalog 扫描程序配置指南*》。

延期

本节介绍版本 10.4.1 中的延期更改。

已提升延期

从版本 10.4.1 开始，以下功能不再延期：

- 流映射中的多个目标和 JMS 源。

IPv6 支持

从版本 10.4.1 开始，除了 IPv4 地址格式外，Informatica 还支持 Internet 协议版本 6 (IPv6) 地址格式。

确保域或服务的地址格式与相应数据库实例的地址格式相同。节点和数据库主机必须采用相同的地址格式。还可以使用 Internet 协议版本 6 (IPv6)，从在配置了 IPv6 的计算机上托管的 Informatica Developer 或 PowerCenter 连接到启用了 IPv6 的数据库。

下表列出了可以从 PowerCenter 客户端或 Developer tool 使用的受支持连接：

连接	工具
DB2	PowerCenter 客户端、Developer tool
JDBC	Developer tool
Microsoft SQL Server	PowerCenter 客户端、Developer tool
ODBC	PowerCenter 客户端、Developer tool
Oracle	PowerCenter 客户端、Developer tool
Sybase ASE	PowerCenter Client
Sybase IQ	PowerCenter Client

在 PowerCenter 客户端或 Developer tool 特定于数据库的连接属性中，指定主机名或 IP 地址可以连接到在 IPv6 网络上托管的数据库。

有关 IPv6 支持的详细信息，请参阅 <https://network.informatica.com/docs/DOC-16182> 和 <https://kb.informatica.com/proddocs/PAM%20and%20EOL/1/Informatica%20Support%20Statement%20for%20IPv6.pdf>。

发行版支持

在每个版本中，Informatica 都可以添加、延迟和删除对非本地发行版和各发行版版本的支持。Informatica 可能会在将来的版本中恢复对延期版本的支持。要查看支持的版本的最新列表，请参阅 Informatica 客户门户上的“产品可用性列表”，网址为：

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

Enterprise Data Catalog

本节介绍版本 10.4.1 中的 Enterprise Data Catalog 支持更改。

资源

本节介绍版本 10.4.1 中的资源支持更改。

资源的许可

从版本 10.4.1 开始，您需要拥有高级许可证才能使用以下资源：

- SQL Server Integration Service
- SAP BW
- SAP BW/4HANA
- IBM InfoSphere Datastage

请联系 Informatica 全球客户支持部门以获取高级许可证。

安装程序更改

Informatica 版本 10.4.1 是 HotFix 版本。您可以将 HotFix 应用到 10.4.0，也可以从受支持的版本升级。如果运行 Informatica 安装程序的计算机上不存在 10.4.1 产品安装二进制文件，则可以运行 Informatica 安装程序来安装这些二进制文件。

Informatica 版本 10.4.1 还支持将 Data Privacy Management 和 Test Data Management 产品纳入统一安装程序。除了应用 HotFix 外，Test Data Management 和 Data Privacy Management 还需要升级各项服务。HotFix 说明相同，但升级前后需具备的条件不相同。有关信息，请参阅各自的升级指南。

Data Privacy Management

安装 Data Privacy Management 时，可以安装 Data Engineering 产品和 Enterprise Data Catalog 并创建服务。如果已安装 Data Engineering 和 Enterprise Data Catalog，可以运行安装程序，在这些产品的基础上再安装 Data Privacy Management。

Test Data Management

创建域时，会安装 Test Data Management 二进制文件，并创建包含 TDM 和 TDW 文件夹的目录结构。要配置 Test Data Management，请参阅《*Test Data Management 配置指南*》。

发布任务

本节介绍版本 10.4.1 中的发布任务。发布任务是指升级到版本 10.4.1 后必须执行的任务。

Informatica PowerExchange 适配器

本节介绍版本 10.4.1 中 PowerExchange 适配器的发布任务。

Informatica PowerExchange for Microsoft Azure Data Lake Storage Gen1 的升级后更改

如果从早于 10.2.2 的版本升级到版本 10.4.1，并且已将现有映射配置为在目标中创建标题行，则在运行现有映射时升级后，不会在目标中创建标题。

要生成标题，可以重新导入目标对象，启用**导入第一行中的列名称**选项，然后再次运行映射。

Informatica PowerExchange for Snowflake 的升级后更改

应用 10.4.1 HotFix 或从早于 10.4.0 的版本升级到版本 10.4.1 之后，使用 AIX 计算机时必须将**其他 JDBC URL 参数**连接属性设置为 `jdbc_query_result_format=json`。

PowerCenter PowerExchange 适配器

本节介绍版本 10.4.1 中 PowerExchange 适配器的发布任务。

适用于 PowerCenter 的 PowerExchange for Snowflake 的升级后更改

应用 10.4.1 HotFix 或从早于 10.4.0 的版本升级到版本 10.4.1 之后，PowerExchange for Snowflake 包括以下更新：

- 使用 AIX 计算机时，必须将**其他 JDBC URL 参数**连接属性设置为 `jdbc_query_result_format=json`。
- 您必须在 PowerCenter 集成服务属性中指定 Java 堆空间内存，避免从 Snowflake 读取数据的映射和映射任务失败。因为 10.4.1 版本对 JDBC 驱动程序的升级导致映射失败。
将 **JVMOption1** 属性中的堆大小值配置为 256 MB 或更高。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica 10.4.1 PowerExchange for Snowflake 用户指南*》。

第 5 章

新增功能 (10.4.1)

本章包括以下主题：

- [连接\(10.4.1\), 48](#)
- [Data Engineering Integration, 48](#)
- [Data Engineering Streaming, 49](#)
- [Data Privacy Management, 51](#)
- [Enterprise Data Catalog, 52](#)
- [Informatica 转换, 56](#)
- [智能结构模型, 57](#)
- [Metadata Manager, 57](#)
- [PowerExchange 适配器, 57](#)

连接(10.4.1)

本节介绍版本 10.4.1 版本中新增的连接功能。

DataDirect ODBC 和 JDBC 驱动程序

从版本 10.4.1 开始，Informatica 随附的 DataDirect ODBC 和 JDBC 驱动程序为最新的可用版本。

Oracle Connection Manager

从版本 10.4.1 开始，您可以使用 Oracle Connection Manager 连接到 Oracle。在 PowerCenter 或 Developer tool 中指定 Oracle 连接的连接字符串，以连接到 Oracle 数据库。

有关详细信息，请参阅《*Informatica 10.4.1 管理员指南*》。

Data Engineering Integration

本节介绍版本 10.4.1 中新增的 Data Engineering Integration 功能。

Hive 源和目标上的层次结构数据

从版本 10.4.1 开始，您可以在 Spark 引擎上运行的映射中从包含层次结构数据的 Hive 表读取数据以及将数据写入到其中。

有关详细信息，请参阅《*Data Engineering Integration 10.4.1 用户指南*》。

Mass Ingestion 的 SAML 发送验证

从版本 10.4.1 开始，您可以配置 SAML 身份验证以登录到 Mass Ingestion 工具。

借助 SAML 身份验证，您可以使用第三方身份提供程序来管理用户凭据和授权。

有关详细信息，请参阅以下指南：

Data Engineering Integration 10.4.1 Mass Ingestion 指南

Informatica® 10.4.1 应用程序服务指南

Informatica® 10.4.1 安全指南

Data Engineering Streaming

本节介绍版本 10.4.1 中新增的 Data Engineering Streaming 功能。

ADLS Gen2 的 FileName 端口

从版本 10.4.1 开始，在您为 ADLS Gen2 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。您可以在 ADLS Gen2 目标中使用文件名端口从 PWX CDC 发布者引入 CDC 数据。

有关详细信息，请参阅《*Data Engineering Streaming 10.4.1 用户指南*》。

从多个 Kafka 主题引入 CDC 数据

从版本 10.4.1 开始，您可以从多个 Kafka 主题将来自 PWX CDC 发布者的 CDC 数据引入到 Data Engineering 系统中的一个或多个映射。

有关详细信息，请参阅《*Data Engineering Streaming 10.4.1 用户指南*》。

JDBC V2 查找转换

从版本 10.4.1 开始，您可以使用 JDBC 数据对象读取操作来查找 JDBC V2 表中的数据。

您可以在映射中添加 JDBC V2 数据对象读取操作以用作查找。然后将查找条件配置为从 JDBC V2 表查找数据。可以在 Databricks 引擎上运行该映射。

有关详细信息，请参阅《*Data Engineering Streaming 10.4.1 用户指南*》。

复杂目标的 Parquet 数据格式

从版本 10.4.1 开始，您可以对复杂目标使用 Parquet 数据格式。

您可以在流映射中对 HDFS、ADLS Gen2 和 Amazon S3 等复杂目标使用 Parquet 数据格式。

有关详细信息，请参阅《*Data Engineering Streaming 10.4.1 用户指南*》。

Amazon S3 目标和 ADLS Gen2 目标中的滚存参数

从版本 10.4.1 开始，您可以对 Amazon S3 目标和 ADLS Gen2 目标使用不同的滚存参数，以确定每个目标的滚存时间或大小。

有关详细信息，请参阅《*Data Engineering Streaming 10.4.1 用户指南*》。

源和目标源和目标

从版本 10.4.1 开始，在 Databricks 环境中，您可以在流映射中使用 Kafka 和 Confluent Kafka 作为源和目标。

在 AWS 云生态系统和 Microsoft Azure 云服务中，您都可以在 Databricks 环境中运行流映射。

有关详细信息，请参阅《*Data Engineering Streaming 10.4.1 用户指南*》。

AWS Databricks 中的流映射

从版本 10.4.1 开始，在 AWS 云生态系统中，您可以在 AWS Databricks 服务中运行流映射。

您可以使用 AWS Databricks 通过以下功能运行映射：

源和目标

可以在 Databricks 环境中针对以下源和目标运行流映射：

- Amazon S3
- Kinesis Streams
- Kinesis Firehose

转换

您可以将以下转换添加到 AWS 中的 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

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

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

有关详细信息，请参阅《*Data Engineering Streaming 10.4.1 用户指南*》。

Data Privacy Management

本节介绍版本 10.4.1 中新增的 Data Privacy Management 功能。

数据存储

本节介绍新的数据存储属性和源。

从版本 10.4.1 开始，您可以使用主题注册表远程代理扫描非结构化数据存储来发现敏感数据。可以将扫描配置为读取所有文件夹或读取列表中的特定文件夹，也可以配置为根据正则表达式读取文件夹。

版本 10.4.1 支持 Google 云端硬盘、SAP HANA 和 Snowflake 数据源。

有关详细信息，请参阅《《*Informatica Data Privacy Management 10.4.1 用户指南*》》中的“数据存储属性”一章和“扫描”一章。

作业

本节介绍新的作业类型。

版本 10.4.1 包括以下作业类型：

- “增量扫描”任务只会扫描数据存储中自上次运行扫描任务以来已更改的记录。
- “导入目录结果”任务将从 Enterprise Data Catalog 导入资源。
- “同步目录更新”任务会将 Enterprise Data Catalog 导入的信息与 Data Privacy Management 中的数据存储同步。

有关详细信息，请参阅《《*Informatica Data Privacy Management 10.4.1 用户指南*》》中的“作业”一章。

隐私仪表板

从版本 10.4.1 开始，Data Privacy Management 包含一个隐私仪表板，用于查看和监视主题注册表中的数据。

使用该仪表板上的摘要指示器和工作区指示器可以查看有关主题总计数、合法保留计数、主题类型和主题请求的信息。您可以通过详细信息页面（可在仪表板上打开）跟踪主题请求并对请求执行进一步操作。

还可以查看包含主题的数据存储的相关信息和位置。

有关隐私仪表板的详细信息，请参阅《《*Informatica Data Privacy Management 10.4.1 用户指南*》》。

主题注册表数据加密

从版本 10.4.1 开始，Data Privacy Management 将对主题扫描中确定的黄金记录数据进行加密，并将加密数据存储在 HBase 存储库中。

每次运行主题扫描时都会对数据进行加密，并且主题注册表存储库中的数据会更新。

结构化数据存储

本节介绍 Data Privacy Management 对结构化数据存储发现敏感数据和个人数据的方式所发生的更改。

版本 10.4.1 包含非结构化数据存储的以下新功能：

- 您可以为最大文件大小指定全局设置，以包含到非结构化数据存储扫描（使用远程代理扫描程序）中。
- 您可以使用 Data Privacy Management 安装中包含的、与预配置的分类策略匹配的数据域来扫描非结构化数据存储。对于安装中不包含的数据域，您可以在多路复用代理程序 (MPA) 中使用正则表达式来配置数据域。
- 您可以在分类策略中添加自定义匹配条件，以与 Google 云端硬盘和 Microsoft SharePoint 数据存储中的文件标记匹配。
- 您可以下载非结构化数据存储扫描的扫描作业报告。

有关详细信息，请参阅《《Informatica Data Privacy Management 10.4.1 用户指南》》。

Enterprise Data Catalog

本节介绍版本 10.4.1 中新增的 Enterprise Data Catalog 功能。

业务术语关联

从版本 10.4.1 开始，您可以在目录中将业务术语与来自 Azure Data Lake Store Gen2 资源的资产相关联。

有关详细信息，请参阅《Informatica 10.4.1 Catalog Administrator 指南》中的“Enterprise Data Catalog 概念”一章。

为数据规则和列名称规则配置冲突解决方案

从版本 10.4.1 开始，您可以选择接受数据规则结果或列名称规则结果，也可以同时接受数据规则结果和列名称规则结果（如果自动接受数据域期间规则相冲突）。

有关详细信息，请参阅《Informatica 10.4.1 Catalog Administrator 指南》中的“Enterprise Data Catalog 概念”一章。

上下文沿袭信息

从版本 10.4.1 开始，您可以使用沿袭和影响选项卡查看资产的上下文沿袭信息。在 Catalog Administrator 中创建自定义资源，以提取包含进程定义或映射执行的资产的上下文信息。

有关详细信息，请参阅《Informatica 10.4.1 Catalog Administrator 指南》中的“将自定义元数据引入到目录中”一章和《Informatica 10.4.1 Enterprise Data Catalog 用户指南》中的“查看沿袭和影响”一章。

Data Asset Analytics

从版本 10.4.1 开始，您可以使用 Data Asset Analytics，以报告形式查看有关目录的分析见解。分析见解包括有关配置用于访问目录的用户、资产和资源、资产使用情况的信息，以及与资产关联的扩充和协作详细信息。

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 用户指南*》。

通过导入和导入进行数据管理

从版本 10.4.1 开始，您可以接受或拒绝数据域的多个推理的资产。您可以将资产导出到逗号分隔 (CSV) 文件，然后将更新的 CSV 文件导入到 Enterprise Data Catalog。

通过从导出的 CSV 文件中的接受或推理的数据域列中删除数据域，可以从资产中删除数据域。

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 用户指南*》中的“查看资产”一章。

数据发现

从版本 10.4.1 开始，您可以对以下资源和文件类型运行配置文件以发现数据域：

- Azure Data Lake Store Gen2 资源。
可以对结构化、非结构化和扩展非结构化文件类型运行配置文件。
- Cassandra 资源。
您可以在本地引擎上执行数据发现。
- 适用于 Amazon S3 和 Azure Data Lake Store 资源的 Parquet 文件。

有关详细信息，请参阅《*Informatica 10.4.1 Catalog Administrator 指南*》中的“Enterprise Data Catalog 概念”一章。

Enterprise Data Catalog 演练

从版本 10.4.1 开始，您可以访问演练，这些演练可帮助您了解和快速开始使用 Enterprise Data Catalog。演练是 Enterprise Data Catalog 中的指导式教程，介绍如何使用特定功能。

Enterprise Data Catalog 中提供以下演练：

- 入门指南和目录主页简介
- 搜索结果简介
- 表概览简介
- 沿袭简介
- 列概览简介
- 提高资产可靠性

要启动演练，请单击工具栏中的 ? 帮助图标，然后选择并单击演练链接。

提取转换端点

从版本 10.4.1 开始，您可以提取具有 Web 服务转换映射的 PowerCenter 资源的转换端点。在 Enterprise Data Catalog 中，转换端点显示为连接，这些连接链接了数据源或目标。

平面文件的字段级别沿袭

从版本 10.4.1 开始，您可以在 Catalog 中查看源或目标平面文件的字段级别沿袭。

下表包含了资源支持的文件系统，您可以配置这些资源以查看字段级别沿袭：

资源	支持的文件系统
服务	Linux Windows
SQL Server Integration Service	Windows Azure Data Lake Store Gen2
Informatica Cloud Service	Azure Data Lake Store Gen2 Azure Data Lake Store Gen1 Amazon S3 Azure Blob 存储 Google Cloud Storage Linux Windows
IBM InfoSphere DataStage	Linux Windows
Informatica Platform	Linux Windows

要查看不带标题的平面文件的字段级别沿袭，您必须在配置资源期间选择**启用引用资源属性**和**保留未解析的引用资产属性**。对于带标题的平面文件，无需选择这些属性便能查看字段级别沿袭。

云存储的文件沿袭

从版本 10.4.1 开始，您可以使用 Informatica Cloud Service 资源查看云存储的文件沿袭。

您可以查看使用 Amazon S3、Microsoft Azure Blob 存储和 Azure Data Lake Store Gen2 等数据源的 Informatica Cloud Service 映射的文件沿袭。

Hive 资源

从版本 10.4.1 开始，Hive 资源包括以下增强功能：

提取连接元数据

您可以使用 Hive 资源来提取不同架构的视图的连接详细信息。

自动分配连接

创建资源时，您可以选择自动为 Hive 资源分配数据库架构。您可以查看自动为资源分配的架构及其连接的列表。可以在自动分配的连接中分配或取消分配架构。

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 扫描程序配置指南*》。

Informatica MDM 资源

从版本 10.4.1 开始，Informatica MDM 资源将通过 API 从 MDM 数据源提取元数据。

Informatica MDM 资源包括以下增强功能：

元数据提取

您可以配置 Informatica MDM 资源，以提取基础对象、登录表和暂存表等元数据以及字段和属性信息。

沿袭增强功能

您可以查看登录表和业务实体之间的沿袭。还可以查看从 Informatica Platform 应用程序的详细沿袭。

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 扫描程序指南*》。

Microsoft SQL Server 资源

从版本 10.4.1 开始，在创建 Microsoft SQL Server 资源时，您可以选择自动为资源分配数据库架构。您可以查看自动为资源分配的架构及其连接的列表。可以在自动分配的连接中分配或取消分配架构。

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 扫描程序配置指南*》。

MicroStrategy 资源

从版本 10.4.1 开始，MicroStrategy 资源包括以下增强功能：

查看报告级别沿袭

您可以查看 MicroStrategy 资源的报告级别沿袭。报告级别沿袭不包含页眉、页脚和页体等容器，以及列、行和度量等分组元素。

支持增量加载

您可以为 MicroStrategy 资源启用增量加载。增量加载使数据源能够加载最近对元数据进行的更改，而不是加载完整的元数据。增量加载可以缩短加载资源所需的时长。

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 扫描程序配置指南*》。

分区文件发现

从版本 10.4.1 开始，可以使用 Amazon S3 和 Azure Data Lake Store Gen2 资源来确定和发布在同一个目录下水平分区的文件，以及在 Hive 样式层次结构目录结构中组织成单个分区文件的多个文件。

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 扫描程序配置指南*》。

引用资源和引用资产

从版本 10.4.1 开始，可以配置 Hive 和 Microsoft SQL Server 资源，以从数据源或该资源引用的目录中的其他资源提取元数据。

有关详细信息，请参阅《*Informatica 10.4.1 Catalog Administrator 指南*》和《*Informatica 10.4.1 Enterprise Data Catalog 用户指南*》。

资源级别权限

从版本 10.4.1 开始，您可以为用户或用户组分配配置权限，以及对资源应用限制。

注意：也可以在创建资源期间对资源应用限制。

有关详细信息，请参阅《*Informatica 10.4.1 Catalog Administrator 指南*》。

提取数据库名称

从版本 10.4.1 开始，可以使用 Oracle 资源通过**导入数据库名称**选项来提取配置为元数据源的 Oracle 数据库的名称。默认情况下，此选项处于启用状态。

不能使用 `-DextractDatabaseName=true` JVM 选项来提取 Oracle 元数据源的数据库名称。

注意: 如果从较早的版本升级，在 Oracle 资源配置中，**导入数据库名称**选项会替换 `-DextractDatabaseName=true` JVM 选项。

REST API

从版本 10.4.1 开始，可以使用以下 Informatica Enterprise Data Catalog REST API：

- 目录事件 REST API。除了现有 REST API 之外，您还可以列出用户订阅的对象的事件类类型。
- 沿袭筛选器 REST API。除了现有 REST API 之外，您还可以创建默认沿袭筛选器。
- 监视信息 REST API。除了现有 REST API 之外，您还可以执行以下任务：
 - 列出作业的日志。
 - 提交和列出备份作业、资源扫描日志作业和服务日志作业。
 - 下载作业的特定日志，以及备份作业、资源扫描日志作业和服务日志作业的 ZIP 文件。

有关 REST API 的详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog REST API 参考*》。

SAP BW 和 SAP BW/4HANA 资源

从版本 10.4.1 开始，SAP BW 和 SAP BW/4HANA 资源从 InfoObject as InfoProvider 资产提取元数据。您可以在 Catalog 中查看 InfoObject as InfoProvider 资产的沿袭和关系信息。

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 扫描程序配置指南*》。

SSIS 资源

从版本 10.4.1 开始，SSIS 资源包括以下增强功能：

查看转换的详细沿袭

可以在 Catalog 中查看汇总、审计和字符映射等转换的详细沿袭。

从 SSIS 数据库提取元数据

您可以配置 SSIS 资源，以从 SSIS 数据库提取元数据。

字段级别和控制沿袭支持

可以在 Catalog 中查看平面文件的字段级别沿袭，以及 SSIS 资产的控制摘要。

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 扫描程序指南*》。

Informatica 转换

本节介绍版本 10.4.1 中新增的 Informatica 转换功能。

宏转换

从版本 10.4.1 开始，您可以在 Informatica 映射中使用宏转换。宏转换可为 Maplet 中包含的转换逻辑添加动态功能。

使用宏转换之前，请创建包含您希望接受动态输入的转换逻辑的 Maplet。然后将宏转换添加到动态映射中，并使用您创建的 Maplet 作为宏指令。

例如，您要从动态源读取数据，执行数据加密，然后将加密数据写入到目标。首先使用数据屏蔽转换创建一个用于执行加密的 Maplet。然后创建一个用于从数据源读取数据以及将数据写入到其中的 Maplet。将宏转换添加到动态映射中，并使用数据屏蔽 Maplet 作为宏指令。

有关详细信息，请参阅《*Informatica 10.4.1 Developer 转换指南*》中的“宏转换”一章。

智能结构模型

本节介绍版本 10.4.1 中新增的智能结构模型功能。

更改数组结构中的根节点

从版本 10.4.1 开始，可以通过选择结构中成为根节点的元素来确定 Intelligent Structure Discovery 平展结构的位置。例如，如果中游映射中的智能结构模型包含复杂的层次结构输入，则选择顶级节点作为根节点来避免数据丢失。

有关详细信息，请参阅《*Data Engineering Integration 10.4.1 用户指南*》。

Metadata Manager

本节介绍版本 10.4.1 中新增的 Metadata Manager 功能。

为业务对象资源配置 Java 位置和堆大小

从版本 10.4.1 开始，您可以在“其他”属性中为业务对象资源配置 Java 位置和堆大小。可以根据 Java SE Runtime Environment 配置这些选项。

有关详细信息，请参阅《*Informatica 10.4.1 Metadata Manager 管理员指南*》中的“业务情报资源”一章。

PowerExchange 适配器

本节介绍版本 10.4.1 中新增的 PowerExchange 适配器功能。

Informatica PowerExchange 适配器

本节介绍版本 10.4.1 中新增的 PowerExchange 适配器功能。

PowerExchange for Amazon Redshift

从版本 10.4.1 开始，PowerExchange for Amazon Redshift 包含以下功能：

- 您指定用于创建暂存文件的 Amazon S3 存储桶可以位于与 Amazon Redshift 群集不同的区域。
- 您可以使用外部帐户的 KMS 客户管理的密钥来创建加密资源，方法是在连接属性中提供外部帐户的 Amazon 资源名称 (ARN)。

有关详细信息，请参阅《*Informatica 10.4.1 PowerExchange for Amazon Redshift 用户指南*》。

PowerExchange for Amazon S3

从版本 10.4.1 开始，PowerExchange for Amazon S3 包含以下功能：

- 您可以使用 .manifest 文件从 Amazon S3 读取多个平面格式类型的文件，并将数据写入到目标。
- 您可以在本地环境和 Spark 引擎中使用 AWS ap-east-1（香港）区域。
- 您可以使用外部帐户的 KMS 客户管理的密钥来创建加密资源，方法是在连接属性中提供外部帐户的 Amazon 资源名称 (ARN)。
- 您可以在本地模式模式下使用 FileName 端口运行映射，以读取 Amazon S3 二进制文件。
- 您可以在本地模式模式下从某个目录读取 Amazon S3 二进制文件。

有关详细信息，请参阅《*Informatica 10.4.1 PowerExchange for Amazon S3 用户指南*》。

PowerExchange for Google BigQuery

从版本 10.4.1 开始，PowerExchange for Google BigQuery 包含以下功能：

- 可以针对 Google BigQuery 表配置缓存查找操作。
- 您可以从位于特定区域的 Google BigQuery 数据集中的表读取数据，以及将数据写入到其中。

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

PowerExchange for Hive

从版本 10.4.1 开始，您可以在 Spark 引擎上运行的映射中使用复杂数据类型从 Hive 表读取层次结构数据以及将层次结构数据写入到其中。

从包含层次结构数据的 Hive 表读取数据以及将数据写入到其中时，您可以对映射执行数据预览和架构同步。

只能在 HDP 3.1 发行版中使用包含层次结构数据的 Hive 表。

有关详细信息，请参阅《*Informatica 10.4.1 PowerExchange for Hive 用户指南*》。

PowerExchange for Microsoft Azure Data Lake Storage Gen2

从版本 10.4.1 开始，PowerExchange for Microsoft Azure Data Lake Storage Gen2 包含以下功能：

- 以下运行时引擎可以读取和写入 ORC 文件：
 - 数据集成服务
 - Spark
 - Databricks Spark
- 您可以在本地环境中运行映射以读取和写入 JSON 文件。
- 在本地环境和 Spark 引擎中，您可以在映射中配置 Azure Government 端点。
- 您可以为数据集成服务配置经过身份验证的代理服务器设置，以连接到 Microsoft Azure Data Lake Storage Gen2。

有关详细信息，请参阅《*Informatica 10.4.1 PowerExchange for Microsoft Azure Data Lake Storage Gen2 用户指南*》。

PowerExchange for Salesforce

从版本 10.4.1 开始，PowerExchange for Salesforce 包含以下功能：

- 您可以使用版本 48.0 的 Salesforce API 创建 Salesforce 连接以及访问 Salesforce 对象。
- 您可以对动态映射进行配置，以根据您定义的参数和规则将运行时对源、目标和转换逻辑的频繁更改包括在内。

有关详细信息，请参阅《*Informatica 10.4.1 PowerExchange for Salesforce 用户指南*》。

PowerExchange for SAP NetWeaver

从版本 10.4.1 开始，在 SAP 表读取器 中使用 HTTP 流数据传输模式时，您可以使用 MB 为单位指定传输数据包大小以从 SAP 表读取数据。

升级后，SAP 表读取器中使用 HTTP 流数据传输模式的现有映射可能会体现出性能改进。要进一步调谐性能，请根据要求指定合适的传输数据包大小。

有关详细信息，请参阅《*Informatica 10.4.1 PowerExchange for SAP NetWeaver 用户指南*》。

PowerExchange for Snowflake

从版本 10.4.1 开始，您可以使用外部表和具体化视图作为 Snowflake 源并执行所有读取操作。

有关详细信息，请参阅《*Informatica 10.4.1 PowerExchange for Snowflake 用户指南*》。

PowerCenter PowerExchange 适配器

本节介绍版本 10.4.1 中新增加的 PowerExchange 适配器功能。

PowerExchange for Db2 Warehouse

从版本 10.4.1 开始，PowerExchange for Db2 Warehouse 包含以下功能：

- 您可以在 Db2 Warehouse 中读取和写入 VarcharForBitData 和 CharForBitData 数据类型的数据。
- 您可以在 Windows 和 AIX 平台上使用 PowerExchange for Db2 Warehouse。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica PowerExchange for Db2 Warehouse 10.4.1 用户指南*》。

PowerExchange for Greenplum

从版本 10.4.1 开始，您可以在 Windows 上使用 PowerExchange for Greenplum。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica 10.4.1 PowerExchange for Greenplum 用户指南*》。

PowerExchange for Google BigQuery

从版本 10.4.1 开始，PowerExchange for Google BigQuery 包含以下功能：

- 您可以创建管道查找转换，用于对 Google BigQuery 表执行查找。

- 您可以创建会话，以从更改数据捕获 (CDC) 源读取实时数据或已更改的数据，并将数据加载到 Google BigQuery 中。
- 您可以从位于特定区域的 Google BigQuery 数据集中的表读取数据，以及将数据写入到其中。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica 10.4.1 PowerExchange for Google BigQuery 用户指南》。

PowerExchange for Google Cloud Storage

从版本 10.4.1 开始，在导入 Google Cloud Storage 目标定义时，您可以删除 Google Cloud Storage 平面文件中的标题行。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica 10.4.1 PowerExchange for Google Cloud Storage 用户指南》。

PowerExchange for Kafka

从版本 10.4.1 开始，PowerExchange for Kafka 包含以下功能：

- 将数据写入到 Kafka 目标时，您可以在实时会话中配置基于文件的恢复。
- 可以配置其他配置属性来通过 SSL 连接到 Kafka 代理。
- 在“其他安全属性”字段中指定配置属性时，指定的值会屏蔽。
- 您可以为 Kafka 代理配置 PLAIN 安全。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Kafka 10.4.1 用户指南》。

PowerExchange for Microsoft Azure SQL Data Warehouse V3

从版本 10.4.1 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《适用于 PowerCenter 的 Informatica 10.4.1 PowerExchange for Microsoft Azure SQL Data Warehouse V3 用户指南》。

PowerExchange for Salesforce

从版本 10.4.1 开始，可以使用版本 48.0 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 10.4.1 PowerExchange for Salesforce 用户指南》。

PowerExchange for SAP NetWeaver

从版本 10.4.1 开始，当您在 PowerExchange for SAP Dynamic ABAP Table Extractor 中使用 HTTP 流数据传输模式时，您可以使用 MB 为单位指定传输数据包大小，以优化从 SAP 表读取数据的性能。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica 10.4.1 PowerExchange for SAP NetWeaver 用户指南》。

PowerExchange for Snowflake

从版本 10.4.1 开始，PowerExchange for Snowflake 包含以下功能：

- 您可以使用外部表和具体化视图作为 Snowflake 源并执行所有读取操作。

- 您可以将 PowerCenter 集成服务生成的默认更新查询替换为自己指定的更新查询。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica 10.4.1 PowerExchange for Snowflake 用户指南》。

第 6 章

更改 (10.4.1)

本章包括以下主题：

- [命令行程序, 62](#)
- [Data Engineering Integration, 63](#)
- [Data Privacy Management, 63](#)
- [文档, 64](#)
- [域安全性, 65](#)
- [Enterprise Data Catalog, 65](#)
- [Informatica Developer , 66](#)
- [Metadata Manager, 66](#)
- [Informatica PowerExchange 适配器, 67](#)
- [Test Data Management, 67](#)

命令行程序

本节介绍版本 10.4.1 中对命令进行的更改。

infacmd ldm 命令

下表介绍了已更改的 infacmd ldm 命令：

命令	更改说明
BackupContents	添加了 -mode 用于备份内容。可以使用脱机、hdfs 或联机模式进行备份。
CreateService	添加了以下选项： <ul style="list-style-type: none">- -ed。启用 Data Asset Analytics。- -ddt。选择 Data Asset Analytics 的数据库。- -ddu。访问 Data Asset Analytics 的数据库。- -ddp。用于访问 Data Asset Analytics 的密码。- -ddl。提供数据库连接字符串。- -dds。提供数据库架构名称。- -dsjdbcp。提供用于保护 JDBC 连接参数的值。

有关详细信息，请参阅《*Informatica 10.4.1 命令参考*》。

Data Engineering Integration

本节介绍版本 10.4.1 中对 Data Engineering Integration 进行的更改。

Intelligent Structure Discovery

从版本 10.4.1 开始，当 Intelligent Structure Discovery 推理 Avro、Parquet 和 ORC 文件的数据类型时，它会使用与本地连接器相同的数据类型，从而在各个平台之间生成统一的输出。

在以前的版本中，Intelligent Structure Discovery 数据类型推理与本地连接器的推理不保持一致。

有关详细信息，请参阅《*Data Engineering Integration 10.4.1 用户指南*》中的“智能结构模型”。

Data Privacy Management

本节介绍版本 10.4.1 中对 Data Privacy Management 进行的更改。

Data Privacy Management 安装程序

从版本 10.4.1 开始，Data Privacy Management 安装程序包含在 Informatica 服务安装程序中。您可以从可用的产品安装选项中为 Data Privacy Management 选择所需的安装选项。

在以前的版本中，Data Privacy Management 安装程序是单独的安装程序。您可以从可用的 Data Privacy Management 安装选项中选择所需的安装选项。

现在，远程代理安装程序文件在安装后位于以下位置：<INFA_HOME>/SecureAtSourceService/InformaticaDiscoveryAgent/InformaticaDiscoveryAgent.zip

您可以使用远程代理对非结构化源执行主题注册表扫描和域发现扫描。

在以前的版本中，远程代理安装程序文件在安装后位于以下位置：<INFA_HOME>/SecureAtSourceService/SubjectRegistryAgent/SubjectRegistryAgent.zip

您只能使用代理对非结构化源执行主题注册表扫描。

有关如何使用安装程序和安装远程代理的信息，请参阅《《*Informatica Data Privacy Management 10.4.1 安装和配置指南*》》

概览工作区

本介绍**概览**工作区名称的更改。

从版本 10.4.1 开始，包含敏感数据、用户和用户行为、仪表板分析、敏感数据位置和扩散以及保护度量等指示器的**概览**工作区视图名为**安全仪表板**。

首次登录到 Data Privacy Management 时，**概览**工作区将显示**隐私仪表板**。要查看**安全仪表板**，请在**设置**工作区中更改默认设置。

在以前的版本中，**概览**工作区只会显示现在出现在**安全仪表板**中的信息。

有关详细信息，请参阅《《*Informatica Data Privacy Management 10.4.1 用户指南*》》中的“安全仪表板”一章。

扫描作业报告

本节介绍您在**作业详细信息**页面中下载的数据存储扫描报告发生的更改。

从版本 10.4.1 开始，您可以下载“加载目录”作业步骤的报告。剖析作业步骤包含一个 .zip 文件，其中包含两个 .csv 文件报告。

在以前的版本中，“加载目录”作业步骤不包含报告，而剖析作业步骤包含单个 .csv 文件报告。

有关详细信息，请参阅《《Informatica Data Privacy Management 10.4.1 用户指南》》中的“作业”一章。

敏感文件页面视图

本节介绍**敏感文件**页面中的新视图和列。

版本 10.4.1 包含非结构化数据存储的以下视图：

平面视图

平面视图列出了包含敏感数据域的非结构化数据存储中的文件和文件夹。文件和文件夹按字母顺序显示。

平面视图包含新的“置信度”列。该列显示与使用远程代理进行扫描的非结构化数据存储中的文件匹配的分类策略的置信水平。

树视图

树视图列出了包含扫描所确定的非结构化文件和文件夹的顶级目录。您可以向下钻取到文件夹和文件。要导航到目录树中的上一个文件夹，请单击显示在列表上方的文件路径中的文件夹名称。

在以前的版本中，“敏感字段”视图和“导入的字段”视图显示所有数据存储类型的敏感数据。在版本 10.4.1 中，只有结构化数据存储的**敏感字段**页面包含这些视图。

有关详细信息，请参阅《《Informatica Data Privacy Management 10.4.1 用户指南》》中的“安全仪表板”一章。

文档

本节介绍版本 10.4.1 中 Informatica 文档中的指南更改。

Data Privacy Management

从版本 10.4.1 开始，Data Privacy Management 发行指南和发行说明与 Informatica 发行指南和发行说明合并到一起。

您可以在 Informatica 发行指南中找到有关新增的和已更改的 Data Privacy Management 功能的信息。有关 Data Privacy Management 中修复的问题和已知限制的信息可在 Informatica 发行说明中找到。

在以前的版本中，Data Privacy Management 具有单独的发行指南和发行说明。

Test Data Management

从版本 10.4.1 开始，Test Data Management 发行指南和发行说明与 Informatica 发行指南和发行说明合并到一起。

您可以在 Informatica 发行指南中找到有关新增的和已更改的 Test Data Management 功能的信息。有关 Test Data Management 中修复的问题和已知限制的信息可在 Informatica 发行说明中找到。

在以前的版本中，Test Data Management 具有单独的发行指南和发行说明。

域安全性

本节介绍版本 10.4.1 中对 Informatica 域安全进行的更改。

启用对启用 Kerberos 的域的访问

从版本 10.4.1 开始，当您 Informatica 域启用 Kerberos 时，您必须将 DNS 条目添加到 krb5.conf 文件中。dns_lookup_kdc 和 dns_lookup_realm 的条目使 Kerberos 能够定位 KDC 和 Kerberos 域。

在以前的版本中，不需要这些条目。

有关详细信息，请参阅《Informatica 安全指南》中的“配置 Kerberos 配置文件”。

Enterprise Data Catalog

本节介绍版本 10.4.1 中对 Enterprise Data Catalog 进行的更改。

更改摘要页面

从版本 10.4.1 开始，您可以对更改摘要通知列表执行导出、排序和筛选。可以根据资产类型和扫描时间筛选列表。在以前的版本中，您可以查看更改摘要通知列表并按资源扫描时间筛选列表。

有关详细信息，请参阅《Informatica 10.4.1 Enterprise Data Catalog 用户指南》中的“资产方面的用户协作”一章。

将资产数据导出到 Tableau 数据提取文件

从版本 10.4.1 开始，此功能包括以下更改：

- 如果您拥有资源的**所有权限**或**元数据**和**数据读取**权限，则可以使用**导出**选项将资产数据导出到 Tableau 数据提取文件。在以前的版本中，用户不需要拥有任何权限和特权便可使用**导出**选项将资产数据导出到 Tableau 数据提取文件。
- 您可以在 Informatica Administrator 中为目录服务配置 LdmCustomOptions.enableTDEExport 自定义属性并将值设置为 false，以在目录中对所有用户（无论为用户分配了什么权限）隐藏**导出**选项。在以前的版本中，**导出**选项在目录中对所有用户可见。

有关详细信息，请参阅《Informatica 10.4.1 Enterprise Data Catalog 用户指南》中的“执行资产任务”一章。

资产的导出路径

从版本 10.4.1 开始，从 Business Glossary、Axon Glossary、Data Domain 和 Domain User 资源导出资产时，您可以查看资产的路径。在以前的版本中，导出的 csv 文件不包含这些资源的资产路径。

有关详细信息，请参阅《Informatica 10.4.1 Enterprise Data Catalog 用户指南》中的“执行资产任务”一章。

显示活动业务术语

从版本 10.4.1 开始，您可以在**添加业务标题**对话框中选择在词汇表列表中显示活动业务术语。在以前的版本中，词汇表列表会同时显示非活动业务术语和活动业务术语。

有关详细信息，请参阅《Informatica 10.4.1 Enterprise Data Catalog 用户指南》中的“执行资产任务”一章。

对列和字段资产进行排序

从版本 10.4.1 开始，您可以在**列**选项卡和**字段**选项卡中按源位置对列和分隔字段资产进行排序。在以前的版本中，无论列和分隔字段资产在数据源中的位置如何，它们都按升序或降序显示。

有关详细信息，请参阅《*Informatica 10.4.1 Enterprise Data Catalog 用户指南*》中的“查看资产”一章。

Catalog Administrator 中的开始选项卡名称

从版本 10.4.1 开始，Catalog Administrator 中的**开始**选项卡重命名为**概览**选项卡。

有关详细信息，请参阅《*Informatica 10.4.1 Catalog Administrator 指南*》。

Informatica Developer

本节介绍版本 10.4.1 中对 Informatica Developer 进行的更改。

关系对象搜索

从版本 10.4.1 开始，Developer tool 中的关系搜索已增强。

创建关系数据对象时，您可以搜索连接中的可用架构。Developer tool 会显示与搜索条件匹配的可用架构数量和表数量。可用资源的列表按资源类型组织。没有成员的类型将不会出现在列表中。例如，如果资源不包含任何视图，则“视图”节点将不会出现在选择列表中。

有关详细信息，请参阅《*Developer tool 指南*》中有关导入关系数据对象的一节。

查看执行计划

从版本 10.4.1 开始，您可以从映射中查看在非本地环境中运行的映射的执行计划。

在以前的版本中，您可以从“数据查看器”视图中查看执行计划。

有关详细信息，请参阅《*Data Engineering Integration 10.4.1 用户指南*》中的“映射执行计划”。

Metadata Manager

本节介绍版本 10.4.1 中对 Metadata Manager 进行的更改。

以 HTML 5 格式查看数据沿袭

从版本 10.4.1 开始，您可以采用 HTML 5 格式查看数据沿袭图表。

在以前的版本中，您可以在 Adobe Flash Player 中查看数据沿袭图表。

有关详细信息，请参阅《*Informatica 10.4.1 Metadata Manager 用户指南*》中的“使用数据沿袭”一章。

Informatica PowerExchange 适配器

本节介绍版本 10.4.1 中对 PowerExchange 适配器进行的更改。

PowerExchange for Amazon S3

从版本 10.4.1 开始，当您在运行时刷新动态映射中的源架构或目标架构时，数据集成服务将使用您为平面文件的分隔符、文本限定符和转义符指定的值。

在以前的版本中，数据集成服务不会使用您指定的值，而会使用默认值，从而导致目标中出现意外结果。

PowerExchange for Microsoft Azure Blob 存储

从版本 10.4.1 开始，当您在运行时刷新动态映射中的源架构或目标架构时，数据集成服务将使用您为平面文件的分隔符、文本限定符和转义符指定的值。

在以前的版本中，数据集成服务不会使用您指定的值，而会使用默认值，从而导致目标中出现意外结果。

PowerExchange for Microsoft Azure Data Lake Storage Gen1 and Gen2

从版本 10.4.1 开始，当您在运行时刷新动态映射中的源架构或目标架构时，数据集成服务将使用您为平面文件的分隔符、文本限定符和转义符指定的值。

在以前的版本中，数据集成服务不会使用您指定的值，而会使用默认值，从而导致目标中出现意外结果。

Test Data Management

本节介绍版本 10.4.1 中对 Test Data Management 进行的更改。

Test Data Management 安装程序

从版本 10.4.1 开始，Test Data Management 安装程序包含在 Informatica 服务安装程序中。安装传统产品和 Data Engineering 产品时均会安装 Test Data Management 二进制文件。

要安装 Test Data Management，请从可用的产品安装选项中选择安装 Informatica 服务。

在以前的版本中，Test Data Management 安装程序是单独的安装程序。您可以从可用的 Test Data Management 安装选项中选择所需的安装选项。

有关如何安装 Test Data Management 的信息，请参阅《*PowerCenter 和 Data Quality 的 Informatica 安装*》和《*Data Engineering 的 Informatica 安装*》。

有关如何配置 Test Data Management 的信息，请参阅《*Test Data Management 配置指南*》。

第 II 部分： 版本 10.4.0

本部分包含以下章节：

- [新增功能 \(10.4.0.2\), 69](#)
- [通知、新增功能和更改 \(10.4.0.1\), 70](#)
- [通知 \(10.4.0\), 73](#)
- [新产品\(10.4.0\), 78](#)
- [新增功能 \(10.4.0\), 80](#)
- [更改 \(10.4.0\), 104](#)

第 7 章

新增功能 (10.4.0.2)

本章包括以下主题：

- [更新群集发行版本, 69](#)

更新群集发行版本

从版本 10.4.0.2 开始，您可以使用 `infacmd dis updateClusterDistributionVersion` 命令为工作流中的“创建群集”任务更新发行版本。

应用 Service Pack 10.4.0.2 时，“创建群集”任务的群集版本属性不会更新到受支持的最新版本。使用该命令可以为应用程序中的所有工作流更新群集发行版本。

有关详细信息，请参阅知识库文章

[“HOW TO: Update cluster distribution version after you apply service pack 10.4.0.2”](#)。

第 8 章

通知、新增功能和更改 (10.4.0.1)

本章包括以下主题：

- [通知 \(10.4.0.1\), 70](#)
- [新增功能 \(10.4.0.1\), 70](#)
- [更改 \(10.4.0.1\), 71](#)

通知 (10.4.0.1)

已提升技术预览

从版本 10.4.0.1 开始，提升了技术预览中的以下功能：

将 Databricks 增量表作为流映射目标

对于 Data Engineering Streaming，您可以使用 Databricks 增量表作为流映射的目标，以引入流数据。

动态流映射

您可以配置动态流映射，以根据您在 Confluent 架构注册表中定义的参数和规则更改运行时 Kafka 源和目标。

Hive 仓库连接器和 Hive LLAP

对于 Data Engineering 产品，可以对 Azure HDInsight 4.x 和 Hortonworks HDP 3.x 群集使用 Hive 仓库连接器和 Hive LLAP 使 Spark 代码与 Hive 表交互并在 Spark 引擎上使用启用了 ACID 的 Hive 表。

有关详细信息，请参阅《*Data Engineering Integration 10.4.1 用户指南*》中的“映射优化”一章。

将 Snowflake 作为流映射目标

对于 Data Engineering Streaming，您可以将 Snowflake 配置为流映射的目标，以将数据写入 Snowflake。

新增功能 (10.4.0.1)

Data Engineering Integration

本节介绍版本 10.4.0.1 中新增的 Data Engineering Integration 功能。

CLAIRE 建议和见解

从版本 10.4.0.1 开始，您可以对多选各个映射运行 CLAIRE 分析。

您可以多选多个映射，然后右键单击并选择**分析**，以读取有关所选映射之间的相似度的 CLAIRE 见解。

在以前的版本中，您只能对一个项目或文件夹运行分析。

有关建议和见解的详细信息，请参阅《*Data Engineering Integration 用户指南*》。

Enterprise Data Catalog

本节介绍版本 10.4.0.1 中新增的 Enterprise Data Catalog 功能。

新资源

从版本 10.4.0.1 开始，Enterprise Data Catalog 中添加了以下新资源：

- Apache Kafka
- Qlik Sense
- MongoDB

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 10.4.0.1 Enterprise Data Catalog 扫描程序配置指南*》。

Informatica PowerExchange 适配器

本节介绍版本 10.4.0.1 中新增的 Informatica 适配器功能。

PowerExchange for Microsoft Azure Data Lake Storage Gen2

从版本 10.4.0.1 开始，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。

有关详细信息，请参阅《*Informatica 10.4.0 PowerExchange for Microsoft Azure Data Lake Storage Gen2 用户指南*》。

更改 (10.4.0.1)

Enterprise Data Catalog (10.4.0.1 更改)

本节介绍版本 10.4.0.1 中对 Enterprise Data Catalog 进行的更改。

精简视图

从版本 10.4.0.1 开始，您可以在精简视图中查看转换逻辑。在以前的版本中，您无法在精简视图中查看转换逻辑。

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

Data Engineering Integration

本节介绍版本 10.4.0.1 中对 Data Engineering Integration 进行的更改。

Spark 引擎上的数据预览

从版本 10.4.0.1 开始，使用 Spark Jobserver 在 Spark 引擎上预览数据时，您可以使用 HTTPS。

如果在数据集成服务上启用 HTTPS 协议，则 Spark Jobserver 也使用 HTTPS 协议。Spark Jobserver 使用您通过 Administrator 工具在数据集成服务进程属性中设置的相同 HTTPS 密钥库配置。

有关详细信息，请参阅《*Informatica® 10.4.0 应用程序服务指南*》。

Hive 仓库连接器和 Hive LLAP

从版本 10.4.0.1 开始，在 Spark 引擎上运行映射时，可以启用 Hive 仓库连接器和 Hive LLAP。以前，这些功能在技术预览版中提供。

使用 Spark 引擎从 Hive 表读取数据以及将数据写入到其中时，使用 Hive 仓库连接器和 Hive LLAP 可以更快执行 Hive 查询。您可以在 Spark 引擎上对 Hortonworks HDP 3.x 和 Microsoft Azure HDInsight 4.x 群集使用 Hive Warehouse Connector 和 Hive LLAP。

要启用 Hive 仓库连接器和 Hive LLAP，则在 Hadoop 连接的 Spark 高级属性中配置以下属性：

- `infaspark.useHiveWarehouseAPI=true`
- `spark.datasource.hive.warehouse.load.staging.dir=<用于写入到 Hive 表的 HDFS 临时目录>`
- `spark.datasource.hive.warehouse.metastoreUri=<Hive 元存储的 URI>`
使用 `hive_site.xml` 群集配置属性中 `hive.metastore.uris` 的值。
- `spark.hadoop.hive.llap.daemon.service.hosts=<LLAP 服务的应用程序名称>`
使用 `hive_site.xml` 群集配置属性中 `hive.llap.daemon.service.hosts` 的值。
- `spark.hadoop.hive.zookeeper.quorum=<Hive LLAP 使用的 Zookeeper 主机>`
使用 `hive_site.xml` 群集配置属性中 `hive.zookeeper.quorum` 的值。
- `spark.sql.hive.hivensever2.jdbc.url=<HiveServer2 Interactive 的 URL。使用 Ambari for HiveServer2 JDBC URL 中的值。>`

有关详细信息，请参阅《*Data Engineering Integration 10.4.1 用户指南*》。

Informatica PowerExchange 适配器

本节介绍版本 10.4.0.1 中对 Informatica 适配器进行的更改。

PowerExchange for Microsoft Azure Data Lake Storage Gen2

从版本 10.4.0.1 开始，FileName 端口将对现有 10.4.0 动态映射中使用的源和目标显示。要成功运行现有动态映射，请添加表达式等转换，以从源对象和目标对象中消除 FileName 端口。

有关详细信息，请参阅《*Informatica 10.4.0 PowerExchange for Microsoft Azure Data Lake Storage Gen2 用户指南*》。

第 9 章

通知 (10.4.0)

本章包括以下主题：

- [产品名称更改, 73](#)
- [升级到版本 10.4.0, 73](#)
- [支持更改, 74](#)
- [发布任务, 77](#)

产品名称更改

从版本 10.4.0 开始，Informatica 实施了某些名称更改。

Big Data 产品系列已重命名为 Data Engineering。以下产品名称已更改：

- Big Data Management 更改为 Data Engineering Integration。
- Big Data Quality 更改为 Data Engineering Quality。
- Big Data Streaming 更改为 Data Engineering Streaming。
- Big Data Masking 更改为 Data Engineering Masking。

Enterprise Data Catalog 和 Enterprise Data Preparation 均已列入 Data Catalog 产品系列。

升级到版本 10.4.0

安装程序更改

从版本 10.4.0 开始，Informatica 安装程序进行了以下更改：

- 可以运行 10.4.0 安装程序来安装 Data Engineering、Data Catalog 和传统产品。尽管您可以在 Data Engineering 和 Data Catalog 产品所在的相同域中安装传统产品，但 Informatica 建议您在单独域中安装传统产品。
- 可以运行 10.4.0 安装程序来升级 Data Engineering、Data Catalog 和传统产品。
- 创建域时，可以选择创建 PowerCenter 存储库服务和 PowerCenter 集成服务。

从版本 10.4.0 开始，Informatica 升级进行了以下更改：

- 搜索服务将创建新的索引文件夹并重新编制搜索对象索引。无需在升级之后执行重新编制索引操作。

在高精度或低精度模式下运行映射

从版本 10.2.2 开始，使用 Spark 引擎的映射默认以高精度模式运行。

您可以使用 Developer tool 在映射运行时配置高级设置中禁用高精度设置：

1. 在 Developer tool 中，选择窗口 > 首选项。
2. 选择 **Informatica > 运行配置 > 映射**。
3. 选择高级选项卡。
4. 取消选择**高精度**。

要在 Spark 引擎上以低精度模式运行早期发行版中开发的映射，建议您执行上述操作。在某些情况下，例如未在映射中指定小数位数，则以高精度模式（默认设置）运行映射会失败，并且需要以低精度模式运行。

从版本 10.4.0 开始，以下额外更改会影响在 Spark 引擎上运行的映射。

- 小数到字符串转换不再将小数附加到整数。例如，小数格式的整数 1 的值在字符串格式中保持为“1”。
- 在小数到字符串转换中，小数点后的尾随零会被截断。例如，小数格式的 1.000 的值在字符串格式中将呈现为“1”。
- 在低精度模式下运行映射时，精度设置超过 15 的小数到字符串转换将导致采用指数计数法。

支持更改

本节介绍版本 10.4.0 中的支持更改。

技术预览支持

已启动技术预览

从版本 10.4.0 开始，Informatica 包含以下技术预览功能：

连接至区块链

对于 Data Engineering Integration，您可以连接至区块链，以在运行于 Spark 引擎上的映射中使用区块链源和目标。

对网络上运行的数据集成服务进行数据预览

对于 Data Engineering Integration，在网络上运行数据集成服务时，您可以预览 Spark 引擎上的层次结构数据。

将 Databricks 增量表作为流映射目标

对于 Data Engineering Streaming，您可以使用 Databricks 增量表作为流映射的目标，以引入流数据。

动态流映射

您可以配置动态流映射，以根据您在 Confluent 架构注册表中定义的参数和规则更改运行时 Kafka 源和目标。

智能结构模型中的 HL7 输入

Intelligent Structure Discovery 可以处理 HL7 输入。

Databricks 上的 Python 转换

对于 Data Engineering Integration，您可以在配置为在 Databricks Spark 引擎上运行的映射中包含 Python 转换。

将 Snowflake 作为流映射目标

对于 Data Engineering Streaming，您可以将 Snowflake 配置为流映射的目标，以将数据写入 Snowflake。

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。

已提升技术预览

从版本 10.4.0 开始，提升了技术预览中的以下功能：

Spark 引擎上的数据预览

在 Data Engineering Integration 中，对于配置为使用 Amazon EMR、Cloudera CDH 和 Hortonworks HDP 运行的映射，您可以通过 Developer tool 预览在 Spark 引擎上运行的映射中的数据。在技术预览中，仍可预览配置为使用 Azure HDInsight 和 MapR 运行的映射中的数据。

PowerExchange for Amazon S3

对于 Data Engineering Integration，您可以在导入数据对象时使用智能结构模型。

PowerExchange for Microsoft Azure Cosmos DB SQL API

对于 Data Engineering Integration，您可以在 Azure Databricks 环境中开发和运行映射。

PowerExchange for Microsoft Azure SQL Data Warehouse

对于 Data Engineering Integration，您可以使用以下功能：

- 创建和运行动态映射。
- 在使用 ODBC 连接来连接 Microsoft Azure SQL Data Warehouse 数据库时，可以使用完全下推优化。

已启用 SSL 的 Kafka 连接

对于 Data Engineering Streaming，您可以对流映射使用已启用 SSL 的 Kafka 连接。

延期

本节介绍版本 10.4.0 中的延期更改。

已提升延期

从版本 10.4.0 开始，以下功能不再延期：

- 流映射中的数据屏蔽转换。
- Kerberos 跨域身份验证。
- 监视流传输作业的统计信息。

删除的支持

从版本 10.4.0 开始，Informatica 停止了对 Solaris 的支持。如果您正在使用 Solaris，Informatica 将要求您进行升级，以使用受支持的操作系统。

有关如何升级到受支持操作系统的详细信息，请参阅 Informatica 10.4.0 升级指南。有关受支持操作系统的信息，请参阅 Informatica Network 上的“产品可用性列表”：

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

PowerCenter

本节介绍版本 10.4.0 中的 PowerCenter 支持更改。

连接

本节介绍版本 10.4.0 中的连接支持更改。

SAP HANA 许可

从版本 10.4.0 开始，您需要 SAP HANA 许可证才能从 SAP HANA 源读取数据以及将数据写入 SAP HANA 目标。

在 ODBC 连接中，如果 **ODBC 子类型** 未设置为 **SAP HANA** 且 SAP HANA 许可证不可用，则会话将在运行时失败。

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

PowerCenter PowerExchange 适配器

本节介绍 10.4.0 中对 PowerCenter 适配器进行的支持更改。

PowerExchange for SAP NetWeaver

从版本 10.4.0 开始，PowerExchange for SAP NetWeaver 包括以下更改：

- Informatica 停止支持非 Unicode 传输。
在以前的版本中，Informatica 支持非 Unicode 传输。
- Informatica 在以下文件夹中提供了用于 SAP Unicode 版本 5.0 及更高版本的传输：
 - Unicode cofiles: Informatica 安装程序 zip 文件/saptrans/mySAP/cofiles
 - Unicode 数据文件: Informatica 安装程序 zip 文件/saptrans/mySAP/data在以前的版本中，Informatica 打包了以下文件夹中用于 SAP Unicode 版本 5.0 及更高版本的传输：
 - Unicode cofile: Informatica 安装程序 zip 文件/saptrans/mySAP/UC/cofiles
 - Unicode 数据文件: Informatica installer zip file/saptrans/mySAP/UC/data
- Informatica 停止支持使用 PowerExchange for SAP NetWeaver 通过 HTTP/HTTPS 流传输从 SAP 表读取数据。请使用 PowerExchange for SAP Dynamic ABAP 表提取程序通过 HTTP/HTTPS 流传输从 SAP 表读取数据。
在以前的版本中，Informatica 支持使用 PowerExchange for SAP NetWeaver 通过 HTTP/HTTPS 流传输从 SAP 表读取数据。

有关详细信息，请参阅《Informatica 10.4.0 PowerExchange for SAP NetWeaver 用户指南》和《PowerExchange for SAP NetWeaver 10.4.0 传输版本安装说明》。

Informatica PowerExchange 适配器

本节介绍 10.4.0 中对 Informatica 适配器进行的支持更改。

PowerExchange for SAP NetWeaver

从版本 10.4.0 开始，PowerExchange for SAP NetWeaver 包括以下更改：

- Informatica 在以下文件夹中提供了用于 SAP Unicode 版本 5.0 及更高版本的传输：
 - Unicode cofiles: Informatica 安装程序 zip 文件/saptrans/mySAP/cofiles
 - Unicode 数据文件: Informatica 安装程序 zip 文件/saptrans/mySAP/data在以前的版本中，Informatica 打包了以下文件夹中用于 SAP Unicode 版本 5.0 及更高版本的传输：
 - Unicode cofile: Informatica 安装程序 zip 文件/saptrans/mySAP/UC/cofiles

- Unicode 数据文件: Informatica installer zip file/saptrans/mySAP/UC/data

有关详细信息, 请参阅 《PowerExchange for SAP NetWeaver 10.4.0 传输版本安装说明》。

发布任务

本节介绍版本 10.4.0 中的发布任务。发布任务是指升级到版本 10.4.0 后必须执行的任务。

Data Engineering Integration

本节介绍版本 10.4.0 中 Data Engineering Integration 的发布任务。

Python 转换

从版本 10.4.0 开始, Python 转换中的 Python 代码组件将分为以下几个选项卡:

- 输入前。定义可以解释一次并在所有数据行之间共享的代码。
- 输入。定义 Python 转换在处理分区时收到输入行后的行为。
- 结尾。定义 Python 转换在处理分区中的所有输入数据后的行为。

在升级后的映射中, 您在 Python 代码组件中输入的代码会在**输入**选项卡中显示。

查看代码, 以确认代码按预期工作。如有必要, 请使用**输入前**、**输入**和**结尾**选项卡重构代码。

有关如何在每个选项卡中配置代码的信息, 请参阅 《Informatica Data Engineering Integration 10.4.0 用户指南》中的“Python 转换”一章。

第 10 章

新产品(10.4.0)

本章包括以下主题：

- [Informatica PowerExchange 适配器, 78](#)
- [PowerCenter PowerExchange 适配器, 79](#)

Informatica PowerExchange 适配器

本节介绍版本 10.4.0 中新的 Informatica 适配器。

PowerExchange for JDBC V2

从版本 10.4.0 开始，您可以创建 JDBC V2 连接以连接至 Aurora PostgreSQL、Azure SQL Database 或任何支持 Type 4 JDBC 驱动程序数据库。您必须在 JDBC V2 连接属性中选择所需的数据库子类型。您可以在映射中使用 JDBC V2 连接，以在使用 Type 4 JDBC 驱动程序数据库读取或写入数据。您可以在本地环境或非本地环境中验证并运行 JDBC V2 映射。在非本地环境中，您可以在 Hadoop 环境中的 Spark 引擎或者在 Databricks 上运行映射。

有关详细信息，请参阅《*Informatica 10.4.0 PowerExchange for JDBC V2 用户指南*》。

PowerExchange for Microsoft Azure Data Lake Storage Gen2

从版本 10.4.0 开始，您可以创建 Microsoft Azure Data Lake Storage Gen2 连接以连接至 Microsoft Azure Data Lake Storage Gen2。您可以在映射中使用 Microsoft Azure Data Lake Storage Gen2 连接，以在 Microsoft Azure Data Lake Storage Gen2 对象中读取或写入数据。您可以在本地环境或非本地环境中验证并运行 Microsoft Azure Data Lake Storage Gen2 映射。在非本地环境中，您可以在 Hadoop 环境中的 Spark 引擎或者在 Databricks 上运行映射。

有关详细信息，请参阅《*Informatica 10.4.0 PowerExchange for Microsoft Azure Data Lake Storage Gen2 用户指南*》。

PowerExchange for Salesforce Marketing Cloud

从版本 10.4.0 开始，您可以创建 Salesforce Marketing Cloud 连接以连接至 Salesforce Marketing Cloud。您可以在映射中使用 Salesforce Marketing Cloud 连接以在 Salesforce Marketing Cloud 中读取或写入联系人数数据。使用 Salesforce Marketing Cloud 数据扩展作为映射中的源或目标。可以在本地环境中验证和运行 Salesforce Marketing Cloud 映射。

有关详细信息，请参阅《*Informatica 10.4.0 PowerExchange for Salesforce Marketing Cloud 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。

PowerExchange for Db2 Warehouse

从版本 10.4.0 开始，您可以创建 Db2 Warehouse 连接以从 PowerCenter 连接至 IBM Db2 Warehouse。您可以将 Db2 Warehouse 对象作为源和目标导入以创建映射，并运行会话以在 Db2 Warehouse 中读取或写入数据。

有关详细信息，请参阅《*Informatica 10.4.0 PowerExchange for Db2 Warehouse 用户指南*》。

PowerExchange for Microsoft Dynamics 365 for Sales

从版本 10.4.0 开始，您可以创建 Microsoft Dynamics 365 for Sales 连接以从 PowerCenter 连接至 Microsoft Dynamics 365 for Sales。您可以将 Microsoft Dynamics 365 for Sales 对象作为源和目标导入以创建映射，并运行会话以在 Microsoft Dynamics 365 for Sales 中提取读取数据或写入数据。在映射中使用 Microsoft Dynamics 365 for Sales 对象时，可以配置特定于 Microsoft Dynamics 365 for Sales 的属性。

有关详细信息，请参阅《*Informatica 10.4.0 PowerExchange for Microsoft Dynamics 365 for Sales 用户指南*》。

PowerExchange for PostgreSQL

从版本 10.4.0 开始，您可以创建 PostgreSQL 连接以从 PowerCenter 连接至 Aurora PostgreSQL。您可以将 PostgreSQL 对象作为源和目标导入以创建映射，并运行会话以在 PostgreSQL 中提取读取数据或写入数据。在映射中使用 PostgreSQL 对象时，可以配置特定于 PostgreSQL 的属性。

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

第 11 章

新增功能 (10.4.0)

本章包括以下主题：

- [CI/CD, 80](#)
- [命令行程序, 84](#)
- [Data Engineering Integration, 85](#)
- [Data Engineering Streaming, 88](#)
- [Enterprise Data Catalog, 90](#)
- [Enterprise Data Preparation, 94](#)
- [Informatica 映射, 95](#)
- [Informatica 转换, 96](#)
- [Informatica 工作流, 96](#)
- [Informatica 安装, 97](#)
- [智能结构模型, 97](#)
- [PowerCenter, 98](#)
- [PowerExchange 适配器, 99](#)
- [安全, 103](#)

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。

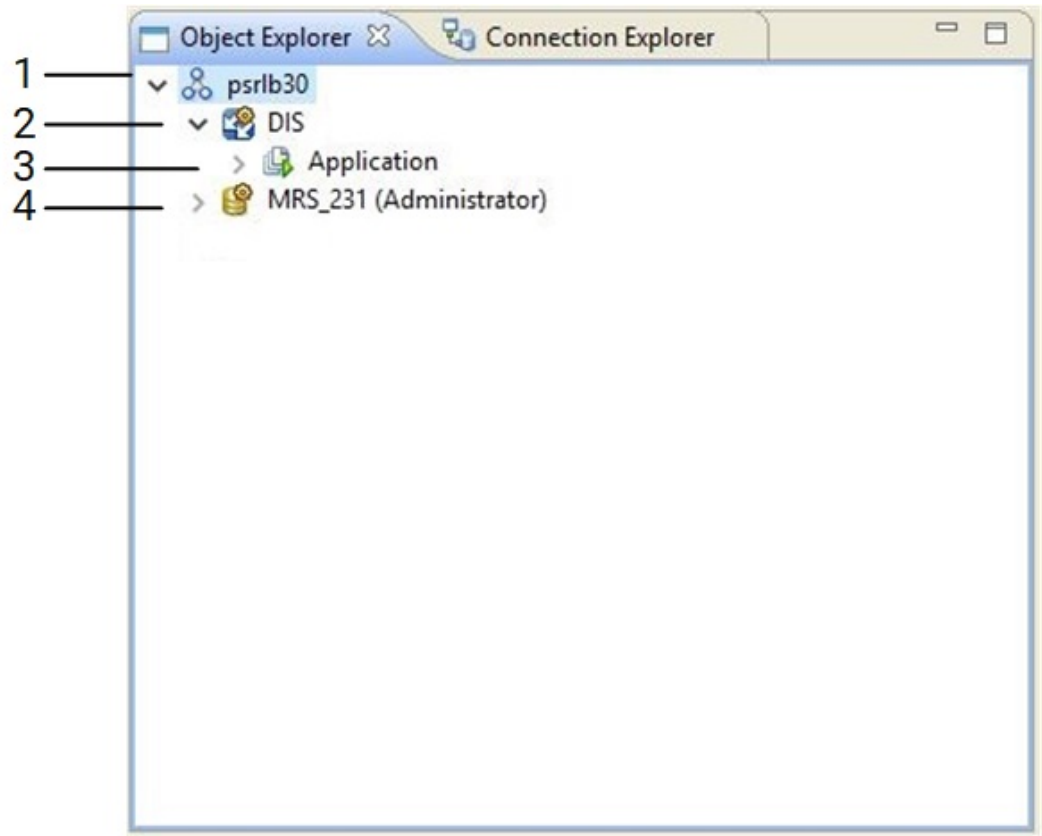
连接到运行时应用程序后，您在 Developer tool 中执行的搜索可查找应用程序中的运行时对象。

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 指南*。有关创建映射以访问 Delta Lake 表的信息，请参阅 *Data Engineering Integration 用户指南*。

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 引擎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。
在通过 Sqoop 连接建立连接的 Oracle 数据库上受支持。
- **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 用户指南*。

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 用于为应用程序中优化级别为“正常”的所有映射设置优化器级别。该命令不会影响应用程序中优化级别不是“正常”的映射。

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 12 章

更改 (10.4.0)

本章包括以下主题：

- [Data Engineering Integration, 104](#)
- [Enterprise Data Preparation, 106](#)
- [Enterprise Data Catalog, 106](#)
- [Informatica Developer, 107](#)
- [Informatica 转换, 107](#)
- [PowerCenter, 108](#)
- [Informatica PowerExchange 适配器, 109](#)
- [安全, 110](#)

Data Engineering Integration

本节介绍版本 10.4.0 中对 Data Engineering Integration 进行的更改。

数据预览

从版本 10.4.0 开始，数据集成服务使用 Spark Jobserver 在 Spark 引擎上预览数据。Spark Jobserver 能够加快数据预览作业，因为它会维护运行中的 Spark 上下文，而不是刷新每个作业的上下文。配置为使用 Amazon EMR、Cloudera CDH 和 Hortonworks HDP 运行的映射使用 Spark Jobserver 预览数据。

在以前的版本中，数据集成服务对 Spark 引擎上的所有数据预览作业使用 spark-submit 脚本。配置为使用 Azure HDInsight 和 MapR 运行的映射使用 spark-submit 脚本在 Spark 引擎上预览数据。在技术预览中，可以预览配置为使用 Azure HDInsight 和 MapR 运行的映射的数据。

有关详细信息，请参阅《*Data Engineering Integration 10.4.0 用户指南*》中的“数据预览”一章。

联合转换

从版本 10.4.0 开始，您在预览数据时可以选择联合转换作为预览点。在以前的版本中，不支持将联合转换作为预览点。

infacmd dp 命令

您可以使用 infacmd dp 插件执行数据预览操作和。使用 infacmd dp 命令可以手动启动和停止 Spark Jobserver。

下表介绍了 infacmd dp 命令：

命令	说明
startSparkJobServer	在集成服务计算机上启动 Spark Jobserver。 默认情况下，当您预览层次结构数据时启动 Spark Jobserver。
stopSparkJobServer	停止指定集成服务上运行的 Spark Jobserver。 默认情况下，如果 Spark Jobserver 空闲时间达到 60 分钟，或者停止或重新启动数据集成服务时，Spark Jobserver 会停止。

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

Databricks 中的日期/时间格式

从版本 10.4.0 开始，Databricks Spark 引擎在读取和写入日期/时间值时使用格式 YYYY-MM-DD HH24:MM:SS.US。

在以前的版本中，在 Developer tool 运行时首选项的映射属性中设置此格式。

您可能需要执行其他任务，才能继续在 Databricks 引擎中使用日期/时间数据。有关详细信息，请参阅《*Data Engineering Integration 10.4.0 指南*》中的“Databricks 集成”一章。

目标中的空值

从版本 10.4.0 开始，向复杂文件写入数据时，以下更改适用：

- 如果映射源包含空值且您使用**创建目标**选项来创建 Parquet 目标文件，则默认架构包含可选字段，您可以向目标中插入空值。
在以前的版本中，默认架构中的所有字段都创建为“必填”字段，您必须在目标架构中手动将数据类型从“必填”更新为“可选”，才能将包含空值的列写入目标。
- 如果映射源包含空值且您使用**创建目标**选项来创建 Avro 目标文件，则在默认架构中定义空值，您可以向目标文件中插入空值。
在以前的版本中，默认架构中不定义空值，您必须手动更新默认目标架构，才能向架构中添加“空”数据类型。

注意：如果您不允许在目标中使用空值，可以手动编辑架构。您不能通过编辑架构来阻止启用了映射流的目标使用空值。

这些更改适用于以下适配器：

- 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

Python 转换

从版本 10.4.0 开始，您可以通过在数组 resourceFilesArray 中引用索引来访问 Python 代码中的资源文件。在您使用版本 10.4.0 创建的新映射中使用 resourceFilesArray。

在以前的版本中，该数组名为 resourceJepFile。使用 resourceJepFile 的升级后的映射将继续成功运行。

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。

Enterprise Data Preparation

本节介绍版本 10.4.0 中对 Enterprise Data Preparation 进行的更改。

空值处理

从版本 10.4.0 开始，Enterprise Data Preparation 会在 Informatica Data Engineering 处理行为之后处理空值。升级应用程序后，包含空值或错误值的工作表列所用的计算配方中的公式会返回空值。

打开在以前版本中创建的项目时，将有一个对话框提示您升级项目中的工作表。如果您选择升级工作表，应用程序会重新计算项目的每个工作表中的公式，然后使用新公式结果更新工作表。

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

Solr 端口属性

从版本 10.4.0 开始，交互数据准备服务中将删除 Solr 端口属性。

在以前的版本中，Enterprise Data Preparation 应用程序在数据准备期间使用 Apache Solr 提供添加到配方的步骤建议。该应用程序现在使用内部算法提供添加到配方的步骤建议。

Enterprise Data Catalog

本节介绍版本 10.4.0 中对 Enterprise Data Catalog 进行的更改。

SAP HANA 资源的沿袭增强功能

从版本 10.4.0 开始，您可以使用 SAP HANA 资源查看以下视图的沿袭：

- 表和普通视图
- 计算视图
- 分析视图

从版本 10.4.0 开始，Enterprise Data Catalog 现在支持剖析 SAP HANA 数据库表和视图，以提取列剖析和域发现统计信息。

剖析和数据域发现

从版本 10.4.0 开始，您可以为 Google BigQuery 资源启用列剖析和数据域发现。

Informatica Data Engineering Streaming

从版本 10.4.0 开始，Informatica Platform 资源支持从包含流源的流映射中提取元数据。流数据源作为引用对象进行创建。

支持的流源为 Apache Kafka。Apache Kafka 和 Hive 数据源支持强类型引用对象。

资源名称更改

从版本 10.4.0 开始，Profiling Warehouse 资源重命名为 Informatica Data Quality。

搜索建议

从版本 10.4.0 开始，Enterprise Data Catalog 现在可将业务标题和资产名称显示为搜索建议中的可能匹配项。在以前的版本中，Enterprise Data Catalog 仅将资产名称显示为搜索建议中的可能匹配项。

配置自定义属性

从版本 10.4.0 开始，您可以使用**高**和**低**选项对自定义属性的搜索等级进行优先级排序。在以前的版本中，您可以使用**高**、**中**和**低**选项对自定义属性的搜索等级进行优先级排序。

Informatica Developer

本节介绍版本 10.4.0 中对 Informatica Developer 的更改。

导入关系数据对象

从版本 10.4.0 开始，当导入的资源与现有物理数据对象同名时，Developer tool 允许您选择如何解决冲突。您可以选择使用其他名称创建对象、重用现有对象或替换现有对象的资源。

在以前的版本中，Developer tool 会停止导入表，且不会尝试导入任何后续表。

有关详细信息，请参阅《*Informatica 10.4.0 Developer tool 指南*》。

Informatica 转换

本节介绍版本 10.4.0 中对 Informatica 转换的更改。

地址验证器转换

本节介绍地址验证器转换的更改。

地址验证器转换对地址功能做了以下更新：

所有国家/地区

从版本 10.4 开始，地址验证器转换使用 5.15.0 版本的 Informatica Address Verification 软件引擎。

在以前的版本中，该转换使用 5.14.0 版本的 Informatica Address Verification 软件引擎。

从版本 10.4 开始，如果引用数据中不含国家/地区的省份信息，地址验证器转换将在输出地址中保留省份信息。如果不含省份数据的输出地址有效，Address Verification 将返回 V2 匹配得分，以指示输入地址正确，但引用数据库不包含地址中的每个元素。

在以前的版本中，如果地址引用数据不含国家/地区的省份信息，Address Verification 会将省份信息移至“其他”字段并返回 Cx 得分。

西班牙

从版本 10.4 开始，对于西班牙地址，地址验证器转换会返回 Ix 状态，指示需要进行有效更正，才能在引用数据中生成匹配项。

在以前的版本中，Address Verification 可能会更正需要进行有效更改的地址，并为该地址返回过于乐观的匹配得分。

美国

从版本 10.4 开始，如果交货地址行中的组织信息位于街道信息之前，地址验证器转换可以验证美国地址。转换可以识别的组织类型包括大学、医院和公司办公室。如果解析操作还能在交货地址行的街道信息中找到门牌号和街道类型，则 Address Verification 可以识别组织信息。

在以前的版本中，如果交货地址行中的组织信息位于街道信息之前，Address Verification 会返回 Ix 匹配得分。

有关 Informatica Address Verification 软件引擎更新的详细信息，请参阅《*Informatica Address Verification 5.15.0 发行指南*》。

PowerCenter

本节介绍 10.4.0 中对 PowerCenter 进行的更改。

刷新 Designer 和 Workflow Manager 中的元数据

从版本 10.4.0 开始，无需中断连接状态，即可刷新 Workflow Manager 和 Designer 中的存储库和文件夹。创建、删除、修改文件夹后或将对象导入 PowerCenter 客户端后，存储库与文件夹将进行更新。

要刷新存储库中的文件夹列表，请右键单击存储库并选择“刷新文件夹列表”。要刷新文件夹及文件夹中的所有内容，请右键单击文件夹并选择“刷新”。

在以前的版本中，您需要断开连接并重新连接 PowerCenter 客户端，才能查看存储库或文件夹的更新。

有关详细信息，请参阅《*PowerCenter 10.4.0 存储库指南*》、《*PowerCenter 10.4.0 Designer 指南*》和《*PowerCenter 10.4.0 工作流基本指南*》。

导入和导出

从版本 10.4.0 开始，您可以导入和导出映射、会话、工作流和工作集。

要将数据从 PowerCenter 导入到模型存储库，请完成以下任务：

1. 使用 PowerCenter 客户端或以下命令将 PowerCenter 对象导出到文件：
`pmrep ExportObject`
2. 使用以下命令将导出文件转换为模型存储库文件：
`infacmd ipc importFromPC`
3. 使用 Developer tool 或以下命令导入对象：
`infacmd tools importObjects`

要将数据从模型存储库导出到 PowerCenter 存储库，请完成以下任务：

1. 使用 Developer tool 或以下命令将模型存储库对象导出到文件：
`infacmd tools ExportObjects`
或者，您也可以直接运行 `infacmd ipc ExportToPC` 执行导出。
2. 使用以下命令将导出文件转换为 PowerCenter 文件：
`infacmd ipc ExporttoPC`
3. 使用 PowerCenter 或以下命令导入对象：
`pmrep importObjects`

在版本 10.2.2 和 10.2.1 中，您可以使用安装程序补丁从 PowerCenter 导入，但无法导出到 PowerCenter。在版本 10.2.1 之前，您可以在 PowerCenter 中导入和导出。

有关详细信息，请参阅《Informatica 10.4.0 Developer tool 指南》、《Informatica 10.4.0 Developer 映射指南》、《Informatica 10.4.0 Developer 工作流指南》和《Informatica 10.4.0 命令引用指南》。

Informatica PowerExchange 适配器

本节介绍版本 10.4.0 中对 Informatica 适配器的更改。

PowerExchange for Amazon Redshift

从版本 10.4.0 开始，PowerExchange for Amazon Redshift 包含以下更改：

- 无需在 Amazon S3 桶策略中添加 GetBucketPolicy 权限即可连接至 Amazon Redshift。在以前的版本中，必须在 Amazon S3 桶策略中添加 GetBucketPolicy 权限才能连接至 Amazon Redshift。具有 GetBucketPolicy 权限的现有 Amazon S3 桶策略可以继续使用，无需进行任何更改。
- Amazon Redshift 数据对象读取操作的以下高级属性已更改：

旧属性	新属性
S3 服务器端加密 S3 客户端加密	加密类型 可以选择下列值之一： <ul style="list-style-type: none">- 无- SSE-S3- SSE-KMS- CSE-SMK

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，无需在 Amazon S3 桶策略中添加 GetBucketPolicy 权限即可连接至 Amazon S3。

在以前的版本中，必须在 Amazon S3 桶策略中添加 GetBucketPolicy 权限才能连接至 Amazon S3。

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

PowerExchange for Microsoft Azure Blob 存储

从版本 10.4.0 开始，在 Databricks Spark 引擎上运行传递映射时，只有连接的字段才会写入平面文件目标。在以前的版本中，数据集成服务会将包含空值的未连接字段名称写入平面文件目标。

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 Data Lake Storage Gen1

从版本 10.4.0 开始，在 Databricks Spark 引擎上运行传递映射时，只有连接的字段才会写入平面文件目标。在以前的版本中，数据集成服务会将包含空值的未连接字段名称写入平面文件目标。

有关详细信息，请参阅《*Informatica 10.4.0 PowerExchange for Microsoft Azure Data Lake Storage Gen1 用户指南*》。

安全

本节介绍版本 10.4.0 中对安全功能的更改。

infacmd isp 命令

下表介绍了更改后的 infacmd isp 命令：

命令	说明
addNameSpace	向命令中添加所需的 -ln 选项。您可以使用此选项指定 LDAP 配置的名称。
listLDAPConnectivity	向命令中添加所需的 -ln 选项。
setLDAPConnectivity	此命令已重命名为 addLDAPConnectivity。使用新命令语法更新使用 setLDAPConnectivity 的所有脚本。 向命令中添加所需的 -ln 选项。
updateNameSpace	向命令中添加所需的 -ln 选项。

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

LDAP 目录服务

从版本 10.4.0 开始，Informatica 支持以下 LDAP 目录服务：

- Oracle Directory Server (ODSEE)
- Oracle Unified Directory (OUD)

您还可以配置域，以使用 Informatica 不认证的 LDAP 目录服务。

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

LDAP 配置

从版本 10.4.0 开始，您可以配置 Informatica 域，以便从一个或多个 LDAP 目录服务导入的用户能够登录 Informatica 节点、服务和应用程序客户端。

在以前的版本中，您可以配置 Informatica 域从单个 LDAP 目录服务导入用户。

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

SAML 身份验证

从版本 10.4.0 开始，Informatica 支持以下身份提供程序：

- Microsoft Active Directory 联合身份验证服务 (AD FS) 2.0
- Microsoft Active Directory 联合身份验证服务 (AD FS) 4.0
- PingFederate9.2.2

在以前的版本中，Informatica 仅支持 AD FS 2.0 身份提供程序。

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

第 III 部分： 版本 10.2.2

本部分包含以下章节：

- [通知、新增功能和更改 \(10.2.2 HotFix 1\), 113](#)
- [通知、新增功能和更改 \(10.2.2 Service Pack 1\), 121](#)
- [通知 \(10.2.2\), 128](#)
- [新产品\(10.2.2\), 134](#)
- [新增功能 \(10.2.2\), 135](#)
- [更改 \(10.2.2\), 162](#)

第 13 章

通知、新增功能和更改 (10.2.2 HotFix 1)

本章包括以下主题：

- [通知 \(10.2.2 HotFix 1\), 113](#)
- [新增功能 \(10.2.2 HotFix 1\), 114](#)
- [更改 \(10.2.2 HotFix 1\), 118](#)

通知 (10.2.2 HotFix 1)

支持更改

本节介绍版本 10.2.2 HotFix 1 Service Pack 1 中的支持更改。

技术预览支持

已启动技术预览

从版本 10.2.2 HotFix 1 Service Pack 1 开始，Informatica 包含以下技术预览功能：

Hive 仓库连接器和 Hive LLAP

对于大数据产品，对 Azure HDInsight 4.x 和 Hortonworks HDP 3.x 使用 Hive 仓库连接器和 Hive LLAP 使 Spark 代码与 Hive 表交互并在 Spark 引擎上使用启用了 ACID 的 Hive 表。

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。

已提升技术预览

从版本 10.2.2 HotFix 1 Service Pack 1 开始，在技术预览中提升了 Hortonworks HDP 3.1。

升级到 HDP 3.1 可能会影响管理的 Hive 表。在升级之前，请查看知识库文章

[What Should Big Data Customers Know About Upgrading to Hortonworks HDP 3.1?](#) 中的 HDP 3.1 升级信息和 Cloudera 已知限制。请联系 Informatica 全球客户服务部门或 Cloudera 专业服务验证您的 HDP 3.1 升级计划。

新增功能 (10.2.2 HotFix 1)

命令程序

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。

Big Data Management

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

Azure Data Lake Storage Gen2

从版本 10.2.2 HotFix 1 Service Pack 1 开始，可以从与 HDInsight 群集关联的 Azure Data Lake Storage Gen2 存储帐户中读取映射，也可以将映射写入其中。

有关如何将 Big Data Management 与使用 ADLS Gen2 存储的群集进行集成的详细信息，请参阅《*Big Data Management 集成指南*》。有关如何对 ADLS Gen2 源和目标使用映射的信息，请参阅《*Big Data Management 用户指南*》。

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 HotFix 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。

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*》。

Informatica PowerExchange 适配器

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

PowerExchange for Google Analytics

从版本 10.2.2 HotFix 1 Service Pack 1 开始，可以在以下发行版的 Spark 引擎上运行 PowerExchange for Google Analytics 映射：

- Amazon EMR 版本 5.20
- Google Cloud Dataproc 版本 1.3
- Hortonworks HDP 版本 3.1

PowerExchange for Google BigQuery

从版本 10.2.2 HotFix 1 Service Pack 1 开始，可以在以下发行版的 Spark 引擎上运行 PowerExchange for Google BigQuery 映射：

- Amazon EMR 版本 5.20
- Google Cloud Dataproc 版本 1.3
- Hortonworks HDP 版本 3.1

PowerExchange for Google Cloud Spanner

从版本 10.2.2 HotFix 1 Service Pack 1 开始，可以在以下发行版的 Spark 引擎上运行 PowerExchange for Google Cloud Spanner 映射：

- Amazon EMR 版本 5.20
- Google Cloud Dataproc 版本 1.3
- Hortonworks HDP 版本 3.1

PowerExchange for Google Cloud Storage

从版本 10.2.2 HotFix 1 Service Pack 1 开始，可以在以下发行版的 Spark 引擎上运行 PowerExchange for Google Cloud Storage 映射：

- Amazon EMR 版本 5.20
- Google Cloud Dataproc 版本 1.3
- Hortonworks HDP 版本 3.1

更改 (10.2.2 HotFix 1)

Data Transformation

本节介绍版本 10.2.2 HotFix 1 中对 Data Transformation 进行的更改。

Business Intelligence and Reporting Tools (BIRT)

从版本 10.2.2 HotFix 1 开始，Business Intelligence and Reporting Tools (BIRT) 不再与 Data Transformation 打包提供。

您可以从以下文件中提到的位置下载 BIRT：

<Data Transformation 安装目录>/readme_Birt.txt

如果在下载之前尝试通过 Data Transformation 使用 BIRT，可能会出现以下错误：

未找到 BIRT Report Engine。请参阅 [DT-home]/readme_Birt.txt 中的下载说明。

之前，BIRT 与 Data Transformation 打包在一起。

Enterprise Data Preparation

本节介绍版本 10.2.2 HotFix 1 中对 Enterprise Data Preparation 进行的更改。

数据资产搜索

从版本 10.2.2 HotFix 1 开始，数据资产搜索功能和用户界面进行了更新，以反映 Enterprise Data Catalog 中所做的更改。

有关详细信息，请参阅《*Informatica 10.2.2 HotFix 1 Enterprise Data Preparation 用户指南*》。

Enterprise Data Catalog

本节介绍版本 10.2.2 HotFix 1 中对 Enterprise Data Catalog 进行的更改。

业务术语传播

从版本 10.2.2 HotFix 1 开始，现在您可以配置资源的**业务术语关联**设置，以自动将职位与资产关联。之前，只能通过配置**运行相似性配置文件**选项自动将职位与资产关联。您还可以自动将职位与表和文件资产相关联。之前，只能将职位自动关联到列和字段资产。

有关业务术语传播的详细信息，请参阅《*Informatica 10.2.2 HotFix 1 Catalog 管理员指南*》中的“Enterprise Data Catalog 概念”一章。

自定义资源

从版本 10.2.2 HotFix 1 开始，自定义资源包括以下增强功能：

查看 ETL 转换的详细或摘要沿袭

您可以配置自定义 ETL 资源，以查看多个数据源之间 ETL 转换的详细和摘要沿袭。

指定元数据 ZIP 文件的远程文件路径

配置自定义资源时，可以提供 ZIP 文件（包括要获取到目录中的自定义元数据）的远程路径。可以使用远程文件路径来自动执行和计划自定义资源扫描。

配置自定义资源以使用用户定义的脚本提取元数据

如果您已使用脚本或命令序列使准备自定义元数据和生成 ZIP 文件的过程自动化，则可以在配置资源时提供详细信息。Enterprise Data Catalog 会在运行资源之前运行脚本。

为自定义资源数据资产配置自定义图标

可以针对自定义模型中定义的类自定义和配置图标。Enterprise Data Catalog 中会显示这些图标来表示来自自定义资源的数据资产。

为自定义资源配置关系视图

可以在 Enterprise Data Catalog 中为自定义资源配置**关系视图**页面。在配置过程中，可以为自定义模型中的类定义一组配置。根据这些定义，可以为每个类类型筛选或分组相关对象，并在**关系视图**页面中查看这些对象。

有关详细信息，请参阅《*Informatica 10.2.2 HotFix 1 Catalog 管理员指南*》。

文档更改

从版本 10.2.2 HotFix 1 开始，所有与创建和配置资源有关的信息都从《Catalog 管理员指南》移到了名为《*Informatica 10.2.2 HotFix 1 Enterprise Data Catalog 扫描程序配置指南*》的新指南中。

有关详细信息，请参阅《*Informatica 10.2.2 HotFix 1 Enterprise Data Catalog 扫描程序配置指南*》。

Microsoft SQL Server 资源

从版本 10.2.2 HotFix 1 开始，Microsoft SQL Server 资源支持使用 Enterprise Data Catalog 代理进行基于 Windows 的身份验证。

有关详细信息，请参阅《*Informatica 10.2.2 HotFix 1 扫描程序配置指南*》。

Informatica PowerExchange 适配器

本节介绍版本 10.2.2 HotFix 1 Service Pack 1 中对 Informatica 适配器进行的更改。

PowerExchange for Amazon Redshift

从版本 10.2.2 HotFix 1 Service Pack 1 开始，无需在 Amazon S3 桶策略中添加 GetBucketPolicy 权限即可连接至 Amazon Redshift。

在以前的版本中，必须在 Amazon S3 桶策略中添加 GetBucketPolicy 权限才能连接至 Amazon Redshift。

有关详细信息，请参阅《*Informatica PowerExchange for Amazon Redshift 10.2.2 HotFix 1 用户指南*》。

第 14 章

通知、新增功能和更改 (10.2.2 Service Pack 1)

本章包括以下主题：

- [通知 \(10.2.2 Service Pack 1\), 121](#)
- [新增功能 \(10.2.2 Service Pack 1\), 122](#)
- [更改 \(10.2.2 Service Pack 1\), 125](#)

通知 (10.2.2 Service Pack 1)

支持更改

本节介绍版本 10.2.2 Service Pack 1 中的支持更改。

延迟的支持

从版本 10.2.2 Service Pack 1 开始，Informatica 延迟了对以下功能的支持：

数据屏蔽转换的流传输

由于对流支持的更改，流映射中的数据屏蔽转换已延迟。

Informatica 打算在即将推出的版本中恢复此项支持，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。

技术预览版支持

从版本 10.2.2 Service Pack 1 开始，Informatica 包含以下技术预览功能：

通过启用 SSL 的 Kafka 连接进行流传输

创建 Kafka 连接时，可以使用 SSL 连接属性来配置 Kafka Broker。

支持 Hortonworks HDP 3.1

可以在 Hortonworks HDP 版本 3.1 上使用 Informatica 大数据产品。

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。

产品和服务名称更改

本节介绍版本 10.2.2 Service Pack 1 中对产品和服务名称的更改。

以下产品和服务名称已更改：

- Enterprise Data Lake 产品已改名为 Enterprise Data Preparation。
- Enterprise Data Lake 服务改名为 Enterprise Data Preparation 服务。
- 数据准备服务改名为交互数据准备服务。

发布任务

本节介绍版本 10.2.2 Service Pack 1 中的发布任务。发布任务是在升级到版本 10.2.2 Service Pack 1 后必须执行的任务。

Sqoop 连接

从版本 10.2.2 Service Pack 1 开始，以下发布任务适用于 Sqoop：

- 使用由 Cloudera Connector Powered by Teradata 连接器在 Spark 或 Blaze 引擎和 Cloudera CDH 版本 6.1.x 上运行现有 Sqoop 映射时，必须下载 junit-4.11.jar 和 sqoop-connector-teradata-1.7c6.jar 文件。在 Cloudera CDH 版本 6.1.x 上运行现有 Sqoop 映射前，请执行以下任务：
 1. 从以下 URL 下载并复制 junit-4.11.jar 文件：
<http://central.maven.org/maven2/junit/junit/4.11/junit-4.11.jar>
 2. 在数据集成服务运行的节点上，将 junit-4.11.jar 文件添加到以下目标：<Informatica 安装目录>\externaljdbcjars
 3. 从 Cloudera 网站下载并提取 Cloudera Connector Powered by Teradata 软件包，然后复制以下文件：sqoop-connector-teradata-1.7c6.jar
 4. 在数据集成服务运行的节点上，将 sqoop-connector-teradata-1.7c6.jar 文件添加到以下目标：
<Informatica 安装目录>\externaljdbcjars
 - 要在 Blaze 或 Spark 引擎和 Cloudera CDH 上运行 Sqoop 映射，不必再为 MapReduce 应用程序在 mapred-site.xml 文件中设置 mapreduce.application.classpath 条目。如果使用 Cloudera CDH 版本 6.1.x 运行现有 Sqoop 映射，请从 mapred-site.xml 文件中删除 mapreduce.application.classpath 条目。
- 有关详细信息，请参阅《*Informatica Big Data Management 10.2.2 Service Pack 1 集成指南*》。

新增功能 (10.2.2 Service Pack 1)

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，可以在联接器转换中使用 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联接器转换。

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*》。

更改 (10.2.2 Service Pack 1)

Big Data Management

本节介绍版本 10.2.2 Service Pack 1 中对 Big Data Management 进行的更改。

数据集成服务队列

从版本 10.2.2 Service Pack 1 开始，数据集成服务默认使用分布式队列来存储作业信息。

只能使用一个自定义属性来启用本地队列。如果需要此功能，请联系 Informatica 全球支持部门。

在以前的版本中，数据集成服务默认在每个节点上使用本地队列，只对启用了大数据恢复时的 Spark 作业使用分布式队列。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.2 Service Pack 1 管理员指南*》中的“数据集成服务处理”一章。

Mass Ingestion

从版本 10.2.2 Service Pack 1 开始，可以选择群集默认作为将数据获取到 Hive 目标的 Mass Ingestion 规范的存储格式。如果选择群集默认，规范会使用 Hadoop 群集的默认存储格式。

在以前的版本中，规范使用选择文本存储格式时群集上的默认存储格式。在 10.2.2 Service Pack 1 中，选择文本存储格式会将数据获取到一个标准文本文件。

大数据 Streaming

本节介绍版本 10.2.2 Service Pack 1 中对 Big Data Streaming 进行的更改。

转换

本节介绍版本 10.2.2 Service Pack 1 中对 Big Data Streaming 中的转换进行的更改。

等级转换

从版本 10.2.2 Service Pack 1 开始，如果流映射包含连接器转换，则其必须满足以下额外条件：

- 流映射不能在管道中包含等级转换和配置了不相等查找条件的被动查找转换。在以前的版本中，可以在同一管道中使用等级转换和配置了不相等查找条件的被动查找转换。
- 流映射中的等级转换不能有下游连接器转换。在以前的版本中，可以在流映射中的连接器转换之前的任意位置使用等级转换。
- 流映射不能在管道中包含多个等级转换。在以前的版本中，可以在流映射中使用多个等级转换。
- 流映射不能在管道中包含汇总器转换和等级转换。在以前的版本中，可以在同一管道中使用汇总器转换和等级转换。

排序器转换

从版本 10.2.2 Service Pack 1 开始，如果流映射包含排序器转换，则必须满足以下额外要求：

- 如果流映射包含排序器转换，则以完整输出模式运行。在以前的版本中，如果流映射包含排序器转换，则以追加输出模式运行。
- 流映射中的排序器转换必须有上游汇总器转换。在以前的版本中，无需上游汇总器转换，即可使用排序器转换。

- 如果映射包含排序器转换，则将忽略汇总器转换上游的窗口转换。在以前的版本中，如果映射包含排序器转换，不会忽略汇总器转换上游的窗口转换。

Informatica Analyst

本节介绍版本 10.2.2 Service Pack 1 中对 Analyst 工具的更改。

默认视图

从版本 10.2.2 Service Pack 1 开始，平面文件和表对象的默认视图为“属性”选项卡。创建或打开平面文件或表数据对象时，会在“属性”选项卡中打开该对象。在以前的版本中，默认视图为“数据查看器”选项卡。

有关详细信息，请参阅《*Informatica 10.2.2 Service Pack 1 Analyst 工具指南*》。

Informatica PowerExchange 适配器

本节介绍版本 10.2.2 Service Pack 1 中对 Informatica 适配器进行的更改。

PowerExchange for Amazon Redshift

从版本 10.2.2 Service Pack 1 开始，PowerExchange for Amazon Redshift 做出了以下更改：

- PowerExchange for Amazon Redshift 在以下发行版上支持**服务器端 KMS 加密**：
 - Amazon EMR 版本 5.20
 - Cloudera CDH 版本 5.16 和 6.1

在以前的版本中，数据集成服务在以下发行版上支持**服务器端 KMS 加密**：

- Amazon EMR 版本 5.16
- Cloudera CDH 版本 5.15
- 不能使用以下发行版运行 Amazon Redshift 映射：
 - MapR 版本 5.2
 - IBM BigInsight

在以前的版本中，可以使用 MapR 版本 5.2 和 IBM BigInsight 发行版运行 Amazon Redshift 映射。

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

PowerExchange for Amazon S3

从版本 10.2.2 Service Pack 1 开始，PowerExchange for Amazon S3 做出了以下更改：

- PowerExchange for Amazon S3 在以下发行版上支持**服务器端 KMS 加密**：
 - Amazon EMR 版本 5.20
 - Cloudera CDH 版本 5.16 和 6.1

在以前的版本中，PowerExchange for Amazon S3 在以下发行版上支持**服务器端 KMS 加密**：

- Amazon EMR 版本 5.16
- Cloudera CDH 版本 5.15。
- 不能使用以下发行版运行 Amazon S3 映射：
 - MapR 版本 5.2

- IBM BigInsight

在以前的版本中，可以使用 MapR 版本 5.2 和 IBM BigInsight 发行版运行 Amazon S3 映射。

有关详细信息，请参阅 《*Informatica PowerExchange for Amazon S3 10.2.2 Service Pack 1 用户指南*》。

第 15 章

通知 (10.2.2)

本章包括以下主题：

- [安装和配置, 128](#)
- [支持更改, 129](#)
- [发布任务, 131](#)

安装和配置

本节介绍版本 10.2.2 中与大数据安装和配置相关的通知。

OpenJDK

从版本 10.2.2 开始，Informatica 安装程序会封装 OpenJDK (AzulJDK)。支持的 Java 版本为 Azul OpenJDK 1.8.192。

可以使用 OpenJDK 在嵌入式群集上部署 Enterprise Data Catalog。要在现有群集上部署 Enterprise Data Catalog，必须在所有群集节点上安装 JDK 1.8。

Informatica 不再支持数据集成服务属性的“JDK 主目录”执行选项。Spark 引擎上的 Sqoop 映射使用与 Informatica 安装程序一起封装的 Java Development Kit (JDK)。

在以前的版本中，安装程序使用的是与自身一起封装的 Oracle Java。您还必须安装 JDK 并在数据集成服务计算机中指定 JDK 安装目录，这样才能运行在 Spark 引擎上使用 Sqoop 连接的 Sqoop 映射或 Mass Ingestion 规范或者在 Spark 引擎上处理 Java 转换。

安装程序代码签名

从版本 10.2.2 开始，Informatica 使用基于证书的数字签名对 Informatica 软件代码进行签名。代码签名有助于验证代码，并保护其在安装程序代码得到数字签名之后不会受到任何形式的篡改。

Informatica 将文件的公钥、签名和哈希封装在安装程序包中。Informatica 对软件包进行签名后，您可以联系 Informatica 全球客户支持获得公钥。

有关安装程序代码签名过程或客户如何验证已签名代码是否可信的详细信息，请参阅《*Informatica Big Data Suite 10.2.2 安装和配置指南*》。

恢复安装程序

从版本 10.2.2 开始，可以从故障点或退出点恢复安装过程。如果服务失败或在服务创建期间安装过程失败，可以使用服务器安装程序来恢复安装过程。如果是在创建服务后运行安装程序以配置服务，则无法恢复安装程序。当您加入域时，也无法恢复安装程序。

有关恢复安装程序的详细信息，请参阅《*Informatica Big Data Suite 10.2.2 安装和配置指南*》。

Informatica Docker 实用程序

从版本 10.2.2 开始，您可以使用 Informatica Docker 实用程序快速安装 Informatica 域。

运行 Informatica Docker 实用程序时，可以构建含基本操作系统和 Informatica 二进制文件的 Informatica Docker 映像。您可以运行现有 Docker 映像以配置 Informatica 域。运行 Informatica Docker 映像时，可以创建域，也可以加入域。在容器创建过程中，您可以创建模型存储库服务、数据集成服务和群集配置。

有关 Informatica Docker 实用程序的详细信息，请参阅 Informatica Network 上的文章 [How to Install Big Data Management 10.2.2 Using a Docker Utility](#)。

安装程序

本节介绍版本 10.2.2 中对 Informatica 安装程序的更改。

配置 Enterprise Data Catalog

如果在安装了 Informatica 服务的节点上安装 Enterprise Data Catalog 二进制文件，则可使用安装程序在该节点上配置 Enterprise Data Catalog 服务。

有关详细信息，请参阅《*Informatica Big Data Suite 10.2.2 安装和配置指南*》。

配置 Enterprise Data Lake

如果在安装了 Informatica 服务的节点上安装 Enterprise Data Lake 二进制文件，则可使用安装程序在该节点上配置 Enterprise Data Lake 服务。

有关详细信息，请参阅《*Informatica Big Data Suite 10.2.2 安装和配置指南*》。

支持更改

本节介绍版本 10.2.2 中的支持更改。

Hive 引擎

从版本 10.2.2 开始，Informatica 不再支持在 Hadoop 环境下运行的作业的 Hive 执行模式。您无法在 Hive 引擎上配置或运行作业。

Informatica 在 Hadoop 环境下继续支持 Blaze 和 Spark 引擎，并增加了对 Databricks 环境下的 Databricks Spark 引擎的支持。

在升级之前，您需要更新所有配置为在 Hive 引擎上运行的映射和配置文件。

有关详细信息，请参阅《*Informatica 10.2.2 升级指南*》。

发行版支持

Informatica 大数据产品支持 Hadoop 和 Databricks 环境。在每个版本中，Informatica 会增加、延期和终止对 Hadoop 发行版版本的支持。Informatica 可能会在将来的版本中恢复对延期版本的支持。

Big Data Management 新增了对 Databricks 环境的支持，并支持 Databricks 发行版版本 5.1。

下表列出了 Informatica 10.2.2 大数据产品支持的 Hadoop 发行版版本：

产品	EMR	HDI	CDH	HDP	MapR
Big Data Management	5.16	3.6.x	5.15 5.16	2.6.x	6.0.x MEP 5.0
Big Data Streaming	5.16	3.6.x 注意: 仅 ADLS 非 Kerberos 支持 HDI。	5.15 5.16	2.6.x	延迟的支持
Enterprise Data Catalog	N/A	3.6.x 注意: 仅 WASB 非 Kerberos 支持 HDI。	5.15 5.16 注意: 只能在部署于 CDH 5.16 Hadoop 发行版之上的 Enterprise Data Catalog 中使用 OpenJDK 1.8.0。	2.6.x	N/A
Enterprise Data Preparation	5.16	3.6.x 注意: 仅 ADLS 和 WASB 非 Kerberos 支持 HDI。	5.15 5.16	2.6.x	6.0.x MEP 5.0

要查看支持的版本的最新列表，请参阅 Informatica 客户门户上的“产品可用性列表”，网址为：
<https://network.informatica.com/community/informatica-network/product-availability-matrices>。

Big Data Management 的支持更改

本节介绍版本 10.2.2 中对 Big Data Management 进行的支持更改。

Python 转换

从版本 10.2.2 开始，对 Python 转换中二进制端口的支持已延迟。将在未来版本中恢复支持。

Big Data Streaming 的支持更改

本节介绍版本 10.2.2 中对 Big Data Streaming 进行的更改。

Azure 事件中心

从版本 10.2.2 开始，升级的流映射变为无效。要在使用 Spark 结构化流的 Spark 引擎上运行映射，必须重新创建物理数据对象。在重新创建物理数据对象之后，以下属性不适用于 Azure 事件中心数据对象：

- 使用者属性

- 分区计数

有关详细信息，请参阅《*Informatica 10.2.2 升级指南*》。

数据对象类型

从版本 10.2.2 开始，对某些数据对象类型的支持已延迟。将在未来版本中恢复支持。

下表介绍了在版本 10.2.2 中延迟支持的数据对象类型：

对象类型	对象
源	JMS MapR 流
目标	MapR 流
转换	数据屏蔽 联接器（主外联接、详细外联接和完整外部联接类型） 等级 排序器

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

Enterprise Data Catalog 中的通用连接框架

从版本 10.2.2 开始，通用连接框架在 Enterprise Data Catalog 中已弃用。

有关详细信息，请参阅《*Informatica Enterprise Data Catalog 对通用连接框架 (UCF) 使用性的支持声明*》。

发布任务

本节介绍版本 10.2.2 中的发布任务。发布任务是指升级到版本 10.2.2 后必须执行的任务。

Big Data Management

本节介绍版本 10.2.2 中 Big Data Management 的发布任务。

Decimal 数据类型

如果升级到版本 10.2.2，则启用高精度模式并在 Spark 引擎上运行的映射必须针对 TO_DECIMAL 和 TO_DECIMAL38 函数使用精度参数。如果函数没有精度参数，映射将失败。

例如，如果升级前的映射使用高精度模式且包含表达式 TO_DECIMAL(3)，则您必须指定精度参数，之后才能在 Spark 引擎上运行已升级的映射。当表达式具有精度参数时，表达式可能为 TO_DECIMAL(3,2)。

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

Mass Ingestion

从版本 10.2.2 开始，可以使用 Mass Ingestion 工具通过增量加载来获取数据。

如果升级到版本 10.2.2，Mass Ingestion 规范会升级到禁用增量加载。请先完成以下任务，然后才能对现有规范运行增量加载：

1. 编辑规范。
2. 在**定义**页面上，选择**启用增量加载**。
3. 在**源**和**目标**页面上，配置增量加载选项。
4. 保存规范。
5. 将规范重新部署到数据集成服务。

注意：重新部署的 Mass Ingestion 规范在 Spark 引擎上运行。

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

Python 转换

如果升级到版本 10.2.2，Python 转换在 Big Data Management 中处理数据的效率将会提高。

要体验性能方面的改进，请在 Hadoop 连接中配置以下 Spark 高级属性：

`infaspark.pythontx.exec`

在 Data Engineering Integration 的 Spark 引擎上运行 Python 转换时需要。Python 可执行二进制文件在 Hadoop 群集中工作节点上的位置。

例如，设置为：

```
infaspark.pythontx.exec=/usr/bin/python3.4
```

如果使用数据集成服务计算机上的 Python 安装，请将值设置为数据集成服务计算机上 Informatica 安装目录中的 Python 可执行二进制文件。

例如，设置为：

```
infaspark.pythontx.exec=INFA_HOME/services/shared/spark/python/lib/python3.4
```

`infaspark.pythontx.executorEnv.PYTHONHOME`

在 Data Engineering Integration 和 Data Engineering Streaming 的 Spark 引擎上运行 Python 转换时需要。Hadoop 群集中工作节点上 Python 安装目录的位置。

例如，设置为：

```
infaspark.pythontx.executorEnv.PYTHONHOME=/usr
```

如果使用数据集成服务计算机上的 Python 安装，请使用数据集成服务计算机上的 Python 安装目录的位置。

例如，设置为：

```
infaspark.pythontx.executorEnv.PYTHONHOME=
INFA_HOME/services/shared/spark/python/
```

配置高级属性之后，Spark 引擎不会使用 Jep 在 Python 转换中运行 Python 代码。

有关安装 Python 的信息，请参阅《*Informatica Big Data Management 10.2.2 集成指南*》。

大数据 Streaming

本节介绍版本 10.2.2 中 Big Data Streaming 的发布任务。

Kafka 目标

从版本 10.2.2 开始，Kafka 目标中键标头端口的数据类型为二进制。在以前的版本中，键标头端口的数据类型为字符串。

升级后，若要运行现有流映射，您必须重新创建数据对象，并使用新创建的数据对象更新流映射。

有关重新创建数据对象的详细信息，请参阅《*Big Data Management 10.2.2 集成指南*》。

Kafka 连接属性

升级后，对于 Kafka 连接，将 Kafka Messaging Broker 版本配置为 0.10.1.x-2.0.0。

Informatica PowerExchange 适配器

本节介绍版本 10.2.2 中 Informatica 适配器的发布任务。

PowerExchange for HBase

从版本 10.2.2 开始，必须在 Spark 引擎上运行映射，才能在 HBase 资源中查找数据。

如果先前已将映射配置为在本地环境中运行才能在 HBase 资源中查找数据，则必须在升级到版本 10.2.2 后将执行引擎更新为 Spark。否则，映射将失败。

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

PowerExchange for Microsoft Azure SQL 数据仓库

从先前版本升级到版本 10.2.2 之后，在运行时，包含以下数据类型的现有映射在 Spark 引擎上将失败：

- Binary
- Varbinary
- Datetime2
- Datetimeoffset

要成功运行现有映射，必须将这些数据类型映射到 String 数据类型或重新导入对象。

- Binary -> String (n)
- Varbinary -> String (n)
- Datetime2 -> String (27)
- Datetimeoffset -> String (34)

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

第 16 章

新产品(10.2.2)

本章包括以下主题：

- [PowerExchange 适配器, 134](#)

PowerExchange 适配器

Informatica PowerExchange 适配器

本节介绍版本 10.2.2 中新增的 Informatica 适配器。

PowerExchange for Cassandra JDBC

从版本 10.2.2 开始，可以创建 Cassandra 连接以连接到 Cassandra 密钥空间，并在数据对象中包括 Cassandra 表。可以使用 Cassandra 连接从 Cassandra 源读取数据，并将数据写入 Cassandra 目标。可以在 Hadoop 环境中的 Spark 引擎上验证并运行映射。

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

PowerExchange for Google Cloud Spanner

从版本 10.2.2 开始，可以创建 Google Cloud Spanner 连接以连接到 Google Cloud Spanner 实例，并在数据对象中包括 Google Cloud Spanner 表。可以使用 Google Cloud Spanner 连接从 Google Cloud Spanner 源读取数据，并将数据写入 Google Cloud Spanner 目标。可以在本地环境中或在 Hadoop 环境中的 Spark 引擎上验证并运行 Google Cloud Spanner 映射。

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

PowerExchange for Tableau V3

从版本 10.2.2 开始，您可以使用 Tableau V3 连接从多个源读取数据，生成 Tableau .hyper 输出文件，以及向 Tableau 中写入数据。可以在本地环境中验证和运行映射。

有关详细信息，请参阅《*Informatica PowerExchange for Tableau V3 用户指南*》。

第 17 章

新增功能 (10.2.2)

本章包括以下主题：

- [应用程序服务, 135](#)
- [Big Data Management, 136](#)
- [大数据 Streaming, 140](#)
- [命令行程序, 141](#)
- [Enterprise Data Catalog, 145](#)
- [Enterprise Data Lake, 147](#)
- [Informatica Developer, 152](#)
- [Informatica 映射, 152](#)
- [Informatica 转换, 153](#)
- [Informatica PowerExchange 适配器, 156](#)

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*》

Spark 引擎上的数据预览

从版本 10.2.2 开始，您可以在 Developer tool 中预览在 Spark 引擎上运行的映射中的数据。预览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

数据预览可用于技术预览版。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。

有关详细信息，请参阅《*Informatica® 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。

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 144](#)。

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 ([a href="#">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

- STDEV
- STDEVIF
- 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*》。

第 18 章

更改 (10.2.2)

本章包括以下主题：

- [应用程序服务, 162](#)
- [Big Data Management, 162](#)
- [大数据 Streaming, 164](#)
- [Enterprise Data Catalog, 165](#)
- [Enterprise Data Lake, 165](#)
- [Informatica Developer, 166](#)
- [Informatica 转换, 166](#)
- [Informatica PowerExchange 适配器, 167](#)

应用程序服务

本节介绍版本 10.2.2 中对应用程序服务功能进行的更改。

元数据访问服务

从版本 10.2.2 开始，可以使用同一元数据访问服务从多个 Hadoop 发行版导入和预览元数据。在以前的版本中，必须为多个 Hadoop 发行版创建多个元数据访问服务实例。

Big Data Management

本节介绍版本 10.2.2 中对 Big Data Management 进行的更改。

Hive 连接

从版本 10.2.2 开始，以下 Hive 连接属性已重命名：

- “遵守细粒度 SQL 授权”属性重命名为“细粒度授权”。
- “用户名”属性重命名为“LDAP 用户名”。

下表介绍了属性：

属性	说明
细粒度授权	<p>选择在 Hive 源中遵守细粒度授权的选项时，映射遵循以下内容：</p> <ul style="list-style-type: none">- 行和列级别限制。应用于启用了 Sentry 或 Ranger 安全模式的 Hadoop 群集。- 数据屏蔽规则。应用于 Dynamic Data Masking 对包含敏感数据的列所设置的屏蔽规则。 <p>如果没有选择此选项，Blaze 和 Spark 引擎会忽略限制和屏蔽规则，并且结果将包括限制数据或敏感数据。</p>
LDAP 用户名	<p>数据集成服务用于模拟在 Hadoop 群集中运行映射的用户的 LDAP 用户名。该用户名取决于在元数据连接字符串或数据访问连接字符串中为本机环境指定的 JDBC 连接字符串。</p> <p>如果 Hadoop 群集使用 Kerberos 身份验证，则 JDBC 连接字符串的主体名称必须与用户名相同。否则，用户名将取决于 JDBC 驱动程序的行为。使用 Hive JDBC 驱动程序时，您可以通过多种方法指定用户名，而且用户名可以成为 JDBC URL 的一部分。</p> <p>如果 Hadoop 群集不使用 Kerberos 身份验证，则用户名将取决于 JDBC 驱动程序的行为。</p> <p>如果不指定用户名，则 Hadoop 群集会根据以下条件对作业进行身份验证：</p> <ul style="list-style-type: none">- Hadoop 群集不使用 Kerberos 身份验证。它根据运行数据集成服务的计算机的操作系统配置文件用户名对作业进行身份验证。- Hadoop 群集 Kerberos 身份验证。它根据数据集成服务的 SPN 对作业进行身份验证。LDAP 用户名将被忽略。

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

Mass Ingestion

从版本 10.2.2 开始，已部署的 Mass Ingestion 规范在 Spark 引擎上运行。在版本 10.2.2 之前部署的已升级 Mass Ingestion 规范将继续在 Blaze 和 Spark 引擎上运行，直到其重新得到部署。

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

Spark 监视

从版本 10.2.2 开始，Spark 监视默认处于启用状态。

以前，Spark 监视默认处于禁用状态。

有关 Spark 监视的详细信息，请参阅《*Informatica Big Data Management 10.2.2 用户指南*》。

Sqoop

从版本 10.2.2 开始，对 Sqoop 应用了以下更改：

- 可以在 Hadoop 连接的 Spark 暂存目录中指定文件路径，以存储 Sqoop 作业的临时文件。Spark 引擎运行 Sqoop 作业时，数据集成服务会在 Spark 暂存目录内创建 Sqoop 暂存目录以存储临时文件：<Spark 暂存目录>/sqoop_staging
以前，Sqoop 暂存目录为硬编码，数据集成服务使用以下暂存目录：/tmp/sqoop_staging
有关详细信息，请参阅《*Informatica Big Data Management 10.2.2 用户指南*》。
- Spark 引擎上的 Sqoop 映射使用与 Informatica 安装程序打包在一起的 OpenJDK (AzulJDK)。不再需要为数据集成服务指定 **JDK 主目录** 属性。
以前，要在 Spark 引擎上运行 Sqoop 映射，必须在运行数据集成服务的计算机上安装 Java Development Kit (JDK)。然后，在 Informatica Administrator 的数据集成服务执行选项下的 **JDK 主目录** 属性中指定 JDK 安装目录的位置。

Hadoop 环境中的转换

本节介绍版本 10.2.2 中对 Hadoop 环境中的转换进行的更改。

Python 转换

从版本 10.2.2 开始，Python 转换可以在 Spark 引擎上更高效地处理数据（与版本 10.2.1 中的 Python 转换相比）。此外，Python 转换无需安装 Jep，您可以使用任何版本的 Python 来运行转换。

以前，Python 转换仅支持与 Jep 兼容的特定 Python 版本。

注意：这些改进功能仅可用于 Big Data Management。

有关安装 Python 的信息，请参阅《*Informatica Big Data Management 10.2.2 集成指南*》。

有关 Python 转换的详细信息，请参阅《*Informatica 10.2.2 Developer 转换指南*》中的“Python 转换”一章。

写入转换

从版本 10.2.2 开始，关系、Netezza 和 Teradata 数据对象的写入转换中的**创建或替换目标表**高级属性已重命名为**目标架构策略**。

配置写入转换时，您可以从目标数据对象的以下目标架构策略选项中进行选择：

- RETAIN - 保留现有目标架构。数据集成服务保留现有目标架构。
- CREATE - 在运行时创建或替换表。数据集成服务在运行时丢弃目标表，然后将其替换为基于您标识的目标数据对象的表。
- 分配参数。指定**目标架构策略**选项作为参数值。

以前，当选中**创建或替换目标表**高级属性时，数据集成服务会在运行时丢弃目标表，然后将其替换为基于您标识的目标表的表。未选中**创建或替换目标表**高级属性时，数据集成服务则保留目标表的现有架构。

在启用了**创建或替换目标表**属性的现有映射中，升级到版本 10.2.2 后，默认情况下，**目标架构策略**属性会显示 **CREATE - 在运行时创建或替换表**选项处于启用状态。在未选中**创建或替换目标表**选项的映射中，升级后，会为**目标架构策略**属性启用 **RETAIN - 保留现有目标架构**选项。在升级之后，如果选中的目标架构策略选项不正确，则必须手动从**目标架构策略**列表中选择所需的选项，然后运行映射。

有关配置目标架构策略的详细信息，请参阅《*Informatica 转换指南*》中的“写入转换”一章或《*Informatica Developer 映射指南*》中的“动态映射”一章。

大数据 Streaming

本节介绍版本 10.2.2 中对 Big Data Streaming 进行的更改。

Big Data Streaming 与 Big Data Management 集成

从版本 10.2.2 开始，要在运行时管理目标文件系统中的目标文件，系统会在指定的目标文件目录中创建临时目录。

临时目录将当前写入数据的目标文件和达到滚存限制后关闭的目标文件分离开来。

以前，所有目标文件都存储在目标文件目录中。

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

Kafka 连接

从版本 10.2.2 开始，Kafka Broker 维护 Kafka Messaging Broker 的配置信息。在以前的版本中，Apache ZooKeeper 维护 Kafka Messaging Broker 的配置信息。

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

转换

本节介绍版本 10.2.2 中对 Big Data Streaming 中的转换进行的更改。

汇总器转换

从版本 10.2.2 开始，如果流映射包含汇总器转换，则其必须满足以下额外要求：

- 流映射具有的窗口转换必须恰好是汇总器转换的相邻上游。以前，您可以在窗口转换之后的管道中任意位置使用汇总器转换。
- 流映射可以具有单个汇总器转换。以前，您会在流映射中使用多个汇总器转换。

联接器转换

从版本 10.2.2 开始，如果流映射包含联接器转换，则其必须满足以下额外条件：

- 流映射具有的窗口转换必须是联接器转换的相邻上游。在以前的版本中，您可以在窗口转换之后的管道中任意位置使用联接器转换。
- 流映射可以具有单个联接器转换。在以前的版本中，可以在流映射中使用多个联接器转换。
- 在流映射中的联接器转换之前的任意位置，流映射都不能包含汇总器转换。在以前的版本中，您可以在流映射中的联接器转换之前的任意位置使用汇总器转换。

Enterprise Data Catalog

本节介绍版本 10.2.2 中对 Enterprise Data Catalog 进行的更改。

Java Development Kit 更改

从版本 10.2.2 开始，Informatica 安装程序会封装 OpenJDK (Azul JDK)。在以前的版本中，安装程序封装的是 Oracle JDK。可以使用 OpenJDK 在嵌入式群集上部署 Enterprise Data Catalog。

要在现有群集上部署 Enterprise Data Catalog，必须在所有群集节点上安装 JDK 1.8。

Enterprise Data Lake

本节介绍版本 10.2.2 中对 Enterprise Data Lake 进行的更改。

MAX 和 MIN 函数

从版本 10.2.2 开始，MAX 和 MIN 汇总函数的行为有变化。要在工作表中保持版本 10.2.1 或更早版本中提供的 MAX 和 MIN 函数的行为，请在工作表脚本中将这两个函数重命名为 MAXINLIST 和 MININLIST。

下表介绍了版本 10.2.2 中提供的函数：

功能	说明
MAX (value)	根据指定表达式中包含的列，返回工作表中所有行的最大值。
MIN (value)	根据指定表达式中包含的列，返回工作表中所有行的最小值。
MAXINLIST (value, [value],...)	返回指定表达式列表中的最大数字或最近日期。
MININLIST (value, [value],...)	返回指定表达式列表中的最小数字或最早日期。

Informatica Developer

本节介绍版本 10.2.2 中对 Informatica Developer 的更改。

Informatica Developer 名称更改

在版本 10.2.2 中，Informatica Developer 现称为 Big Data Developer。

在大数据版本中，该工具的名称改为 Big Data Developer。大数据版本包括 Big Data Management 和 Big Data Quality 之类的产品。

在传统版本中，该工具的名称仍为 Informatica Developer。传统版本包括 PowerCenter 和 Data Quality 之类的产品。

Informatica 转换

本节介绍版本 10.2.2 中对 Informatica 转换的更改。

地址验证器转换

本节介绍地址验证器转换的更改。

地址验证器转换对地址功能做了以下更新：

所有国家/地区

从版本 10.2.2 开始，地址验证器转换纳入了 Informatica Address Verification 软件引擎版本 5.14.0 中的功能。

以前，该转换使用 5.12.0 版本的 Informatica Address Verification 软件引擎。

日本

从版本 10.2.2 开始，Informatica 根据客户反馈改进了日本地址的解析和验证。

例如，在版本 10.2.2 中，当日本地址中缺少邮政编码或邮政编码与区域信息不匹配时，Informatica 会拒绝该地址。

以前，Informatica 会尝试更正地址。

西班牙

从版本 10.2.2 开始，Informatica 改进了西班牙地址的解析和验证。

例如，在版本 10.2.2 中，当西班牙地址的街道信息需要多次更正才能与参考数据相匹配时，地址验证器转换会拒绝该地址。

以前，该转换会更正街道数据多次，这可能导致对输入地址准确性的评估过高。

类似地，在版本 10.2.2 中，如果地址与参考数据中的多个候选项匹配，地址验证器转换则会以批处理模式返回地址的 I3 结果。

以前，该转换可能会尝试更正输入地址。

Parcel Centroid 和 Rooftop 地理编码

自 2018 年 10 月起，Informatica 不再支持包含 Parcel Centroid 和 Rooftop 地理编码数据的参考数据文件。

有关 Informatica Address Verification 软件引擎更新的详细信息，请参阅《*Informatica Address Verification 5.14.0 发行指南*》。

写入转换

从版本 10.2.2 开始，关系、Netezza 和 Teradata 数据对象的写入转换中的**创建或替换目标表**高级属性已重命名为**目标架构策略**。

配置写入转换时，您可以从目标数据对象的以下目标架构策略选项中进行选择：

- RETAIN - 保留现有目标架构。数据集成服务保留现有目标架构。
- CREATE - 在运行时创建或替换表。数据集成服务在运行时丢弃目标表，然后将其替换为基于您标识的目标数据对象的表。
- 分配参数。指定**目标架构策略**选项作为参数值。

以前，当选中**创建或替换目标表**高级属性时，数据集成服务会在运行时丢弃目标表，然后将其替换为基于您标识的目标表的表。未选中**创建或替换目标表**高级属性时，数据集成服务则保留目标表的现有架构。

在启用了**创建或替换目标表**属性的现有映射中，升级到版本 10.2.2 后，默认情况下，**目标架构策略**属性会显示 **CREATE - 在运行时创建或替换表**选项处于启用状态。在未选中**创建或替换目标表**选项的映射中，升级后，会为**目标架构策略**属性启用 **RETAIN - 保留现有目标架构**选项。在升级之后，如果选中的目标架构策略选项不正确，则必须手动从**目标架构策略**列表中选择所需的选项，然后运行映射。

有关配置目标架构策略的详细信息，请参阅《*Informatica 转换指南*》中的“写入转换”一章或《*Informatica Developer 映射指南*》中的“动态映射”一章。

Informatica PowerExchange 适配器

本节介绍版本 10.2.2 中对 Informatica 适配器的更改。

PowerExchange for Amazon Redshift

从版本 10.2.2 开始，PowerExchange for Amazon Redshift 具有以下更改：

- Amazon Redshift 数据对象写入操作的以下高级属性名称已更改：

旧属性名称	新属性名称
CHAR 和 VARCHAR 数据类型为空值	Char 和 Varchar 需要空值
S3 上文件一致性的等待时间(秒)	S3 文件一致性的等待时间(秒)
开启 S3 服务器端加密	S3 服务器端加密
开启 S3 客户端加密	S3 客户端加密
S3 上暂存文件的保留前缀	为 S3 上的暂存文件保留的前缀

- 以下卸载命令的默认值已更改：

卸载命令	旧默认值	新默认值
DELIMITER	竖线 ()	\036

- 以下复制命令的默认值已更改：

复制命令	旧默认值	新默认值
DELIMITER	竖线 ()	\036
QUOTE	双引号 (")	\037

- 在 Developer Tool 中导入 Amazon Redshift 表时，无法将表中可为空的列添加为主键。
在以前的版本中，可以在 Developer Tool 中将表中可为空的列添加为主键。

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 具有以下更改：

- 分多个部分下载 S3 文件高级源会话属性的名称已更改为多下载阈值。
- 无需在 Amazon S3 桶策略中添加 GetBucketAcl 权限即可连接至 Amazon S3。
在以前的版本中，必须在 Amazon S3 桶策略中添加 GetBucketAcl 权限才能连接至 Amazon S3。

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

PowerExchange for Google Analytics

从版本 10.2.2 开始，PowerExchange for Google Analytics 随 Informatica 服务一起安装。

以前，PowerExchange for Google Analytics 有一个单独的安装程序。

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

PowerExchange for Google Cloud Storage

从版本 10.2.2 开始，PowerExchange for Google Cloud Storage 随 Informatica 服务一起安装。

以前，PowerExchange for Google Cloud Storage 有一个单独的安装程序。

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

PowerExchange for HBase

从版本 10.2.2 开始，必须在 Spark 引擎上运行映射，才能在 HBase 资源中查找数据。

以前，在本地环境中或在 Spark 引擎上运行映射均可在 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，在 Blaze 引擎上运行映射将数据写入 HDFS 平面文件目标时，数据集成服务会按以下命名约定创建目标文件：

`<FileName>-P1, <FileName>-P2, ..., <FileName>-P100, ..., <FileName>-PN`

在以前的版本中，按以下格式生成目标文件：

`Target1.out, Target2.out, ..., Target<PartitionNo>.out`

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，Hive 连接中的**用户名**属性重命名为 **LDAP 用户名**。

PowerExchange for Microsoft Azure Blob 存储

从版本 10.2.2 开始，当您从 Microsoft Azure Blob 存储导入对象以及从子目录选择对象时，可以在容器结构中导航。以前，只能选择存在于容器中的对象。

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

第 IV 部分： 版本 10.2.1

本部分包含以下章节：

- [新增功能 \(10.2.1\), 171](#)
- [更改 \(10.2.1\), 201](#)
- [发布任务 \(10.2.1\), 218](#)

第 19 章

新增功能 (10.2.1)

本章包括以下主题：

- [应用程序服务, 171](#)
- [Big Data Management, 173](#)
- [Big DataStreaming, 180](#)
- [命令行程序, 181](#)
- [Enterprise Data Catalog, 185](#)
- [Enterprise Data Lake, 188](#)
- [Informatica Developer, 189](#)
- [Informatica 映射, 191](#)
- [Informatica 转换语言, 194](#)
- [Informatica 转换, 194](#)
- [Informatica 工作流, 196](#)
- [Informatica PowerExchange 适配器, 197](#)
- [安全, 200](#)

应用程序服务

本节介绍版本 10.2.1 中新增的系统服务功能。

内容管理服务

从版本 10.2.1 开始，您可以选择性地指定一个架构，以将引用数据数据库中的引用表标识为内容管理服务上的属性。

要指定架构，请在 Informatica Administrator 中使用内容管理服务上的"引用数据位置架构"属性。或者使用 DataServiceOptions.RefDataLocationSchema 选项运行 infacmd cms updateServiceOptions 命令。

如果没有为内容管理服务上的引用表指定架构，该服务将使用数据库连接指定的架构。如果未在数据库连接中显式设置架构，内容管理服务将使用默认数据库架构。

注意：创建受管引用表之前，请先建立内容管理服务将用于引用数据的数据库和架构。

有关详细信息，请参阅《*Informatica 10.2.1 应用程序服务指南*》中的"内容管理服务"一章和《*Informatica 10.2.1 命令参考*》中的"infacmd cms 命令参考"一章。

数据集成服务

从版本 10.2.1 开始，数据集成服务属性包括一个新的执行选项。

JDK 主目录

运行数据集成服务的计算机上的 JDK 安装目录。需要此目录才能在 Spark 引擎上运行使用 Sqoop 连接的 Sqoop 映射或 Mass Ingestion 规范或者在 Spark 引擎上处理 Java 转换。默认值为空。

Mass Ingestion 服务

从版本 10.2.1 开始，可以创建 Mass Ingestion 服务。Mass Ingestion 服务是 Informatica 域中的一个应用程序服务，用于管理 Mass Ingestion 规范。可以在 Mass Ingestion 工具中配置 Mass Ingestion 规范，以从关系源将大量数据引入到 Hive 或 HDFS 目标。

为管理 Mass Ingestion 规范，Mass Ingestion 服务将执行以下任务：

- 管理并验证 Mass Ingestion 规范。
- 将 Mass Ingestion 作业计划为在数据集成服务上运行。
- 监视 Mass Ingestion 作业的结果和统计信息。
- 重新启动 Mass Ingestion 作业。

有关 Mass Ingestion 服务的详细信息，请参阅《*Informatica 10.2.1 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 10.2.1 开始，可以创建元数据访问服务。元数据访问服务是一个应用程序服务，使 Developer tool 可以访问 Hadoop 连接信息以导入和预览元数据。从 Hadoop 群集导入对象时，以下适配器在设计时使用元数据访问服务来提取对象元数据：

- PowerExchange for HBase
- PowerExchange for HDFS
- PowerExchange for Hive
- PowerExchange for MapR-DB

有关详细信息，请参阅《*Informatica 10.2.1 应用程序服务指南*》中的"元数据访问服务"一章。

模型存储库服务

使用 Azure SQL 数据库作为模型存储库

从版本 10.2.1 开始，可以使用 Azure SQL 数据库作为模型存储库。

有关详细信息，请参阅《*Informatica 10.2.1 应用程序服务指南*》中的"模型存储库服务"一章。

Git 版本控制系统

从版本 10.2.1 开始，可以将模型存储库与 Git 版本控制系统集成。Git 是一个分布式版本控制系统。签出和签入对象时，版本的副本会保存到本地存储库和 Git 服务器。如果 Git 服务器已关闭，则本地存储库会保留该对象的所有版本。要使用 Git 版本控制系统，请在 **URL** 字段中输入 Git 全局存储库的 URL，在**用户名**和**密码**字段中输入全局存储库的登录凭据，在 **VCS 本地存储库路径** 字段中输入模型存储库服务的本地存储库路径。

有关详细信息，请参阅《*Informatica 10.2.1 应用程序服务指南*》中的"模型存储库服务"一章。

Big Data Management

本节介绍版本 10.2.1 中新增的 Big Data Management 功能。

Blaze 引擎资源节约

从版本 10.2.1 开始，可以保留 Blaze 引擎基础结构使用的资源。

设置 `infagrid.blaze.service.idle.timeout` 属性可以指定 Blaze 引擎在释放资源之前保持空闲的分钟数。设置 `infagrid.orchestrator.svc.sunset.time` 属性可以指定 Blaze Orchestrator 服务持续运行的最大小时数。可以使用 `infacmd isp createConnection` 命令，或者也可以使用 Administrator 工具或 Developer tool 在 Hadoop 连接中设置 "Blaze 高级属性" 中的属性。

有关这些属性的详细信息，请参阅《*Informatica Big Data Management 10.2.1 管理员指南*》。

群集工作流

可以使用新的工作流任务来创建群集工作流。

群集工作流将在云平台上创建一个群集，并在该群集上运行映射及其他工作流任务。工作流任务完成后，您可以选择终止并删除该群集以节省群集资源。

借助两项新的工作流任务，您可以在群集工作流中创建和删除 Hadoop 群集：

创建群集任务

借助 "创建群集" 任务，您可以在以下云平台上创建、配置和启动 Hadoop 群集：

- Amazon Web Services (AWS)。可以创建 Amazon EMR 群集。
- Microsoft Azure。可以创建 HDInsight 群集。

删除群集任务

借助可选的 "删除群集" 任务，您可以在映射任务和工作流中的任何其他任务完成后删除群集。您可能需要执行此操作以节省成本。

在以前的版本中，您可以在工作流中使用命令任务在云平台上创建群集。有关群集工作流和工作流任务的详细信息，请参阅《*Informatica 10.2.1 Developer 工作流指南*》。

注意：在版本 10.2.1 中，使用命令任务来创建和删除群集的方法现在支持 AWS 上的 Cloudera Altus 群集。有关详细信息，请参阅 Informatica Network 上的文章 "How to Create Cloudera Altus Clusters with a Cluster Workflow on Big Data Management"（如何在 Big Data Management 中使用群集工作流创建 Cloudera Altus 群集）。

映射任务

映射任务高级属性包括新的 `ClusterIdentifier` 属性。`ClusterIdentifier` 标识要用于运行映射任务的群集。

有关群集工作流的详细信息，请参阅《*Informatica 10.2.1 Developer 工作流指南*》。

云置备配置

云置备配置是一个对象，其中包含有关连接到 Hadoop 群集的信息。

云置备配置包含有关如何将域和 Hadoop 帐户身份验证以及存储资源集成的信息。群集工作流使用云置备配置中的信息来连接到云平台（例如 Amazon Web Services 或 Microsoft Azure）并在其上创建群集。

有关云置备的详细信息，请参阅《*Informatica Big Data Management 10.2.1 管理员指南*》中的 "云置备配置" 一章。

高可用性

从版本 10.2.1 开始，可以在 Cloudera CDH、Hortonworks HDP 和 MapR Hadoop 发行版上为 Hadoop 环境中的以下服务和安全系统启用高可用性：

- Apache Ranger
- Apache Ranger KMS
- Apache Sentry
- Cloudera Navigator Encrypt
- HBase
- Hive Metastore
- HiveServer2
- NameNode
- 资源管理器

Hadoop 环境中的 Hive 功能

本节介绍版本 10.2.1 中为 Hadoop 环境新增的 Hive 功能。

Hive 表截断

从版本 10.2.1 开始，可以在所有运行时引擎上截断外部的已分区 Hive 表。

可以截断以下 Hive 存储格式的表：

- Avro
- ORC
- Parquet
- RCFile
- 序列
- 文本

可以截断以下 Hive 外部表格式的表：

- Hive on HDFS
- Hive on Amazon S3
- Hive on Azure Blob
- Hive on WASB
- Hive on ADLS

有关截断 Hive 目标的详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》中的"Hadoop 环境中的映射目标"一章。

映射前和映射后 SQL 命令

从版本 10.2.1 开始，可以针对 Spark 引擎上运行的映射中的 Hive 源和目标配置 PreSQL 和 PostSQL 命令。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》。

从 PowerCenter 导入

本节介绍版本 10.2.1 中新增的"从 PowerCenter 导入"功能。

从 PowerCenter 导入会话属性

从版本 10.2.1 开始，可以从 PowerCenter 存储库将会话属性（例如关系源和目标中基于 SQL 的替代和查找转换的替代）导入到模型存储库。

有关"从 PowerCenter 导入"功能的详细信息，请参阅《*Informatica 10.2.1 Developer 映射指南*》中的"从 PowerCenter 导入"一章。

SQL 参数

从版本 10.2.1 开始，可以指定 SQL 参数类型以将所有基于 SQL 的替代导入到模型存储库。其余会话替代属性将映射到 String 或相应的参数类型。

有关详细信息，请参阅《*Informatica 10.2.1 Developer 映射指南*》中的"从 PowerCenter 导入"一章。

从 PowerCenter 导入命令任务

从版本 10.2.1 开始，可以从 PowerCenter 将命令任务导入到模型存储库。

有关详细信息，请参阅《*Informatica 10.2.1 Developer 工作流指南*》中的"工作流"一章。

智能结构模型

从版本 10.2.1 开始，可以在 Big Data Management 中使用智能结构模型。

使用智能结构模型的数据对象的 Spark 引擎支持

可以在 Amazon S3、Microsoft Azure Blob 或复杂文件数据对象中纳入智能结构模型。向 Spark 引擎上运行的映射添加数据对象时，您可以处理该模型所能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。

复杂文件数据对象、Amazon S3 数据对象和 Microsoft Azure Blob 数据对象中的智能结构模型可用于技术预览版。支持技术预览版功能，但无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 10.2.1 用户指南*》。

Mass Ingestion

从版本 10.2.1 开始，可以执行 Mass Ingestion 作业来引入或复制大量数据以在数据库或存储库中使用，或者将数据存储在数据库或存储库中。要执行 Mass Ingestion 作业，可以使用 Mass Ingestion 工具创建 Mass Ingestion 规范。可以将 Mass Ingestion 规范配置为从关系数据库将输入引入到 Hive 或 HDFS 目标。还可以指定参数来清理引入的数据。

借助 Mass Ingestion 规范，您无需手动创建并运行映射。可以创建一个同时引入所有数据的 Mass Ingestion 规范。

有关 Mass Ingestion 的详细信息，请参阅《*Informatica Big Data Management 10.2.1 Mass Ingestion 指南*》。

监视

本节介绍版本 10.2.1 中与 Big Data Management 监视相关的新功能。

Hadoop 群集监视

从版本 10.2.1 开始，可以配置为 Hadoop 群集监视的应用程序日志中显示的信息量。

应用程序日志中的信息量取决于在 Developer tool 中为映射配置的跟踪级别。下表介绍了每个跟踪级别的应用程序日志中显示的信息量：

跟踪级别	消息
无	日志显示 FATAL 消息。FATAL 消息包括不可恢复的系统故障，这些故障导致服务关闭或变为不可用。
简洁	日志显示 FATAL 和 ERROR 代码消息。ERROR 消息包括连接故障、无法保存或检索元数据、服务错误。
普通	日志显示 FATAL 消息、ERROR 消息和 WARNING 消息。WARNING 错误包括可恢复系统故障或警告。
详细初始化	日志显示 FATAL 消息、ERROR 消息、WARNING 消息和 INFO 消息。INFO 消息包括系统和服务更改消息。
详细数据	日志显示 FATAL 消息、ERROR 消息、WARNING 消息、INFO 消息和 DEBUG 消息。DEBUG 消息是用户请求日志。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》中的"监视 Hadoop 环境中的映射"一章。

Spark 监视

从版本 10.2.1 开始，作为 Spark 监视支持的一部分，Spark 执行程序在某个端口上侦听 Spark 事件，您不需要配置 SparkMonitoringPort。

数据集成服务有一个可用端口范围，Spark 执行程序将从该可用范围中选择端口。出现故障时，端口连接保持可用，您不需要在运行映射之前重新启动数据集成服务。

系统会保留监视端口的自定义属性。如果您配置该属性，数据集成服务将使用指定的端口来侦听 Spark 事件。

在以前的版本中，数据集成服务自定义属性"Spark 监视端口"可用于配置 Spark 侦听端口。如果未配置该属性，则默认情况下会禁用 Spark 监视。

Tez 监视

从版本 10.2.1 开始，可以查看与 Tez 引擎监视支持相关的属性。可以使用 Hive 引擎在 MapReduce 或 Tez 上运行映射。Tez 引擎可以处理 Hortonworks HDP、Azure HDInsight 和 Amazon Elastic MapReduce 上的作业。要在 Tez 上运行 Spark 映射，可以使用 Tez 支持的任何群集。

在 Administrator 工具中，还可以在监视 Hive 引擎时检查 Tez 的 Hive 查询属性。可以在 Hive 会话日志以及 Tez 中查看与 Tez 统计信息相关的信息，例如 DAG 跟踪 URL、顶点总数和 DAG 进度。

可以监视 Tez 引擎上的任何 Hive 查询。为日志记录启用了详细数据或详细初始化时，可以在 Administrator 工具或会话日志中查看 Tez 引擎信息。还可以在 Administrator 工具的"监视"选项卡中监视 Tez 引擎上的映射的状态。

有关 Tez 监视的详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》和《*Informatica Big Data Management 10.2.1 Hadoop 集成指南*》。

在 Spark 引擎上处理层次结构数据

从版本 10.2.1 开始，Spark 引擎包含以下用于处理层次结构数据的附加功能：

map 数据类型

可以使用 map 数据类型来生成和处理复杂文件中的映射数据。

Amazon S3 上的复杂文件

可以使用复杂数据类型来读取 Amazon S3 上的 Avro 和 Parquet 文件中的层次结构数据以及向这些文件写入层次结构数据。可以在数据对象读取操作和写入操作中将列投影为复杂数据类型。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》中的“在 Spark 引擎上处理层次结构数据”一章。

Spark 引擎上的规则规范支持

从版本 10.2.1 开始，除了在 Blaze 和 Hive 引擎上之外，还可以在 Spark 引擎上运行包含规则规范的映射。

除了在 Blaze 和 Hive 引擎上之外，还可以在 Spark 引擎上运行包含从规则规范生成的 Mapplet 的映射。

有关规则规范的详细信息，请参阅《*Informatica 10.2.1 规则规范指南*》。

安全

本节介绍版本 10.2.1 中与 Big Data Management 安全相关的新功能。

Cloudera NavigatorEncrypt

从版本 10.2.1 开始，可以使用 Cloudera Navigator Encrypt 来保护部分数据的安全并对其余数据实施透明加密。

EMR 文件系统授权

从版本 10.2.1 开始，可以在 Spark 引擎上使用 EMR 文件系统 (EMR File System, EMRFS) 授权来访问 Amazon S3 中的数据。

IAM 角色

从版本 10.2.1 开始，可以使用 EMR 文件系统的 IAM 角色从 Amazon EMR 群集版本 5.10 的 Amazon S3 群集读取数据以及向其中写入数据。

Kerberos 身份验证

从版本 10.2.1 开始，可以为以下群集启用 Kerberos 身份验证：

- Amazon EMR
- 将 WASB 用作存储的 Azure HDInsight

LDAP 身份验证

从版本 10.2.1 开始，可以为 Amazon EMR 群集版本 5.10 配置轻量级目录访问协议 (LDAP) 身份验证。

Sqoop

从版本 10.2.1 开始，可以使用以下新的 Sqoop 功能：

对 MapR Connector for Teradata 的支持

可以在 Spark 引擎上使用 MapR Connector for Teradata 从 Teradata 读取数据或向其写入数据。MapR Connector for Teradata 是专用于 Sqoop 的 Teradata Connector for Hadoop (TDCH) 连接器。在 Spark 引擎上运行 Sqoop 映射时，数据集成服务在默认情况下会调用该连接器。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》。

Sqoop 传递映射的 Spark 引擎优化

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

- 您从 Sqoop 源读取数据并将数据写入到使用文本格式的 Hive 目标。
- 您从 Sqoop 源读取数据并将数据写入到使用平面格式、Avro 格式或 Parquet 格式的 HDFS 目标。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》。

高可用性功能和安全功能的 Spark 引擎支持

Sqoop 采用 Spark 引擎支持的所有高可用性功能和安全功能，例如 Kerberos keytab 登录和 KMS 加密。

有关详细信息，请参阅《*Informatica 10.2.1 应用程序服务指南*》中的“数据集成服务”一章和《*Informatica 10.2.1 命令参考指南*》中的“infacmd dis 命令参考”一章。

Teradata 数据对象的 Spark 引擎支持

如果您使用 Teradata 数据对象并在 Hortonworks 或 Cloudera 群集中运行映射，则数据集成服务会通过 Sqoop 运行映射。

如果使用的是 Hortonworks 群集，则数据集成服务会在运行时调用 Hortonworks Connector for Teradata。如果使用的是 Cloudera 群集，则数据集成服务会在运行时调用 Cloudera Connector Powered by Teradata。

有关详细信息，请参阅《*Informatica PowerExchange for Teradata Parallel Transporter API 10.2.1 用户指南*》。

Hadoop 环境中的转换支持

本节介绍版本 10.2.1 中 Hadoop 环境新增的转换功能。

Spark 引擎上的转换支持

本节介绍版本 10.2.1 中 Spark 引擎新增的转换功能。

转换支持

从版本 10.2.1 开始，Spark 引擎支持以下转换：

- 大小写转换器
- 分类器
- 比较
- 键生成器
- 标签创建器
- 合并

- 解析器
- Python
- 标准创建器
- 加权平均值

从版本 10.2.1 开始，Spark 引擎有限制地支持以下转换：

- 地址验证器
- 合并
- 判定
- 匹配
- 序列生成器

从版本 10.2.1 开始，Spark 引擎还支持以下转换：

- Java 支持用于处理层次结构数据的复杂数据类型，例如数组、映射和结构。

有关转换支持的详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》中的"Hadoop 环境中的映射转换"一章。

有关转换操作的详细信息，请参阅《*Informatica 10.2.1 Developer 转换指南*》。

Python 转换

从版本 10.2.1 开始，可以在 Developer tool 中创建 Python 转换。使用 Python 转换可在 Spark 引擎上运行的映射中执行 Python 代码。

可以使用 Python 转换对您在转换中传递的数据实施某个计算机模型。例如，使用 Python 转换编写可加载预训练模型的 Python 代码。可以使用预训练模型对输入数据进行分类或创建预测。

注意：Python 转换可用于技术预览版。支持技术预览版功能，但尚未做好生产准备。Informatica 建议您只在非生产环境中使用。

有关详细信息，请参阅《*Informatica 10.2.1 Developer 转换指南*》中的"Python 转换"一章。

更新策略转换

从版本 10.2.1 开始，可以对 Spark 引擎上运行的映射使用 Hive MERGE 语句来执行更新策略任务。在查询中使用 MERGE 通常更高效，并且有助于提高性能。

以下 Hadoop 发行版支持 Hive MERGE 语句：

- Amazon EMR 5.10
- Azure HDInsight 3.6
- Hortonworks HDP 2.6

要使用 Hive MERGE，请在更新策略转换的高级属性中选择选项。

在以前的版本中，数据集成服务使用 INSERT、UPDATE 和 DELETE 语句在运行时引擎上执行此任务。更新策略转换在下列情况下仍使用这些语句：

- 您未选择 Hive MERGE 选项。
- 映射在 Hive 或 Blaze 引擎上运行。
- Hadoop 发行版不支持 Hive MERGE。

有关在更新策略转换中使用 MERGE 语句的详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》中有关更新策略转换一章。

Blaze 引擎上的转换支持

本节介绍版本 10.2.1 中 Blaze 引擎新增的转换功能。

汇总器转换

从版本 10.2.1 开始，汇总器转换的数据缓存使用可变长度在 Blaze 引擎上存储 binary 和 string 数据类型。运行汇总器转换时，可变长度可以减少数据缓存所存储的数据量。

如果使用可变长度将汇总器转换中传递的数据存储在数据存储中，则汇总器转换会优化以使用已排序输入，并且排序器转换会在运行时映射中插入到汇总器转换之前。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》中的"Hadoop 环境中的映射转换"一章。

匹配转换

从版本 10.2.1 开始，可以在 Blaze 引擎上运行包含为标识分析配置的匹配转换的映射。

请将匹配转换配置为将标识索引数据写入到缓存文件。如果将匹配转换配置为将索引数据写入到数据库表，映射验证将失败。

有关转换支持的详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》中的"Hadoop 环境中的映射转换"一章。

等级转换

从版本 10.2.1 开始，等级转换的数据缓存使用可变长度在 Blaze 引擎上存储 binary 和 string 数据类型。运行等级转换时，可变长度可以减少数据缓存所存储的数据量。

如果使用可变长度将等级转换中传递的数据存储在数据存储中，则等级转换会优化以使用已排序输入，并且排序器转换会在运行时映射中插入到等级转换之前。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》中的"Hadoop 环境中的映射转换"一章。

有关转换操作的详细信息，请参阅《*Informatica 10.2.1 Developer 转换指南*》。

Big DataStreaming

本节介绍版本 10.2.1 中新增的 Big Data Streaming 功能。

源和目标

从版本 10.2.1 开始，可以在流映射中从以下源读取数据以及向以下目标写入数据：

- Azure 事件中心。创建 Azure EventHub 数据对象以从事件中心事件读取数据或向其写入数据。可以使用 Azure EventHub 连接来访问作为源或目标的 Microsoft Azure 事件中心。可以通过 Developer tool 或 infacmd 创建和管理 Azure EventHub 连接。
- Microsoft Azure Data Lake Store。创建 Azure Data Lake Store 数据对象以向 Azure Data Lake Store 写入数据。可以使用 Azure Data Lake Store 连接来访问作为目标的 Microsoft Azure Data Lake Store 表。可以在 Developer tool 中创建和管理 Microsoft Azure Data Lake Store 连接。
- 符合 JDBC 的数据库。使用 JDBC 连接创建关系数据对象。

有关详细信息，请参阅《*Informatica Big Data Streaming 10.2.1 用户指南*》中的"流映射中的源"一章和"流映射中的目标"一章。

流映射中的有状态计算

从版本 10.2.1 开始，通过在表达式转换中使用窗口函数，可以在流映射中执行有状态计算。

有关详细信息，请参阅《*Informatica Big Data Streaming 10.2.1 用户指南*》中的"流映射"一章。

转换支持

从版本 10.2.1 开始，可以在流映射中使用以下转换：

- 数据屏蔽
- 规范器
- Python

可以使用查找转换在流映射中对 HBase 数据执行未缓存的查找。

有关详细信息，请参阅《*Informatica Big Data Streaming 10.2.1 用户指南*》中的"流映射"一章。

截断已分区的 Hive 目标表

从版本 10.2.1 开始，可以截断具有或不具有分区的外部或受管 Hive 表。

有关截断 Hive 目标的详细信息，请参阅《*Informatica Big Data Streaming 10.2.1 用户指南*》中的"流映射中的目标"一章。

命令行程序

本节介绍版本 10.2.1 中的新命令。

infacmd autotune 命令

autotune 是新的 infacmd 插件，用于调整 Informatica 域中的服务和连接。

下表介绍了新的 infacmd autotune 命令：

命令	说明
autotune	根据大小说明使用建议的设置配置 Informatica 域中的服务和连接。

有关详细信息，请参阅《*Informatica 10.2.1 命令参考*》中的"infacmd autotune 命令参考"一章。

infacmd ccps 命令

ccps 是新的 infacmd 插件，用于对云平台群集执行操作。

下表介绍了新的 infacmd ccps 命令：

命令	说明
deleteClusters	从云平台上删除由群集工作流创建的群集。
listClusters	列出云平台上由群集工作流创建的群集。
updateADLSCertificate	更新 Azure Data Lake 服务主体证书。

有关详细信息，请参阅《*Informatica 10.2.1 命令参考*》中的"infacmd ccps 命令参考"一章。

infacmd cluster 命令

下表介绍了新的 infacmd cluster 命令：

命令	说明
updateConfiguration	更新群集配置的 Hadoop 发行版版本。 使用 -dv 选项可更新群集配置的 Hadoop 发行版版本。

下表介绍了对 infacmd cluster 命令进行的更改：

命令	更改说明
listConfigurationProperties	从版本 10.2.1 开始，使用 -cs 选项以返回常规配置集中的属性值时可以指定常规配置集。 在以前的版本中，-cs 选项仅接受 .xml 文件名。
createConfiguration	从版本 10.2.1 开始，在创建群集配置时，您可以选择性地使用 -dv 选项来指定 Hadoop 发行版版本。如果未指定版本，命令将使用指定 Hadoop 发行版的默认版本创建群集配置。 在以前的版本中，createConfiguration 命令不包含用于指定 Hadoop 版本的选项。

有关详细信息，请参阅《*Informatica 10.2.1 命令参考*》中的"infacmd cluster 命令参考"一章。

infacmd cms 命令

下表介绍了用于 infacmd cms updateServiceOptions 的新的内容管理服务选项：

命令	说明
DataServiceOptions.RefDataLocationSchema	标识一个架构，该架构指定引用数据数据库中的引用数据表。

有关详细信息，请参阅《*Informatica 10.2.1 命令参考*》中的"infacmd cms 命令参考"一章。

infacmd dis 命令

下表介绍了新的 infacmd dis 命令：

命令	说明
listMappingEngines	列出数据集成服务上已部署的映射的执行引擎。

有关详细信息，请参阅《*Informatica 10.2.1 命令参考*》中的"infacmd dis 命令参考"一章。

infacmd ihs 命令

下表介绍了新的 infacmd ihs 命令：

命令	说明
ListServiceProcessOptions	列出 Informatica 群集服务的进程选项。
UpdateServiceProcessOptions	更新 Informatica 群集服务的服务选项。

有关详细信息，请参阅《*Informatica 10.2.1 命令引用*》中的"infacmd ihs 命令引用"一章。

infacmd isp 命令

下表介绍了新的 infacmd isp 命令：

命令	说明
PingDomain	对域、服务、域网关主机或节点执行 Ping 操作。
GetPasswordComplexityConfig	返回域用户的密码复杂度配置。
ListWeakPasswordUsers	列出密码不符合密码策略的用户。

有关详细信息，请参阅《*Informatica 10.2.1 命令引用*》中的"infacmd isp 命令引用"一章。

infacmd ldm 命令

下表介绍了新的 infacmd ldm 命令：

命令	说明
ListServiceProcessOptions	列出 Catalog Administrator 进程的选项。
UpdateServiceProcessOptions	更新目录服务的进程选项。

有关详细信息，请参阅《*Informatica 10.2.1 命令引用*》中的"infacmd ldm 命令引用"一章。

infacmd mi 命令

mi 是新的 infacmd 插件，用于执行 Mass Ingestion 操作。

下表介绍了新的 infacmd mi 命令：

命令	说明
abortRun	中止 Mass Ingestion 规范的运行实例中的引入映射作业。
createService	创建 Mass Ingestion 服务。默认已禁用。 要启用 Mass Ingestion 服务，请使用 infacmd isp enableService。
deploySpec	部署 Mass Ingestion 规范。
exportSpec	将 Mass Ingestion 规范导出到应用程序存档文件中。
extendedRunStats	获取已部署 Mass Ingestion 规范中的映射的扩展统计信息。
getSpecRunStats	获取已部署 Mass Ingestion 规范的详细运行统计信息。
listSpecRuns	列出已部署 Mass Ingestion 规范的运行实例。
listSpecs	列出 Mass Ingestion 规范。
restartMapping	重新启动 Mass Ingestion 规范中的引入映射作业。
runSpec	运行已部署到数据集成服务的 Mass Ingestion 规范。

有关详细信息，请参阅《*Informatica 10.2.1 命令参考*》中的"infacmd mi 命令参考"一章。

infacmd mrs 命令

下表介绍了新的 infacmd mrs 命令：

命令	说明
listMappingEngines	列出存储在模型存储库中的映射的执行引擎。
listPermissionOnProject	列出组和用户对多个项目的所有权限。
updateStatistics	更新 Microsoft SQL Server 上的监视模型存储库的统计信息。

有关详细信息，请参阅《*Informatica 10.2.1 命令参考*》中的"infacmd mrs 命令参考"一章。

infacmd wfs 命令

下表介绍了新的 infacmd wfs 命令：

命令	说明
pruneOldInstances	从工作流数据库中删除工作流处理数据。

要删除处理数据，您必须具有该域的"管理服务"特权。

有关详细信息，请参阅《Informatica 10.2.1 命令参考》中的"infacmd wfs 命令参考"一章。

infasetup 命令

下表介绍了新增的 infasetup 命令：

命令	说明
UpdatePasswordComplexityConfig	启用或禁用域的密码复杂度配置。

有关详细信息，请参阅《Informatica 10.2.1 命令引用》中的“infasetup 命令引用”一章。

Enterprise Data Catalog

本节介绍版本 10.2.1 中新增的 Enterprise Data Catalog 功能。

向资产添加业务标题

从版本 10.2.1 开始，Business Glossary 资产和 Axon 词汇表资产除外，可以向目录中的任何其他资产添加业务标题。可以通过关联业务术语或提供显示名称来向资产添加业务标题。

有关添加业务标题的详细信息，请参阅《Informatica 10.2.1 Enterprise Data Catalog 用户指南》。

安装程序中的群集验证实用程序

从版本 10.2.1 开始，在安装 Enterprise Data Catalog 时，安装程序会提供运行群集验证实用程序的选项。该实用程序可帮助您验证在嵌入式群集或现有群集中安装 Enterprise Data Catalog 的先决条件。该实用程序还能验证 Informatica 域、群集主机和 Hadoop 群集服务的配置设置。

有关该实用程序的详细信息，请参阅《Informatica Enterprise Data Catalog 10.2.1 安装和配置指南》和以下知识库文章：

- 如何使用 Enterprise Information Catalog 中的验证实用程序验证嵌入式群集先决条件
- 如何验证 Informatica 域、群集主机和群集服务的配置

数据域发现类型

从版本 10.2.1 开始，在配置数据域发现配置文件设置时，可以选择以下数据域发现类型之一：

- 对源数据运行发现。扫描程序将对源数据运行数据域发现。
- 对源元数据运行发现。扫描程序将对源元数据运行数据域发现。
- 同时对源元数据和源数据运行发现。扫描程序将对源数据和源元数据运行数据域发现。
- 对元数据与其匹配的源数据运行发现。扫描程序将对源元数据运行数据域发现，以确定具有推理数据域的列。扫描程序随后将对具有推理数据域的列的源数据运行发现。

有关数据域发现类型的详细信息，请参阅《Informatica 10.2.1 Catalog Administrator 指南》。

筛选器设置

从版本 10.2.1 开始，可以使用“应用程序配置”页面中的筛选器设置来自定义搜索结果页面的**筛选条件**面板中显示的搜索筛选器。

有关搜索筛选器的详细信息，请参阅《*Informatica Enterprise Data Catalog 10.2.1 用户指南*》。

缺少的链接报表

从版本 10.2.1 开始，将资源的架构分配给连接后，可以生成缺少的链接报表来标识缺少的连接链接。

有关缺少的链接报表的详细信息，请参阅《*Informatica 10.2.1 Catalog Administrator 指南*》。

新的资源类型

从版本 10.2.1 开始，Informatica Enterprise Data Catalog 可从几个新的数据源提取元数据。

您可以在 Informatica Catalog Administrator 中创建资源，以便从以下数据源中提取元数据：
Azure Data Lake Store

在线云文件存储平台。

数据库脚本

数据库脚本用于提取沿袭信息。数据库脚本资源可用于技术预览版。支持技术预览版功能，但无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。

Microsoft Azure Blob 存储

基于云的文件存储 Web 服务。

QlikView

使用业务情报工具，可以从 QlikView 源系统提取元数据。

SharePoint

从 SharePoint 中的文件导入元数据。

OneDrive

从 OneDrive 中的文件导入元数据。

有关新资源的详细信息，请参阅《*Informatica 10.2.1 Catalog Administrator 指南*》。

REST API

从版本 10.2.1 开始，可以使用 Informatica Enterprise Data Catalog REST API 来加载和监视资源。

有关 REST API 的详细信息，请参阅《*Informatica 10.2.1 Enterprise Data Catalog REST API 参考*》。

Enterprise Data Catalog 应用程序的 SAML 身份验证

从版本 10.2.1 开始，可以为 Enterprise Data Catalog 应用程序启用使用 SAML 身份验证的单点登录。可以将基于 OKTA 的 SAML 身份验证与 Active Directory 一起使用，或者将 Active Directory 联合身份验证服务与 Active Directory 一起使用。

有关详细信息，请参阅《*Informatica Enterprise Data Catalog 10.2.1 安装和配置指南*》。

SAP 资源

从版本 10.2.1 开始，可以为 SAP R/3 资源选择**为数据访问启用流**选项，以便使用 HTTP 协议来提取数据。

有关该选项的详细信息，请参阅《*Informatica 10.2.1 Catalog Administrator 指南*》。

从 ServiceNow 导入

从版本 10.2.1 开始，Catalog Administrator 连接到 ServiceNow 以导入连接并将配置元数据提取到目录中。

"从 ServiceNow 导入"功能可用于技术预览版。支持技术预览版功能，但无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。

有关从 ServiceNow 导入元数据的详细信息，请参阅《*Informatica 10.2.1 Catalog Administrator 指南*》。

类似列

从版本 10.2.1 开始，可以查看"类似列"部分，此部分显示与正在查看的列类似的所有列。Enterprise Data Catalog 基于列名称、列模式、唯一值和值频率发现类似列。

有关列相似性的详细信息，请参阅《*Informatica 10.2.1 Enterprise Data Catalog 用户指南*》。

指定目录服务的加载类型

从版本 10.2.1 开始，在创建目录服务时，您可以选择指定要部署的数据大小的选项。

在以前的版本中，您必须创建目录服务并使用目录服务的自定义属性来指定数据大小。

有关详细信息，请参阅《*Informatica Enterprise Data Catalog 10.2.1 安装和配置指南*》。

数据发现支持的资源类型

从版本 10.2.1 开始，可以为以下资源启用数据发现以提取剖析元数据：

- 非结构化文件类型：
 - Apple 文件。支持的扩展名类型包括 .key、.pages、.numbers、.ibooks 和 .ipa。
 - OpenOffice 文件。支持的扩展名类型包括 .odt、.ott、.odm、.ods、.ots、.odp、.odg、.otp、.odg、.otg 和 .odf。
- 结构化文件类型：
 - Avro。支持的扩展名类型为 .avro。
此文件类型可用于 HDFS 资源和文件系统资源。对于文件系统资源，只能选择"本地文件"协议。
 - Parquet。支持的扩展名类型为 .parquet。
此文件类型可用于 HDFS 资源和文件系统资源。对于文件系统资源，只能选择"本地文件"协议。
- 其他资源：
 - Azure Data Lake Store
 - 文件系统。支持的协议包括"本地文件"、SFTP 和 SMB/CIFS 协议。
 - HDFS。支持的发行版包括 MapR FS。
 - Microsoft Azure Blob 存储
 - OneDrive

- SharePoint

有关新资源的详细信息，请参阅《*Informatica 10.2.1 Catalog Administrator 指南*》。

Enterprise Data Lake

本节介绍版本 10.2.1 中新增的 Enterprise Data Lake 功能。

列数据

从版本 10.2.1 开始，可以在处理工作表中的列时使用以下功能：

- 可以将列中的相关值分类或分组为各个类别，以使分析变得更简单。
- 可以查看工作表中的所选列的数据源。您可能需要查看列中的数据源来帮助排除问题。
- 可以将执行列采样期间推理的类型或数据域还原为源类型。如果要在公式中使用列数据，您可能需要将推理的类型或数据域还原为源类型。

有关详细信息，请参阅《*Informatica 10.2.1 Enterprise Data Lake 用户指南*》中的"准备数据"一章。

管理 Data Lake 资源

从版本 10.2.1 开始，可以使用 Enterprise Data Lake 应用程序来添加和删除 Enterprise Data Catalog 资源。目录资源代表外部数据源和元数据存储库，扫描程序将从中提取可在 Data Lake 中使用的元数据。

有关详细信息，请参阅《*Informatica 10.2.1 Enterprise Data Lake 管理员指南*》中的"管理 Data Lake"一章。

数据准备操作

从版本 10.2.1 开始，可以在数据准备期间执行以下操作：

透视数据

可以使用透视操作将工作表中所选列的数据结构更改为汇总格式。使用透视操作，您可以对用于分析的数据进行分组和汇总，例如汇总年中前六个月在每个城市销售的单式家庭房屋的平均价格。

逆透视数据

可以使用逆透视操作将工作表中的列转换为包含键值格式列数据的行。要基于键和相应的值将工作表中的数据汇总到行中，可以使用逆透视操作。

应用 One-Hot 编码

可以使用 One-Hot 编码操作来确定所选列中的字符串值在工作表中的每个行上是否存在。可以使用 One-Hot 编码操作将工作表中的分类值转换为机器学习算法所需的数值。

有关详细信息，请参阅《*Informatica 10.2.1 Enterprise Data Lake 用户指南*》中的"准备数据"一章。

准备 JSON 文件

从版本 10.2.1 开始，在数据准备过程的第一步将 JavaScript Object Notation Lines (JSONL) 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 Lake 会将 JSON 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对数据进行采样。

有关详细信息，请参阅《*Informatica 10.2.1 Enterprise Data Lake 用户指南*》中的"准备数据"一章。

脚本步骤

从版本 10.2.1 开始，可以在处理工作表中的脚本时使用以下功能：

- 可以重用在工作表中创建的脚本步骤，包括包含复杂公式或规则定义的步骤。可以重用同一个或不同工作表中的脚本步骤，包括其他项目中的工作表。可以复制并重用从脚本中选择的步骤，或者也可以重用整个脚本。
- 可以在脚本中的任何位置插入步骤。
- 可以添加筛选器或修改已应用到脚本步骤的筛选器。

有关详细信息，请参阅《*Informatica 10.2.1 Enterprise Data Lake 用户指南*》中的"准备数据"一章。

计划导出、导入和发布活动

从版本 10.2.1 开始，可以计划数据资产的导出、导入和发布。通过计划活动，您可以定期导入、导出或发布已更新的数据资产。

计划活动时，可以创建新计划，或者也可以选择现有计划。您可以使用其他用户创建的计划，其他用户也可以使用您创建的计划。

有关详细信息，请参阅《*Informatica 10.2.1 Enterprise Data Lake 用户指南*》中的"计划导出、导入和发布活动"一章。

安全断言标记语言身份验证

从版本 10.2.1 开始，Enterprise Data Lake 应用程序支持安全断言标记语言 (SAML) 身份验证。

有关配置 SAML 身份验证的详细信息，请参阅《*Informatica 10.2.1 安全指南*》。

查看项目流程和项目历史记录

从版本 10.2.1 开始，可以查看项目流程图以及检查在项目中执行的活动。

可以查看流程图，其中显示项目中工作表之间的关系以及工作表是如何派生的。当处理包含大量工作表和大量资产的复杂项目时，该图特别有用。

还可以检查在项目中执行的活动的完整历史记录，包括对项目中的工作表执行的活动。查看项目历史记录可以帮助您确定项目中出现的问题的根本原因。

有关详细信息，请参阅《*Informatica 10.2.1 Enterprise Data Lake 用户指南*》中的"创建和管理项目"一章。

Informatica Developer

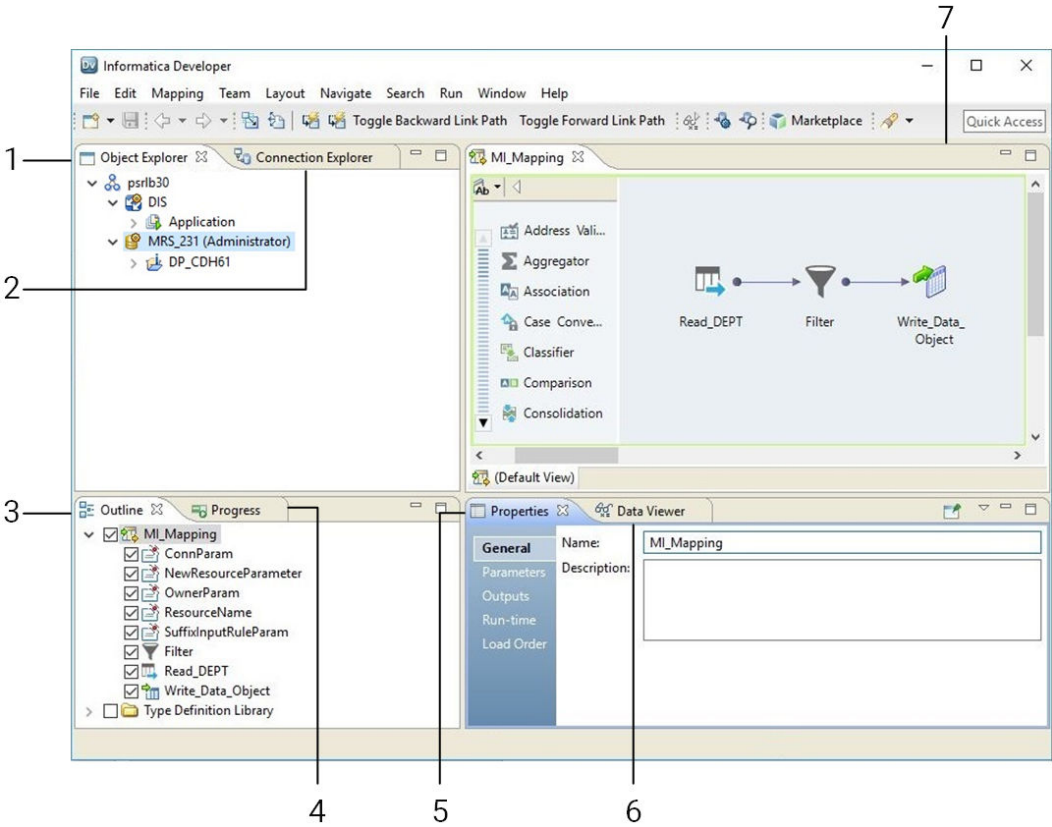
本节介绍版本 10.2.1 中新增的 Developer tool 功能。

默认布局

从版本 10.2.1 开始，默认情况下 Developer tool 工作台会显示以下附加视图：

- "连接浏览器"视图
- "进程"视图

下图显示了版本 10.2.1 中默认的 Developer tool 工作台：



1. "对象浏览器"视图
2. "连接浏览器"视图
3. "大纲"视图
4. "进程"视图
5. "属性"视图
6. "数据查看器"视图
7. 编辑器

有关详细信息，请参阅《Informatica 10.2.1 Developer tool 指南》中的"Informatica Developer"一章。

编辑器搜索

从版本 10.2.1 开始，可以在编辑器视图中搜索映射和 Maplet 中的复杂数据类型定义。还可以使用复杂数据类型定义显示链接路径。

有关详细信息，请参阅《Informatica 10.2.1 Developer tool 指南》中的"在 Informatica Developer 中搜索"一章。

从 PowerCenter 导入会话属性

从版本 10.2.1 开始，可以从 PowerCenter 存储库将会话属性（例如关系源和目标中基于 SQL 的替代和查找转换的替代）导入到模型存储库。

有关"从 PowerCenter 导入"功能的详细信息，请参阅《Informatica 10.2.1 Developer 映射指南》中的"从 PowerCenter 导入"一章。

视图

从版本 10.2.1 开始，可以扩展复杂数据类型，以便在以下视图中查看复杂数据类型定义：

- 编辑器视图
- "大纲"视图
- "属性"视图

有关详细信息，请参阅《*Informatica 10.2.1 Developer tool 指南*》中的"Informatica Developer"一章。

Informatica 映射

本节介绍版本 10.2.1 中新增的 Informatica 映射功能。

动态映射

本节介绍版本 10.2.1 中新增的动态映射功能。

输入规则

从版本 10.2.1 开始，可以在创建输入规则时执行以下任务：

- 按复杂数据类型定义创建输入规则。
- 还原源端口名称（如果已重命名生成的端口）。
- 按源名称选择端口（如果按列名称或模式创建输入规则）。
- 在端口预览中查看源名称和复杂数据类型定义。

有关详细信息，请参阅《*Informatica 10.2.1 Developer 映射指南*》中的"动态映射"一章。

端口选择器

从版本 10.2.1 开始，可以对端口选择器进行配置，使其按复杂数据类型定义选择端口。

有关详细信息，请参阅《*Informatica 10.2.1 Developer 映射指南*》中的"动态映射"一章。

验证动态源和目标

从版本 10.2.1 开始，可以验证动态源和目标。要验证动态源和目标，请解析映射参数以查看运行时映射实例。验证运行时映射实例。

有关详细信息，请参阅《*Informatica 10.2.1 Developer 映射指南*》中的"动态映射"一章。

映射参数

本节介绍版本 10.2.1 中新增的映射参数功能。

分配参数

从版本 10.2.1 开始，可以为以下映射对象和对象字段分配参数：

对象	字段
自定义数据对象读取操作	自定义查询 筛选条件 联接条件 PreSQL PostSQL
自定义数据对象写入操作	PreSQL PostSQL 更新替代
平面文件数据对象	压缩编解码器 压缩格式
查找转换	自定义查询。仅限关系。
读取转换	自定义查询。仅限关系。 筛选条件。仅限关系。 联接条件。仅限关系。 PreSQL。仅限关系。 PostSQL。仅限关系。
写入转换	PreSQL。仅限关系。 PostSQL。仅限关系。 更新替代。仅限关系。

有关详细信息，请参阅《Informatica 10.2.1 Developer 映射指南》中的"映射参数"一章。

解析映射参数

从版本 10.2.1 开始，可以在 Developer tool 中解析映射参数。解析映射参数时，Developer tool 会生成一个运行时映射实例，该实例显示数据集成服务在运行时如何解析参数。您可以运行其中的参数已解析的映射实例，以使用所选参数运行映射。

下表介绍了可用于解析映射参数的选项：

映射参数	说明
在映射中应用默认值	根据映射中为参数配置的默认值解析映射参数。如果没有为映射配置参数，则不会在映射中解析任何参数。
应用参数集	根据指定参数集中指定的参数值解析映射参数。
应用参数文件	根据指定参数文件中指定的参数值解析映射参数。

要根据参数集快速解析映射参数，请从 Object Explorer 视图将参数集拖放到映射编辑器中，以便在运行时映射实例中查看已解析的参数。

有关详细信息，请参阅《*Informatica 10.2.1 Developer 映射指南*》中的“映射参数”一章。

验证映射参数

从版本 10.2.1 开始，可以在 Developer tool 中验证映射参数。要验证映射参数，请先解析映射参数。解析映射参数时，Developer tool 会生成一个运行时映射实例，该实例显示已解析的参数。请验证运行时映射实例，以验证映射参数。

有关详细信息，请参阅《*Informatica 10.2.1 Developer 映射指南*》中的“映射参数”一章。

运行映射

本节介绍版本 10.2.1 中用于运行映射的新功能。

从 Object Explorer 视图运行映射

从版本 10.2.1 开始，可以从 Object Explorer 视图运行映射。不必在映射编辑器中打开映射。在 Object Explorer 视图中右键单击该映射，然后单击“运行”。

有关详细信息，请参阅《*Informatica 10.2.1 Developer tool 指南*》。

使用高级选项运行映射

从版本 10.2.1 开始，可以在 Developer tool 中使用高级选项运行映射。在高级选项中，可以指定映射配置和映射参数。每次运行映射时都应指定映射配置和映射参数。

下表介绍了可用于指定映射配置的选项：

选项	说明
选择映射配置	从下拉菜单中选择映射配置。要创建新的映射配置，请选择“新建配置”。
指定自定义映射配置	创建在运行当前映射期间持续存在的自定义映射配置。

下表介绍了可用于指定映射参数的选项：

映射参数	说明
在映射中应用默认值	根据映射中为参数配置的默认值解析映射参数。如果没有为映射配置参数，则不会在映射中解析任何参数。
应用参数集	根据指定参数集中指定的参数值解析映射参数。
应用参数文件	根据指定参数文件中指定的参数值解析映射参数。

有关详细信息，请参阅《*Informatica 10.2.1 Developer 映射指南*》。

截断已分区的 Hive 目标表

从版本 10.2.1 开始，可以截断具有或不具有分区的外部或受管 Hive 表。

在以前的版本中，可以将映射设计为截断 Hive 目标表而不截断外部的已分区 Hive 目标表。

有关截断 Hive 目标的详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》中的“Hadoop 环境中的映射目标”一章。

Informatica 转换语言

本节介绍版本 10.2.1 中新增的 Informatica 转换语言功能。

适用于 map 数据类型的复杂函数

从版本 10.2.1 开始，转换语言引入了适用于 map 数据类型的复杂函数。可以使用适用于 map 数据类型的复杂函数在 Spark 引擎上生成或处理映射数据。

转换语言包含以下适用于 map 数据类型的复杂函数：

- COLLECT_MAP
- MAP
- MAP_FROM_ARRAYS
- MAP_KEYS
- MAP_VALUES

从版本 10.2.1 开始，可以使用 SIZE 函数来确定映射数据的大小。

有关复杂函数的详细信息，请参阅《*Informatica 10.2.1 Developer 转换语言参考*》中的"函数"一章。

适用于 map 数据类型的复杂运算符

从版本 10.2.1 开始，可以在 Spark 引擎上运行的映射中使用复杂运算符来访问 map 数据类型中的元素。

map 数据类型包含键值对元素的无序集合。使用下标运算符 [] 可以访问 map 数据类型中的给定键对应的值。

有关复杂运算符的详细信息，请参阅《*Informatica 10.2.1 Developer 转换语言参考*》中的"运算符"一章。

Informatica 转换

本节介绍版本 10.2.1 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：

阿根廷

从版本 10.2.1 开始，可以对 Informatica 进行配置，使其返回对输入到单行中的阿根廷地址有效的建议。

请按以下格式输入阿根廷地址：

[Street] [House Number] [Dependent Locality] [Post Code] [Locality]

要验证单行地址，请在完整地址端口中输入地址。

巴西

从版本 10.2.1 开始，可以对 Informatica 进行配置，使其返回对输入到单行中的巴西地址有效的建议。

请按以下格式输入巴西地址：

[Street] [House Number] [Locality] [State Code] [Post Code]

要验证单行地址，请在完整地址端口中输入地址。

哥伦比亚

从版本 10.2.1 开始，Informatica 在验证哥伦比亚地址时将验证到门牌号级别。

香港

从版本 10.2.1 开始，Informatica 对香港地址支持屋顶级地理编码。对于以中文或英语提交的香港地址，Informatica 可以返回其屋顶级地理坐标。

生成地理坐标时，Informatica 会将三个级别的建筑物信息全部考虑在内。它会提供直到所验证地址中可用的最低级别的屋顶级地理坐标。

要检索香港地址的屋顶级地理坐标，请安装 HKG5GCRT.MD 数据库。

印度

从版本 10.2.1 开始，Informatica 在验证印度地址时将验证到门牌号级别。

墨西哥

从版本 10.2.1 开始，可以对 Informatica 进行配置，使其返回对输入到单行中的墨西哥地址有效的建议。

请按以下格式输入墨西哥地址：

[Street] [House Number] [Sub-locality] [Post Code] [Locality] [Province]

要验证单行地址，请在完整地址端口中输入地址。

南非

从版本 10.2.1 开始，Informatica 改进了对南非地址中的投递服务描述符的解析和验证。

Informatica 通过以下方式改进了对投递服务描述符的解析和验证：

- Address Verification 将"专用邮袋"、"团体信箱"、"邮政信箱"和 Postnet Suite 识别为不同的投递服务类型。Address Verification 不会将一个投递服务描述符标准化为其他投递服务描述符。例如，Address Verification 不会将 Postnet Suite 标准化为"邮政信箱"。
- Address Verification 会将 Postnet Box 解析为非标准投递服务描述符，并将 Postnet Box 更正为有效的描述符 Postnet Suite。
- Address Verification 不会将子建筑物描述符 Flat 标准化为 Fl。

韩国

从版本 10.2.1 开始，Informatica 引入了以下适用于韩国地址的功能和增强功能：

- 韩国地址引用数据包含建筑物信息。Informatica 可以读取、验证和更正韩国地址中的建筑物信息。
- Informatica 将返回旧地址所代表的房屋的所有当前地址。旧地址可能代表单个当前地址，或者可能代表多个地址，例如，如果多位居民居住在该房屋地点。

要返回当前地址，请先查找旧房屋的地址 ID。以地址代码查找模式提交最后一个字符为 A 的地址 ID 时，Informatica 将返回与该地址 ID 匹配的所有当前地址。

注意：地址验证器转换使用 Max Result Count 属性来确定输入的地址 ID 返回的最大地址个数。Count Overflow 属性指示数据库是否包含该地址 ID 的其他地址。

泰国

从版本 10.2.1 开始，Informatica 引入了以下适用于泰国地址的功能和增强功能：

对泰国地址进行的改进

Informatica 改进了对采用拉丁语种的泰国地址的解析和验证。

此外，Informatica 在验证地址时会验证到门牌号级别。

对泰国地址的本地支持

Informatica 可以读取和写入采用本地泰语语种和拉丁语种的泰国地址。Informatica 更新了泰国地址的引用数据，并添加了采用本地泰语语种的引用数据。

Informatica 为每种语种的泰国地址提供单独的引用数据库。要验证采用本地泰语语种的地址，请安装本地泰语数据库。要验证采用拉丁语种的地址，请安装拉丁语数据库。

注意: 如果要验证泰国地址，请勿同时安装这两种类型的数据库。可以接受 Preferred Script 属性的默认选项。

阿拉伯联合酋长国

从版本 10.2.1 开始，Informatica 将验证阿拉伯联合酋长国地址中的街道名称。要验证阿拉伯联合酋长国地址中的街道名称，请安装当前的阿拉伯联合酋长国地址引用数据库。

英国

从版本 10.2.1 开始，Informatica 可以返回英国领地名称。

Informatica 在 Country_2 元素中返回领地名称。Informatica 在 Country_1 元素中返回国家/地区名称。您可以为输出地址配置这两个元素；或者如果是在英国境内投递邮件，则可以省略 Country_1 元素。在信封或标签上，领地名称显示在英国地址中的邮政编码上方。

要返回领地名称，请安装当前的英国地址引用数据。

美国

从版本 10.2.1 开始，Informatica 最多可以识别美国地址中的三个子建筑物级别。

按照美国邮政署的要求，Informatica 将单个子建筑物元素中的信息与引用数据匹配。如果 Sub-building_1 中的信息不匹配，则 Informatica 会比较 Sub-building_2 中的信息。如果 Sub-building_2 中的信息不匹配，则地址验证会比较 Sub-building_3 中的信息。地址验证会将输入地址中不匹配的子建筑物信息复制到输出地址。

奥地利、德国和瑞士

从版本 10.2.1 开始，Informatica 支持在奥地利地址、德国地址和瑞士地址中使用大写字符 ß。

Informatica 通过以下方式支持字符 ß：

- 如果将 Casing 属性设置为 UPPER，则 Informatica 会为德语字符 ß 返回 ß。如果将 Casing 属性设置为 LOWER，则 Informatica 会为德语字符 ß 返回 ß。
- Informatica 将地址中的 ß 和 ß 视为同样有效的字符。在引用数据匹配中，当相同值包含 ß 或 ß 时，Informatica 会将其确定为完美匹配。
- Informatica 将地址中的 ß 和 ss 视为同样有效的字符。在引用数据匹配中，当相同值包含 ß 或 ss 时，Informatica 会将其确定为标准匹配。
- 如果将 Preferred Script 属性设置为 ASCII_SIMPLIFIED，则 Informatica 会字符 ß 返回 S。
- 如果将 Preferred Script 属性设置为 ASCII_EXTENDED，则 Informatica 会字符 ß 返回 SS。

有关 Informatica 嵌入到版本 10.2.1 的地址验证软件引擎版本的功能和操作的全面信息，请参阅《*Informatica Address Verification 5.12.0 开发人员指南*》。

Informatica 工作流

本节介绍版本 10.2.1 中新增的 Informatica 工作流功能。

从 PowerCenter 导入命令任务

从版本 10.2.1 开始，可以从 PowerCenter 将命令任务导入到模型存储库。

有关详细信息，请参阅《*Informatica 10.2.1 Developer 工作流指南*》中的"工作流"一章。

Informatica PowerExchange 适配器

本节介绍版本 10.2.1 中新增的 Informatica 适配器功能。

PowerExchange for Amazon Redshift

从版本 10.2.1 开始，PowerExchange for Amazon Redshift 包含以下功能：

- 可以在 Spark 引擎上配置已缓存的查找操作以缓存查找表，而在本地环境中配置未缓存的查找操作。
- 对于服务器端加密，可以在本地环境和 Spark 引擎的连接中配置由 AWS Key Management Service 生成的客户主密钥 ID。

有关详细信息，请参阅《*Informatica PowerExchange for Amazon Redshift 10.2.1 用户指南*》。

PowerExchange for Amazon S3

从版本 10.2.1 开始，PowerExchange for Amazon S3 包含以下功能：

- 对于客户端加密，可以在本地环境的连接中配置由 AWS Key Management Service 生成的客户主密钥 ID。对于服务器端加密，可以在本地环境和 Spark 引擎的连接中配置由 AWS Key Management Service 生成的客户主密钥 ID。
- 对于服务器端加密，可以配置由 Amazon S3 管理的加密密钥或由 AWS KMS 管理的客户主密钥，并将文件上载到存储桶中。
- 可以在 Amazon S3 中使用以下数据源格式创建 Amazon S3 文件数据对象：
 - 智能结构模型
PowerExchange for Amazon S3 的智能结构模型功能可用于技术预览版。支持技术预览版功能，但尚未做好生产准备。Informatica 建议您只在非生产环境中使用。
 - JSON
 - ORC
- 在本地环境和 Spark 引擎中向 Amazon S3 写入数据时，可以使用 Zlib 压缩格式来压缩 ORC 数据。
- 可以在目标会话属性中使用**创建目标**选项创建 Amazon S3 目标。
- 可以在 Spark 引擎上使用复杂数据类型来读取和写入 Avro 和 Parquet 文件格式的层次结构数据。
- 可以在映射中使用 Amazon S3 源作为动态源。PowerExchange for Amazon S3 源的动态映射支持可用于技术预览版。支持技术预览版功能，但无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 Amazon S3 10.2.1 用户指南*》。

PowerExchange for Cassandra

从版本 10.2.1 开始，Informatica Cassandra ODBC 驱动程序支持异步写入。

要在 Linux 操作系统上启用异步写入，必须在 `odbc.ini` 文件中添加 **EnableAsynchronousWrites** 项名称并将值设置为 1。

要在 Windows 操作系统上启用异步写入，必须在 Windows 注册表中为 Cassandra ODBC 数据源名称添加 **EnableAsynchronousWrites** 属性并将值设置为 1。

有关详细信息，请参阅《*Informatica PowerExchange for Cassandra 10.2.1 用户指南*》。

PowerExchange for HBase

从版本 10.2.1 开始，可以使用 HBase 数据对象读取操作在 HBase 资源中查找数据。在本地环境中或在 Spark 引擎上运行映射，以在 HBase 资源中查找数据。可以启用查找缓存，还可以将查找条件参数化。

PowerExchange for HBase 的查找功能可用于技术预览版。支持技术预览版功能，但尚未做好生产准备。Informatica 建议您只在非生产环境中使用。

有关详细信息，请参阅《*Informatica PowerExchange for HBase 10.2.1 用户指南*》。

PowerExchange for HDFS

从版本 10.2.1 开始，可以使用以下新的 PowerExchange for HDFS 功能：

复杂文件数据对象的智能结构模型支持

可以在复杂文件数据对象中纳入智能结构模型。向 Spark 引擎上运行的映射添加数据对象时，您可以处理该模型所能解析的任何输入类型。

PowerExchange for HDFS 的智能结构模型功能可用于技术预览版。支持技术预览版功能，但尚未做好生产准备。Informatica 建议您只在非生产环境中使用。

有关详细信息，请参阅《*Informatica PowerExchange for HDFS 10.2.1 用户指南*》。

复杂文件源的动态映射支持

可以在映射中使用复杂文件源作为动态源。

复杂文件源的动态映射支持可用于技术预览版。支持技术预览版功能，但无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。

有关动态映射的详细信息，请参阅《*Informatica Developer 映射指南*》。

PowerExchange for Hive

从版本 10.2.1 开始，PowerExchange for Hive 支持在 Spark 引擎上对 Hive 源和目标运行 PreSQL 和 PostSQL 查询的映射。

有关详细信息，请参阅《*Informatica PowerExchange for Hive 10.2.1 用户指南*》。

PowerExchange for Microsoft Azure Blob 存储

从版本 10.2.1 开始，PowerExchange for Microsoft Azure Blob 存储包含以下功能：

- 可以在 Spark 引擎上运行映射。
- 在 Spark 引擎和本地环境中运行映射时，可以读取和写入 .csv、Avro 和 Parquet 文件。
- 在 Spark 引擎上运行映射时，可以读取和写入 JSON 和智能结构文件。
- 在 Spark 引擎上运行映射时，可以读取目录。
- 在本地环境中运行映射时，可以生成或跳过标题行。在 Spark 引擎上，默认情况下会创建标题行。

- 可以附加现有 Blob。附加操作仅适用于附加 Blob，且仅在本地环境中适用。
- 可以替代 Blob 名称或容器名称。在"Blob 容器替代"字段中，指定容器名称或使用绝对路径指定根容器中的子文件夹。
- 可以读取和写入以 gzip 格式压缩的 .csv 文件。

PowerExchange for Microsoft Azure Blob 存储的所有新功能可用于技术预览版。支持技术预览版功能，但尚未做好生产准备。Informatica 建议您只在非生产环境中使用。

有关详细信息，请参阅《*Informatica PowerExchange for Microsoft Azure Blob 存储 10.2.1 用户指南*》。

PowerExchange for Microsoft Azure SQL 数据仓库

从版本 10.2.1 开始，PowerExchange for Microsoft Azure SQL 数据仓库包含以下功能：

- 可以在 Spark 引擎上运行映射。
- 从 Microsoft Azure SQL 数据仓库对象读取数据时，可以配置键范围分区。
- 从 Microsoft Azure SQL 数据仓库对象读取数据时，可以替代 SQL 查询并定义约束。
- 可为映射中的源对象和目标对象配置 pre-SQL 和 post-SQL 查询。
- 可以为源数据对象操作配置本地表达式筛选器。
- 可以对 Microsoft Azure SQL 数据仓库表执行更新、插入和删除操作。
- 可以在 Spark 引擎上配置已缓存的查找操作以缓存查找表，而在本地环境中配置未缓存的查找操作。

有关详细信息，请参阅《*Informatica PowerExchange for Microsoft Azure SQL Data Warehouse 10.2.1 用户指南*》。

PowerExchange for Salesforce

从版本 10.2.1 开始，可以使用版本 41 的 Salesforce API 来创建 Salesforce 连接以及访问 Salesforce 对象。可以在源转换和目标转换中使用大型对象。

有关详细信息，请参阅《*Informatica PowerExchange for Salesforce 10.2.1 用户指南*》。

PowerExchange for SAP NetWeaver

从版本 10.2.1 开始，可以在 Spark 引擎上运行映射以从 SAP 表读取数据。

有关详细信息，请参阅《*Informatica PowerExchange for SAP NetWeaver 10.2.1 用户指南*》。

PowerExchange for Snowflake

从版本 10.2.1 开始，PowerExchange for Snowflake 包含以下功能：

- 可以针对 Snowflake 表配置查找操作。还可以为查找操作启用查找缓存来提高查找性能。数据集成服务会缓存查找源，并在缓存中对行运行查询。
- 可以将 Snowflake 连接以及数据对象读取操作属性和写入操作属性参数化。
- 可以为读取操作或写入操作中的 Snowflake 数据对象配置键范围分区。数据集成服务将根据您定义为分区键的端口或端口集分发数据。
- 可以在高级目标属性中指定表名称，以替代 Snowflake 连接属性中的表名称。

有关详细信息，请参阅《*Informatica PowerExchange for Snowflake 10.2.1 用户指南*》。

安全

本节介绍版本 10.2.1 中的新安全功能。

密码复杂度

从版本 10.2.1 开始，您可以启用密码复杂度以验证密码强度。默认情况下，禁用此选项。

有关详细信息，请参阅《*Informatica 10.2.1 安全指南*》中的“Informatica Administrator 中的安全管理”一章。

第 20 章

更改 (10.2.1)

本章包括以下主题：

- [支持更改, 201](#)
- [安装程序更改, 204](#)
- [产品名称更改, 205](#)
- [应用程序服务, 205](#)
- [Big Data Management, 206](#)
- [Big DataStreaming, 210](#)
- [命令行程序, 211](#)
- [内容安装程序, 212](#)
- [Enterprise Data Catalog, 212](#)
- [Informatica Analyst, 214](#)
- [Informatica Developer, 214](#)
- [Informatica 转换, 215](#)
- [Informatica PowerExchange 适配器, 216](#)

支持更改

本节介绍版本 10.2.1 中的支持更改。

升级支持更改

在版本 10.2.1 中，Informatica 仅支持升级 Informatica 大数据产品，例如 Big Data Management 和 Big Data Quality。升级域时，传统产品（例如 PowerCenter 和 Informatica Data Quality）的功能不可用。

如果在同一个域中运行传统产品和大数据产品，则必须先拆分域，然后再升级。拆分域时，应创建域副本，以便您可以在单独的域中运行大数据产品和传统产品。应在域中的每台计算机上复制节点。还应复制传统产品和大数据产品通用的服务。拆分域后，可以升级运行大数据产品的域。

注意: 虽然 Informatica 传统产品在版本 10.2.1 中不受支持，但本文档确实包含某些对 PowerCenter 服务和 Metadata Manager 服务的引用。

Big Data Hadoop 发行版支持

Informatica 大数据产品支持各种 Hadoop 发行版。在每个版本中，Informatica 会增加、延续和终止对 Hadoop 发行版版本的支持。Informatica 可能会在将来的版本中恢复对延续版本的支持。

下表列出了 Informatica 10.2.1 大数据产品支持的 Hadoop 发行版版本：

产品	EMR	HDI	CDH	HDP	MapR
Big Data Management	5.10、5.14 ³	3.6.x	5.11 ¹ 、5.12 ¹ 、5.13、5.14、5.15	2.5、2.6	6.x MEP 5.0.x ²
Big Data Streaming	5.10、5.14 ³	3.6.x	5.11 ¹ 、5.12 ¹ 、5.13、5.14、5.15	2.5、2.6	6.x MEP 4.0.x
Enterprise Data Catalog	N/A	3.6.x	5.13	2.6.x	N/A
Enterprise Data Lake	5.10	3.6.x	5.13	2.6.x	N/A
¹ Big Data Management 和 Big Data Streaming 对 CDH 5.11 和 5.12 的支持需要 EBF-11719。请参阅知识库文章 533310 。					
² Big Data Management 对 MapR 6.x 与 MEP 5.0.x 的支持需要 EBF-12085。请参阅知识库文章 553273 。					
³ Big Data Management 和 Big Data Streaming 对 Amazon EMR 5.14 的支持需要 EBF-12444。请参阅 KB article 560632 。					

注意: Informatica 终止了对 IBM BigInsights 的支持。

要查看支持的版本的最新列表，请参阅 Informatica 客户门户上的“产品可用性列表”，网址为：
<https://network.informatica.com/community/informatica-network/product-availability-matrices>。

Big Data Management Hadoop 发行版

下表列出了 Big Data Management 10.2.1 中支持的 Hadoop 发行版版本和更改：

Hadoop 发行版	支持的发行版版本	10.2.1 中的更改
Amazon EMR	5.10、5.14	增加了对版本 5.10 和 5.14 的支持。 终止了对版本 5.8 的支持。
Azure HDInsight	3.6.x	增加了对版本 3.6 的支持。 终止了对版本 3.5.x 的支持。
Cloudera CDH	5.11、5.12、5.13、5.14、5.15	增加了对版本 5.13、5.14、5.15 的支持。
Hortonworks HDP	2.5.x、2.6.x	增加了对版本 2.6 的支持。 终止了对版本 2.4.x 的支持。
MapR	6.x MEP 5.0.x	增加了对版本 6.x MEP 5.0.x 的支持。 终止了对版本 5.2 MEP 2.0.x 和 5.2.MEP 3.0.x 的支持。

注意: Informatica 终止了对 IBM BigInsights 的支持。

Informatica 大数据产品支持各种 Hadoop 发行版。在每个版本中, Informatica 会增加、延期和终止对 Hadoop 发行版版本的支持。Informatica 可能会在将来的版本中恢复对延续版本的支持。

要查看支持的版本的最新列表, 请参阅 Informatica Network 上的"产品可用性列表", 网址为:
<https://network.informatica.com/community/informatica-network/product-availability-matrices>。

Big Data Streaming Hadoop 发行版

下表列出了 Big Data Streaming 10.2.1 中支持的 Hadoop 发行版版本和更改:

Hadoop 发行版	支持的发行版版本	10.2.1 中的更改
Amazon EMR	5.10、5.14	增加了对 5.10、5.14 的支持。 终止了对版本 5.4 的支持。
Azure HDInsight	3.6.x	增加了对版本 3.6 的支持。。
Cloudera CDH	5.11、5.12、5.13、5.14、5.15	增加了对版本 5.13、5.14、5.15 的支持。
Hortonworks HDP	2.5.x、2.6.x	增加了对版本 2.6 的支持。。 终止了对版本 2.4.x 的支持。
MapR	6.x MEP 4.0.x	增加了对版本 6.x MEP 4.0 的支持。 终止了对版本 5.2 MEP 2.0.x 和 5.2.MEP 3.0.x 的支持。

Informatica 大数据产品支持各种 Hadoop 发行版。在每个版本中, Informatica 会增加、延期和终止对 Hadoop 发行版版本的支持。Informatica 可能会在将来的版本中恢复对延续版本的支持。

要查看支持的版本的最新列表, 请参阅 Informatica Network 上的"产品可用性列表", 网址为:
<https://network.informatica.com/community/informatica-network/product-availability-matrices>。

Hive 运行时引擎

从版本 10.2.1 开始, Hive 运行时引擎的 MapReduce 模式已弃用, Informatica 将在将来的版本中停止对它的支持。Tez 模式仍受支持。

映射

如果您选择在 Hadoop 环境中运行映射, 默认情况下则会选择 Blaze 和 Spark 运行时引擎。

在以前的版本中, 还会选择 Hive 运行时引擎。

如果您选择 Hive 来运行映射, 数据集成服务将会使用 Tez。只有在以下 Hadoop 发行版上才能使用 Tez 引擎:

- Amazon EMR
- Azure HDInsight
- Hortonworks HDP

在将来的版本中, 当 Informatica 停止对 MapReduce 的支持时, 数据集成服务将忽略 Hive 引擎选择项, 并在 Blaze 或 Spark 上运行映射。

配置文件

从版本 10.2.1 开始, Hive 运行时引擎已弃用, Informatica 将在将来的版本中停止对它的支持。

Hive 选项在 Informatica Analyst、Informatica Developer 和 Catalog Administrator 中显示为“Hive (已弃用)”。您仍然可以选择在 Hive 引擎上运行配置文件。Informatica 建议您选择 Hadoop 选项以在 Blaze 引擎上运行配置文件。

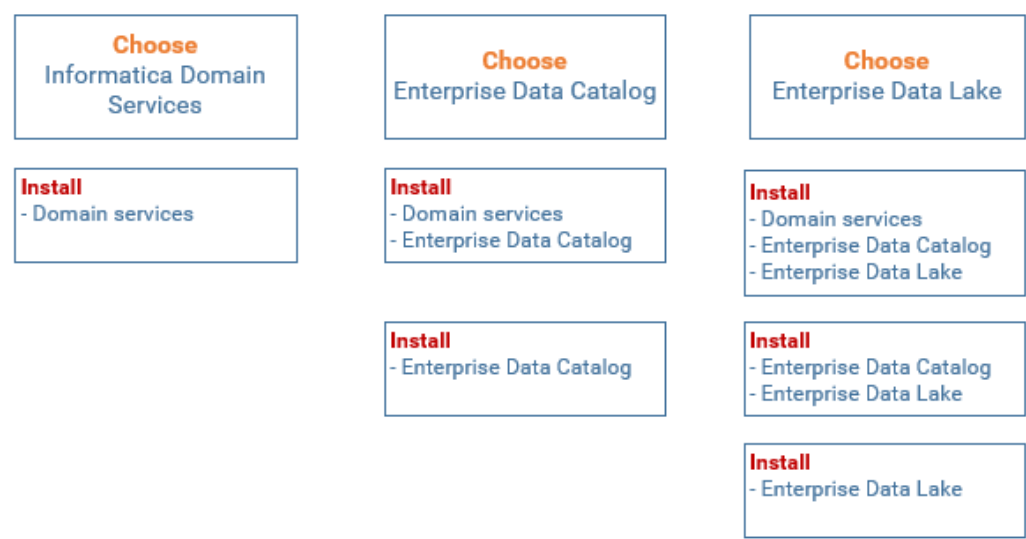
安装程序更改

从版本 10.2.1 开始，安装程序包含新功能，并已更新为包含所有大数据产品的安装和升级。Enterprise Data Catalog 安装和 Enterprise Data Lake 安装已与 Informatica 平台安装程序合并。

安装选项

运行安装程序时，可以选择符合要求的安装选项。

下图说明了版本 10.2.1 的安装选项和不同的安装程序任务：



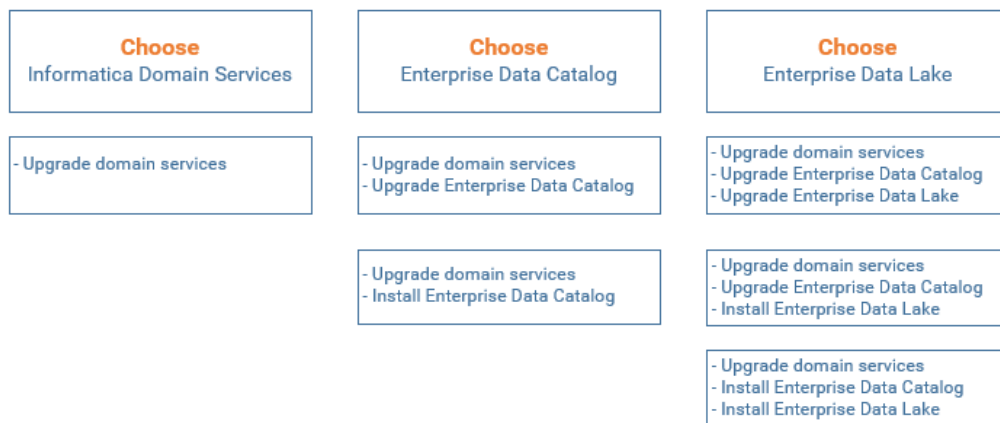
注意: 安装域服务时，安装程序还会安装应用程序服务二进制文件来支持 Big Data Management、Big Data Quality 和 Big Data Streaming。

升级选项

运行安装程序时，可以根据当前安装选择升级选项和操作。选择要升级的产品后，安装程序会根据需要升级父产品，并安装或升级所选产品。

例如，当选择 Enterprise Data Catalog 时，如果它正在运行以前的版本，则安装程序会升级域。如果已安装了 Enterprise Data Catalog，则安装程序会升级它。如果未安装 Enterprise Data Catalog，则安装程序会安装它。

下图说明了版本 10.2.1 的升级选项和不同的安装程序任务：



注意: 安装程序执行升级后，您需要在 Administrator 工具中完成部分应用程序服务的升级。

安装程序任务增强功能

统一安装程序已增强，可执行以下任务：

- 创建单独的监视模型存储库服务（当安装 Informatica 域服务时）。
- 根据 Big Data Management 部署规模调整数据集成服务和模型存储库服务。
- 创建 Enterprise Data Lake 所需的群集配置和关联连接。
- 为 Enterprise Data Lake 启用数据准备服务。

安装程序会限制传统产品

安装程序只包含大数据产品，不包含诸如 PowerCenter 和 Informatica Data Quality 等传统产品。传统产品和大数据产品位于不同的版本系列中。如果要进行升级，并且域包含传统产品和大数据产品，您必须在升级之前拆分域。

产品名称更改

本节介绍版本 10.2.1 中对产品名称进行的更改。

以下产品名称已更改：

- Intelligent Data Lake 已重命名为 Enterprise Data Lake。
- Intelligent Streaming 已重命名为 Big Data Streaming。
- Enterprise Information Catalog 已重命名为 Enterprise Data Catalog。

应用程序服务

本节介绍版本 10.2.1 中对应用程序服务进行的更改。

模型存储库服务

监视模型存储库服务

从版本 10.2.1 开始，可以将模型存储库服务配置为监视模型存储库服务，以监视 Ad Hoc 作业、应用程序、逻辑数据对象、SQL 数据服务、Web 服务和工作流的统计信息。配置监视模型存储库和模型存储库时应使用不同的数据库用户帐户。

在以前的版本中，可以使用模型存储库服务将设计时对象和运行时对象存储在模型存储库中。

有关详细信息，请参阅《*Informatica 10.2.1 应用程序服务指南*》中的“模型存储库服务”一章。

Big Data Management

本节介绍版本 10.2.1 中对 Big Data Management 进行的更改。

Azure 存储访问

从版本 10.2.1 开始，在 Azure HDInsight 群集上运行映射之前必须替代群集配置 core-site.xml 中的属性。WASB

如果您使用的群集将 WASB 用作存储，您可以从管理员处获取与 HDInsight 群集关联的存储帐户密钥，或者对加密的存储帐户密钥进行解密，然后在群集配置 core-site.xml 替代解密的值。

ADLS

如果您使用的群集将 ADLS 用作存储，您必须从 Web 应用程序复制客户端凭据，然后在群集配置 core-site.xml 中替代值。

在以前的版本中，您可以从 Hadoop 群集将文件复制到运行数据集成服务的计算机。

配置 Hadoop 发行版

本节介绍对 Hadoop 发行版配置进行的更改。

Hadoop 发行版配置

从版本 10.2.1 开始，可以在群集配置属性中配置 Hadoop 发行版。

从群集导入群集配置时，系统会填充“发行版名称”属性和“发行版版本”属性。可以在完成导入过程后编辑发行版版本。

在以前的版本中，Hadoop 发行版由承载数据集成服务的计算机上的发行版目录路径标识。

从版本 10.2.1 开始，以下属性已从数据集成服务属性中删除：

- 数据集成服务 Hadoop 分发版目录

有关“发行版名称”属性和“发行版版本”属性的详细信息，请参阅《*Big Data Management 10.2.1 管理指南*》。

MapR 配置

从版本 10.2.1 开始，配合使用 Big Data Management 和 MapR 时，不再需要为域配置数据集成服务进程属性。Big Data Management 支持 Kerberos 身份验证，无需用户操作。

以前，需要在“数据集成服务”自定义属性中配置 JVM 选项属性以及配置环境变量才能启用对 Kerberos 身份验证的支持。

有关将域与 MapR 群集集成的详细信息，请参阅《*Big Data Management 10.2.1 Hadoop 集成指南*》。

Developer tool 配置

从版本 10.2.1 开始，可以创建元数据访问服务。元数据访问服务是一个应用程序服务，使 Developer tool 可以访问 Hadoop 连接信息以导入和预览元数据。从 Hadoop 群集导入对象时，以下适配器在设计时使用元数据访问服务来提取对象元数据：

- PowerExchange for HBase
- PowerExchange for HDFS
- PowerExchange for Hive
- PowerExchange for MapR-DB

在以前的版本中，您需要在每台 Developer tool 计算机上手动执行以下步骤，以便在设计时在 Developer tool 计算机和 Hadoop 群集之间建立通信：

- 提取群集配置文件。
- 运行 krb5.ini 文件以从启用 Kerberos 的 Hadoop 群集中的 Hive、HBase 和复杂文件源导入元数据。

借助元数据访问服务，您无需为每台 Developer tool 计算机配置与 Hadoop 群集的设计时连接。

有关详细信息，请参阅《*Informatica 10.2.1 应用程序服务指南*》中的“元数据访问服务”一章。

Hadoop 连接更改

从版本 10.2.1 开始，Hadoop 连接包含新的不同属性和功能。其中包括以前可以在其他连接或配置文件中配置的几个属性，以及其他一些更改。

本节列出了版本 10.2.1 中对 Hadoop 连接进行的更改。

已从 `hadoopEnv.properties` 移到 Hadoop 连接的属性

从版本 10.2.1 开始，以前在 `hadoopEnv.properties` 文件中配置的属性现在可以在 Hadoop 连接的高级属性中进行配置。

有关 Hive 连接和 Hadoop 连接的详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》。有关配置 Big Data Management 的详细信息，请参阅《*Informatica Big Data Management 10.2.1 Hadoop 集成指南*》。

已从 Hive 连接移到 Hadoop 连接的属性

在 Hadoop 群集上运行映射所需的以下 Hive 连接属性现在位于 Hadoop 连接中：

- 数据库名称。表的命名空间。对于未指定数据库名称的表，请使用名称 default。
- 高级 Hive/Hadoop 属性。配置或替代运行数据集成服务的计算机上 `hive-site.xml` 配置集中的 Hive 或 Hadoop 群集属性。可以指定多个属性。
- 临时表压缩编解码器。压缩编解码器类名的 Hadoop 压缩库。
- 编解码器类名。可在临时暂存表上启用数据压缩并提高性能的编解码器类名。

在以前的版本中，您可以在 Hive 连接中配置这些属性。

有关 Hive 连接和 Hadoop 连接的详细信息，请参阅《*Informatica Big Data Management 10.2.1 管理员指南*》。

Hadoop 运行时引擎的高级属性

从版本 10.2.1 开始，可以在 Hadoop 连接属性中为 Blaze、Spark 和 Hive 运行时引擎配置高级属性。

Informatica 已将运行时引擎的相关属性的属性名称标准化。下表显示了旧名称和新名称：

版本 10.2.1 之前的属性名称	10.2.1 Hadoop 连接属性部分	10.2.1 属性名称
Blaze 服务自定义属性	Blaze 配置	高级属性
Spark 执行参数	Spark 配置	高级属性
Hive 自定义属性	Hive 下推配置	高级属性

在以前的版本中，可以在 `hadoopRes.properties` 或 `hadoopEnv.properties` 文件中为运行时引擎配置高级属性，或者也可以在 Administrator 工具的**共同属性**下的"Hadoop 引擎自定义属性"字段中进行配置。

Blaze 引擎的附加属性

从版本 10.2.1 开始，可以在 Hadoop 连接属性的"Blaze 配置属性"部分中配置一个附加属性。

下表介绍了该属性：

属性	说明
Blaze YARN 节点标签	用于确定 Blaze 引擎在其上运行的 Hadoop 群集节点的节点标签。如果未指定节点标签，Blaze 引擎将在默认分区中的节点上运行。 如果 Hadoop 群集支持对节点标签使用逻辑运算符，则您可以指定节点标签列表。要列出节点标签，可以使用 && (AND)、 (OR) 和 ! (NOT)。

有关在 Blaze 引擎上使用节点标签的详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》中的"Hadoop 环境中的映射"一章。

Hive 连接属性

从版本 10.2.1 开始，Hive 连接的属性已更改。

以下 Hive 连接属性已删除：

- 访问作为源或目标的 Hive
- 使用 Hive 在 Hadoop 群集中运行映射

在以前的版本中，这些属性已弃用。从版本 10.2.1 开始，这些属性已过时。

请在 Hadoop 连接中配置以下 Hive 连接属性：

- 数据库名称
- 高级 Hive/Hadoop 属性
- 临时表压缩编解码器
- 编解码器类名

在以前的版本中，您可以在 Hive 连接中配置这些属性。

有关 Hive 连接和 Hadoop 连接的详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》。

监视

本节介绍版本 10.2.1 中对 Big Data Management 监视进行的更改。

Spark 监视

从版本 10.2.1 开始，Spark 监视更改与以下方面相关：

- 事件更改
- "摘要统计信息"视图中的更新

事件更改

从版本 10.2.1 开始，只有监视信息会签入到会话日志中的 Spark 事件。

在以前的版本中，所有 Spark 事件均按原样从 Spark 应用程序中继到 Spark 执行程序。如果中继的事件耗费很长时间，则会出现性能问题。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》。

"摘要统计信息"视图

从版本 10.2.1 开始，可以根据运行阶段查看 Spark 执行的统计信息。例如，"Spark 运行阶段"显示 Spark 应用程序运行阶段的统计信息。Stage_0 显示 Spark 应用程序中 ID 为 0 的运行阶段的相关统计信息。"行数"和"每秒平均行数"分别显示该阶段写入的行数和相应的吞吐量。"字节数"和"每秒平均字节数"分别显示该阶段播发的字节数和吞吐量。

在以前的版本中，您只能查看 Spark 运行的源行和目标行以及每秒处理的平均行数。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》。

Hive 引擎上的精度和小数位

从版本 10.2.1 开始，如果以下条件属实，则用户定义的在 Hive 引擎上执行乘法的函数的输出具有最大小数位数 6：

- 精度和小数位数之差大于等于 32。
- 结果精度大于 38。

以前，小数位数可能低至 0。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》中的“Hadoop 环境中的映射”一章。

Sqoop

从版本 10.2.1 开始，对 Sqoop 应用了以下更改：

- 在 Spark 引擎上运行 Sqoop 映射时，数据集成服务会将 Sqoop 日志事件输出到映射日志。在以前的版本中，数据集成服务将 Sqoop 日志事件输出到 Hadoop 群集日志。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.1 用户指南*》。

- 如果将 Sqoop 连接所需的 Type 4 JDBC 驱动程序 .jar 文件添加到 externaljdbcjars 目录中或从中删除该文件，所做更改将在重新启动数据集成服务之后生效。如果在 Blaze 引擎上运行映射，所做更改将在重新启动数据集成服务和 Blaze 网格管理器之后生效。

注意: 首次运行映射时，无需重新启动数据集成服务和 Blaze 网格管理器之后生效。只有在后续运行映射时才需要重新启动数据集成服务和 Blaze 网格管理器之后生效。

在以前的版本中，不必在添加或删除 Sqoop .jar 文件之后重新启动数据集成服务和 Blaze 网格管理器。

有关详细信息，请参阅《*Informatica Big Data Management 10.2.1 Hadoop 集成指南*》。

Hive 引擎上的转换支持

从版本 10.2.1 开始，执行概率分析的标签创建器转换或解析器转换要求运行转换的任何节点都具有 Java 8 Development Kit。

在以前的版本中，这些转换需要 Java 7 Development Kit。

如果要运行的映射包含为概率分析配置的标签创建器转换或解析器转换，请验证 Hive 节点上的 Java 版本。

注意: 在 Blaze 或 Spark 节点上，数据集成服务使用随 Informatica 引擎一起安装的 Java Development Kit。Java 8 Development Kit 随 Informatica 10.2.1 一起安装。

有关详细信息，请参阅适用于要升级的 Informatica 版本的《*Informatica 10.2.1 安装指南*》或《*Informatica 10.2.1 升级指南*》。

Big DataStreaming

本节介绍版本 10.2.1 中对 Big Data Streaming 进行的更改。

配置 Hadoop 发行版

从版本 10.2.1 开始，可以在群集配置属性中配置 Hadoop 发行版。

从群集导入群集配置时，系统会填充"发行版名称"属性和"发行版版本"属性。可以在完成导入过程后编辑发行版本。

在以前的版本中，Hadoop 发行版由承载数据集成服务的计算机上的发行版目录路径标识。

有关"发行版名称"属性和"发行版版本"属性的详细信息，请参阅《*Informatica Big Data Management 10.2.1 管理指南*》。

Developer tool 配置

从版本 10.2.1 开始，可以创建元数据访问服务。元数据访问服务是一个应用程序服务，使 Developer tool 可以访问 Hadoop 连接信息以导入和预览元数据。

以下源和目标在设计时使用元数据访问服务来提取元数据：

- HBase
- HDFS
- Hive
- MapR-DB

- MapRStreams

在以前的版本中，您需要在每台 Developer tool 客户端计算机上手动执行以下步骤，以便在设计时在 Developer tool 计算机和 Hadoop 群集之间建立通信：

- 提取群集配置文件。
- 运行 krb5.ini 文件以从启用 Kerberos 的 Hadoop 群集中的 Hive、HBase 和复杂文件源导入元数据。

借助元数据访问服务，您无需为每台 Developer tool 计算机配置与 Hadoop 群集的设计时连接。

有关详细信息，请参阅《Informatica 10.2.1 应用程序服务指南》中的"元数据访问服务"一章。

Kafka 连接属性

从版本 10.2.1 开始，Kafka 连接的属性已更改。

现在可以在连接属性中配置 Kafka 代理版本。

在以前的版本中，可以在 hadoopEnv.properties 文件和 hadoopRes.properties 文件中配置该属性。

有关 Kafka 连接的详细信息，请参阅《Informatica Big Data Streaming 10.2.1 用户指南》中的"连接"一章。

命令行程序

本节介绍版本 10.2.1 中对命令进行的更改。

infacmd ihs 命令

已更改命令

下表介绍了已更改的 infacmd ihs 命令：

命令	说明
createservice	从版本 10.2.1 开始，createservice 命令添加了 -kc 选项。
createservice	从版本 10.2.1 开始，createservice 命令添加了 -bn 选项。

infacmd ldm 命令

已更改命令

下表介绍了已更改的 infacmd ldm 命令：

命令	更改说明
CreateService	从版本 10.2.1 开始，CreateService 命令添加了 -lt 选项。
CreateService	从版本 10.2.1 开始，CreateService 命令删除了 -dis 选项。
CreateService	从版本 10.2.1 开始，CreateService 命令删除了 -cms 选项。

有关详细信息，请参阅《*Informatica 10.2.1 命令参考*》。

内容安装程序

从版本 10.2.1 开始，Informatica 不再为加速器文件和引用数据文件提供内容安装程序实用程序。要将加速器文件或引用数据文件添加到 Informatica 安装中，请提取文件并将其复制到安装中的适当目录。

在以前的版本中，可以使用内容安装程序来提取文件并将其复制到 Informatica 目录。

有关详细信息，请参阅《*Informatica 10.2.1 内容指南*》。

Enterprise Data Catalog

本节介绍版本 10.2.1 中对 Informatica Enterprise Data Catalog 进行的更改。

常规选项卡中的附加属性部分

从版本 10.2.1 开始，在创建资源时，可以在"常规"选项卡的**附加属性**部分中为资源分配自定义属性值。可分配的自定义属性值包括 Department、Data Owner、Data Steward 和 Subject Matter Experts。

有关分配自定义属性的详细信息，请参阅《*Informatica 10.2.1 Catalog Administrator 指南*》和《*Informatica 10.2.1 Enterprise Data Catalog 用户指南*》。

连接分配

从版本 10.2.1 开始，可以将数据库分配给 PowerCenter 资源的连接。

有关连接分配的详细信息，请参阅《*Informatica 10.2.1 Catalog Administrator 指南*》。

列相似性

从版本 10.2.1 开始，可以基于列名称、列模式、唯一值和值频率在资源中发现类似列。

在以前的版本中，相似性发现系统在源数据中确定类似列。

有关列相似性的详细信息，请参阅《*Informatica 10.2.1 Catalog Administrator 指南*》。

创建目录服务

从版本 10.2.1 开始，在创建目录服务时，您不必提供要与目录服务集成的数据集成服务和内容管理服务的详细信息。

有关详细信息，请参阅《*Informatica Enterprise Data Catalog 10.2.1 安装和配置指南*》。

HDFS 资源类型增强功能

从版本 10.2.1 开始，可以对 HDFS 资源使用以下 Hadoop 发行版类型之一：

- Hortonworks

- IBM BigInsights
- Azure HDInsight
- Amazon EMR
- MapR FS

Hive 资源

从版本 10.2.1 开始，如果您在创建 Hive 资源时为**运行于**选项选择 Hive，则需要选择一个 Hadoop 连接才能在 Hive 引擎上运行剖析扫描程序。

在以前的版本中，不需要 Hadoop 连接即可对 Hive 资源运行剖析扫描程序。

有关 Hive 资源的详细信息，请参阅《*Informatica 10.2.1 Catalog Administrator 指南*》。

Informatica 平台扫描程序

从版本 10.2.1 开始，可以使用参数文件和参数集通过 Informatica 平台扫描程序提取详细的沿袭。

概览选项卡

从版本 10.2.1 开始，"资产详细信息"视图在 Enterprise Data Catalog 中的标题为"概览"。

现在可以在"概览"选项卡中查看资产的详细信息。"概览"选项卡中显示多个不同部分，例如源说明、说明、人员、业务术语、业务分类、系统属性及其他属性。"概览"选项卡中显示的各个部分取决于资产类型。

有关资产概览的详细信息，请参阅《*Informatica Enterprise Data Catalog 10.2.1 用户指南*》中的"查看资产"一章。

产品名称更改

从版本 10.2.1 开始，Enterprise Data Catalog 包括以下名称更改：

- 产品名称已更改为 Informatica Enterprise Data Catalog。在以前的版本中，产品名称为 Enterprise Information Catalog。
- 安装程序名称已更改为 Enterprise Data Catalog。在以前的版本中，安装程序名称为 Enterprise Information Catalog。

邻近数据域

从版本 10.2.1 开始，在创建或编辑具有数据规则或列规则的数据域时，可以添加一个或多个数据域作为邻近数据域。剖析扫描程序将扫描数据源以发现资源中的数据域和邻近数据域，并在 Enterprise Data Catalog 中显示匹配得分。匹配得分是数据源中发现的邻近数据域个数与已为推理数据域配置的邻近数据域个数之比。

在以前的版本中，可以向具有数据规则的数据域添加邻近规则。如果在源表中找不到数据域，则数据域的数据遵从性百分比在源表中会下降指定的百分比。

有关邻近数据域的详细信息，请参阅《*Informatica 10.2.1 Catalog Administrator 指南*》。

搜索结果

从版本 10.2.1 开始，搜索结果页面包括以下更改：

- 现在可以基于资产名称和相关性对搜索结果进行排序。在以前的版本中，可以基于资产名称、相关性、系统属性和自定义属性对搜索结果进行排序。
- 现在可以向搜索结果中的资产添加业务标题。在以前的版本中，只能关联业务术语。
- 搜索结果页面现在显示资产的资产详细信息，例如资源名称、源说明、说明、资产路径和资产类型。在以前的版本中，可以查看诸如资产类型、资源类型、资产的上次更新日期和资产大小等详细信息。

有关搜索结果的详细信息，请参阅《*Informatica Enterprise Data Catalog 10.2.1 用户指南*》。

通用连接框架

从版本 10.2.1 开始，所有使用通用连接框架创建的资源都要求 Catalog Agent 已启动且正常运行。

在以前的版本中，只有在 Microsoft Windows 上运行的资源才要求 Catalog Agent 已启动且正常运行。

有关详细信息，请参阅《*Informatica 10.2.1 Catalog Administrator 指南*》。

Informatica Analyst

本节介绍版本 10.2.1 中对 Analyst 工具的更改。

结果卡

本节介绍版本 10.2.1 中对结果卡行为进行的更改。

编辑结果卡中的现有度量

从版本 10.2.1 开始，向现有结果卡添加列时无法编辑现有度量或度量组。要修改结果卡中的现有度量或度量组，请导航到“结果卡”工作区，编辑该结果卡，然后修改度量。

在以前的版本中，向现有结果卡添加列时可以查看和编辑现有度量或度量组。

有关结果卡的详细信息，请参阅《*Informatica 10.2.1 Data Discovery 指南*》。

配置度量的阈值

从版本 10.2.1 开始，可以将最多两个小数位数的十进制数字配置为结果卡中的度量的阈值。

在以前的版本中，只能将整数值配置为度量的阈值。

有关结果卡的详细信息，请参阅《*Informatica 10.2.1 Data Discovery 指南*》。

Informatica Developer

本节介绍版本 10.2.1 中对 Developer tool 的更改。

从 PowerCenter 导入对象以及将对象导出到其中

从版本 10.2.1 开始，Developer tool 不包含从 PowerCenter 导入对象的选项以及将对象导出到其中的选项。

Informatica 转换

本节介绍版本 10.2.1 中对 Informatica 转换进行的更改。

地址验证器转换

本节介绍版本 10.2.1 中对地址验证器转换进行的更改。

地址验证器转换对地址功能做了以下更新：

所有国家/地区

从版本 10.2.1 开始，地址验证器转换使用 5.12.0 版本的 Informatica Address Verification 软件引擎。该引擎支持 Informatica 在版本 10.2.1 中向地址验证器转换添加的功能。

以前，该转换使用 5.11.0 版本的 Informatica Address Verification 软件引擎。

英国

从 2017 年 11 月开始，Informatica 停止交付包含英国企业名称和地址的引用数据文件。Informatica 停止支持对业务名称和地址的验证。

有关 Informatica 嵌入到版本 10.2.1 的地址验证软件引擎版本的功能和操作的全面信息，请参阅《*Informatica Address Verification 5.12.0 开发人员指南*》。

数据转换

本节介绍版本 10.2.1 中对数据处理器转换进行的更改。

从版本 10.2.1 开始，数据处理器转换对层次结构输入执行严格验证。应用严格验证时，层次结构输入文件必须严格遵守其架构。当数据处理器模式设置为"输出映射"时，可以应用此选项，该模式将为关系输出创建输出端口。

此选项不适用于具有来自低于 10.2.1 版本的 JSON 输入的映射。

有关详细信息，请参阅《*Data Transformation 10.2.1 用户指南*》。

序列生成器转换

本节介绍版本 10.2.1 中对序列生成器转换进行的更改。

保持行顺序

从版本 10.2.1 开始，序列生成器转换的"保持行顺序"属性默认设置为 False。

在以前的版本中，默认值为 True。

从早期版本升级时，存储库中任何序列生成器转换的"保持行顺序"属性不会更改。

有关详细信息，请参阅《*Informatica 10.2.1 Developer 转换指南*》中的"序列生成器转换"一章。

排序器转换

本节介绍版本 10.2.1 中对排序器转换的更改。

排序器缓存

从版本 10.2.1 开始，排序器转换的排序器缓存使用可变长度在本地环境和 Hadoop 环境中的 Blaze 引擎上最多存储 8 MB 数据。

在以前的版本中，排序器缓存使用可变长度最多存储 64 KB 数据。对于超出 64 KB 的数据，排序器缓存使用固定长度存储该数据。

有关详细信息，请参阅《Informatica 10.2.1 Developer 转换指南》中的"排序器转换"一章。

排序器性能

从版本 10.2.1 开始，排序器转换已优化，可以更快对最多 8 MB 的数据执行排序键比较。

在下列情况下，排序键比较速度未优化：

- 未选择二进制排序顺序。
- 排序键是数据类型为 time zone 的时间戳。
- 您执行区分大小写的字符串比较，并且任意排序键列为 string 数据类型。

在以前的版本中，排序器转换可以更快对最多 64 KB 的数据执行排序键比较。对于超出 64 KB 的数据，排序键比较速度未优化。

有关详细信息，请参阅《Informatica 10.2.1 Developer 转换指南》中的"排序器转换"一章。

Informatica PowerExchange 适配器

本节介绍版本 10.2.1 中对 Informatica 适配器的更改。

PowerExchange for Amazon Redshift

从版本 10.2.1 开始，连接到 PowerExchange for Amazon Redshift 后，系统会自动完成以下先决任务：

- 下载所需的 Amazon Redshift JDBC .jar 文件。
- 将该 .jar 文件复制到运行数据集成服务的节点上和客户端计算机上。

在以前的版本中，您必须先手动执行先决任务并重新启动数据集成服务，然后才能使用 PowerExchange for Amazon Redshift。

PowerExchange for Cassandra

从版本 10.2.1 开始，PowerExchange for Cassandra 具有以下更改：

- Informatica PowerExchange for Cassandra ODBC 驱动程序文件的名称和目录已更改。

下表根据 Linux 和 Windows 操作系统列出了 Cassandra ODBC 驱动程序文件名称与目录：

操作系统	Cassandra ODBC 驱动程序文件名称	文件目录
Linux	libcassandraodbc_sb64.so	<Informatica 安装目录>\tools\cassandra\lib\libcassandraodbc_sb64.so
Windows	CassandraODBC_sb64.dll	<Informatica 安装目录>\tools\cassandra\lib\CassandraODBC_sb64.dll

在 Linux 操作系统上，必须为 odbc.ini 文件中的现有 Cassandra 数据源将 Driver 属性的值更新为 <Informatica 安装目录>\tools\cassandra\lib\libcassandraodbc_sb64.so。

在 Windows 上，必须在 Windows 注册表中为现有 Cassandra 数据源名称更新以下属性：

```
Driver=<Informatica installation directory>\tools\cassandra\lib\CassandraODBC_sb64.dll
Setup=<Informatica installation directory>\tools\cassandra\lib\CassandraODBC_sb64.dll
```

- "负载均衡策略"选项新的项名称为 LoadBalancingPolicy。
在以前的版本中，"负载均衡策略"选项的项名称为 COLoadBalancingPolicy。
- 以下 Cassandra ODBC 驱动程序属性的默认值已更改：

驱动程序属性名称	项名称	新的默认值
并发请求数	NumConcurrentRequests	100
插入查询线程	NumInsertQueryThreads	2
每个插入线程迭代次数	NumIterationsPerInsertThread	50

有关详细信息，请参阅《*Informatica PowerExchange for Cassandra 10.2.1 用户指南*》。

PowerExchange for Snowflake

从版本 10.2.1 开始，PowerExchange for Snowflake 随 Informatica 10.2.1 一起安装。

在以前的版本中，PowerExchange for Snowflake 有单独的安装程序。

有关详细信息，请参阅《*Informatica PowerExchange for Snowflake 10.2.1 用户指南*》。

第 21 章

发布任务 (10.2.1)

本章包括以下主题：

- [Informatica PowerExchange 适配器, 218](#)

Informatica PowerExchange 适配器

本节介绍版本 10.2.1 中 Informatica 适配器的发布任务。

适用于 Amazon S3 的 PowerExchange 适配器

从版本 10.2.1 开始，要成功预览 Avro 和 Parquet 文件中的数据或在本地环境中使用 Avro 和 Parquet 文件运行映射，必须在 Informatica Administrator 中为数据集成服务配置 INFA_PARSER_HOME 属性。要配置 INFA_PARSER_HOME 属性，请执行以下步骤：

- 登录到 Informatica Administrator。
- 单击数据集成服务，然后单击右窗格上的**进程**选项卡。
- 单击**环境变量**部分中的**编辑**。
- 单击**新建**以添加一个环境变量。
- 输入 **INFA_PARSER_HOME** 作为该环境变量的名称。
- 将该环境变量的值设置为运行数据集成服务的计算机上 Hadoop 发行版目录的绝对路径。验证 INFA_PARSER_HOME 属性中定义的 Hadoop 发行版目录的版本是否与群集配置中定义的版本相同。

有关详细信息，请参阅《*Informatica PowerExchange for Amazon S3 10.2.1 用户指南*》。

第 V 部分： 版本 10.2

本部分包含以下章节：

- [新功能、更改和发布任务 \(10.2 HotFix 2\), 220](#)
- [新功能、更改和发布任务 \(10.2 HotFix 1\), 231](#)
- [新产品 \(10.2\), 249](#)
- [新增功能 \(10.2\), 250](#)
- [更改 \(10.2\), 282](#)
- [发布任务 \(10.2\), 297](#)

第 22 章

新功能、更改和发布任务 (10.2 HotFix 2)

本章包括以下主题：

- [支持更改\(10.2 HotFix 2\), 220](#)
- [新产品\(10.2 HotFix 2\), 221](#)
- [新增功能 \(10.2 HotFix 2\), 221](#)
- [更改 \(10.2 HotFix 2\), 227](#)
- [发布任务\(10.2 HotFix 2\), 230](#)

支持更改(10.2 HotFix 2)

本节介绍版本 10.2 HotFix 2 中的支持更改。

验证 Hadoop 发行版支持

验证 Hadoop 环境中 Hadoop 发行版的版本。

Big Data Management、Big Data Streaming、Big Data Quality 和 PowerCenter 支持以下 Hadoop 发行版：

- Amazon EMR
- Azure HDInsight
- Cloudera CDH
- Hortonworks HDP
- MapR

在每个版本中，Informatica 都可以添加、延迟和删除对非本地发行版和各发行版版本的支持。Informatica 可能会在将来的版本中恢复对延期版本的支持。要查看支持的版本的最新列表，请参阅 Informatica 客户门户上的“产品可用性列表”，网址为：

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

OpenJDK

从版本 10.2 HotFix 2 开始，Informatica 安装程序会封装 OpenJDK (AzulJDK)。支持的 Java 版本为 Azul OpenJDK 1.8.192。

Spark 引擎上的 Sqoop 映射不使用与 Informatica 安装程序打包在一起的 Azul OpenJDK。Spark 和 Blaze 引擎上的 Sqoop 映射继续使用 `hadoopEnv.properties` 文件中的 `infapdo.env.entry.hadoop_node_jdk_home` 属性。`HADOOP_NODE_JDK_HOME` 代表您运行群集服务的目录和群集节点使用的 JDK 版本。

在以前的版本中，安装程序使用的是与自身打包在一起的 Oracle Java。

DataDirect SQL Server 旧版 ODBC 驱动程序

从版本 10.2 HotFix 2 开始，Informatica 不再支持 DataDirect SQL Server 旧版 ODBC 驱动程序文件 `DWmss27.x`，因为 DataDirect 不再支持此驱动程序。

当您使用 ODBC 连接以连接到 Microsoft SQL Server 时，可以使用与 Informatica 安装程序一起打包的 DataDirect 8.0 SQL Server Wire Protocol，也可以使用来自第三方供应商的任何 ODBC 驱动程序。

PowerExchange for SAP NetWeaver

从版本 10.2 HotFix 2 开始，PowerExchange for SAP NetWeaver 支持适用于 PowerCenter 的 SAP NetWeaver RFC SDK 7.50 库。

使用 SAP NetWeaver RFC SDK 7.20 库的现有映射不会失败。但是，Informatica 建议您下载并安装 SAP NetWeaver RFC SDK 7.50 库。

新产品(10.2 HotFix 2)

PowerCenter PowerExchange 适配器

本节介绍版本 10.2 HotFix 2 中新增的 PowerCenter 适配器。

PowerExchange for Tableau V3

从版本 10.2 HotFix 2 开始，您可以使用 Tableau V3 连接从多个源读取数据，生成 Tableau .hyper 输出文件，以及向 Tableau 中写入数据。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Tableau V3 10.2 HotFix 2 用户指南》。

新增功能 (10.2 HotFix 2)

命令行程序

本节介绍版本 10.2 HotFix 2 中新增的命令。

pmrep 命令

创建或更新 ODBC 连接时，可以将 pmrep createconnection 和 pmrep updateconnection 命令选项 -S 与参数 odbc_subtype 配合使用以启用 ODBC 子类型选项。

下表描述了新的 pmrep createconnection 和 pmrep updateconnection 命令选项：

选项	参数	说明
-S	odbc_subtype	可选。为 ODBC 连接启用 ODBC 子类型。 ODBC 子类型可以是以下类型之一： - AWS Redshift - Azure DW - Greenplum - Google Big Query - PostgreSQL - SAP HANA - Snowflake - 无 默认值为“无”。

有关详细信息，请参阅《Informatica 10.2 HotFix 2 命令引用》中的“pmrep 命令引用”一章。

Informatica Analyst

本节介绍版本 10.2 HotFix 2 中新增的 Analyst 工具功能。

结果卡

从版本 10.2 HotFix 2 开始，您可以输入要从结果卡趋势图表中导出的无效行数。

有关详细信息，请参阅《Informatica 10.2 HotFix2 Data Discovery 指南》中的“Informatica Analyst 中的结果卡”一章。

Informatica 转换

本节介绍版本 10.2 HotFix 2 中新增的 Informatica 转换功能。

地址验证器转换

对于以下国家/地区，地址验证器转换具有附加的地址功能：

澳大利亚

从版本 10.2 HotFix 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。

以色列

从版本 10.2 HotFix 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。

支持在以色列地址中使用多个字符集

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。

美国

从版本 10.2 HotFix 2 开始，Informatica 引入了以下适用于美国地址的功能和增强功能：

有关无法投递邮件的其他信息

您可以配置地址验证器转换，以返回有关为什么美国邮政署未将邮件投递到看起来有效地址的信息。

可能由于以下原因之一，美国邮政署未将邮件投递到某个地址：

- 地址所在地正在建筑施工。
- 客户通过邮件转寄服务收到了邮件。
- 邮政运营商返回未投递的邮件。

美国邮政署设有一份不允许投递的地址表。该表称为“无统计信息”表。要返回指示美国邮政署为何将地址添加到“无统计信息”表的代码，请选择“交付序列文件，第二代无统计信息原因”端口。在“基本”模型的“特定于美国”端口组中查找该端口。

该转换会读取 USA5C131.MD 数据库文件中的“无统计信息”表数据。

改进对具有尾随门牌号字符的地址的更正

您可以配置地址验证器转换，以识别门牌号值中有无法识别的尾随字母字符的有效街道地址。如果地址可以提供没有尾随字符的有效收件人地点，则该转换将返回“收件人地点验证脚注”端口中的值 TA。

有关将邮件转寄到邮政信箱的街道地址的信息

您可以配置地址验证器转换，以识别美国邮政署将邮件转寄到邮政信箱的街道地址。要识别此类地址，请选择“收件人地点验证退回”端口。在“基本”模型的“特定于美国”端口组中查找该端口。

该转换会读取 USA5C132.MD 数据库文件中的退回表数据。

关于在一周内的一天或多天无法接收邮件的地址的信息

您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。

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。

有关详细信息，请参阅《*Informatica 10.2 HotFix 2 Developer 转换指南*》和《*Informatica 10.2 HotFix 2 地址验证器端口参考*》。

有关版本 10.2 HotFix 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*》。

Metadata Manager

本节介绍版本 10.2 HotFix 2 中新增的 Metadata Manager 功能。

Cognos

从版本 10.2 HotFix 2 开始，可以为 Cognos 资源设置以下配置属性：

- Cognos SDK 目录。输入 Cognos SDK 目录位置。
- 其他。指定以逗号分隔的多个其他选项。

有关详细信息，请参阅《*Informatica 10.2 HotFix 2 Metadata Manager 管理员指南*》中的“业务情报资源”一章。

Microstrategy

从版本 10.2 HotFix 2 开始，可以为 Microstrategy 资源设置 **其他** 配置属性：使用 **其他** 配置属性可以指定以逗号分隔的多个其他选项。

有关详细信息，请参阅《*Informatica 10.2 HotFix 2 Metadata Manager 管理员指南*》中的“业务情报资源”一章。

PowerCenter

本节介绍版本 10.2 HotFix 2 中新增的 PowerCenter 功能。

下推优化适用于 PostgreSQL

从版本 10.2 HotFix 2 开始，如果连接类型为 ODBC，则可选择 ODBC 子类型作为 **PostgreSQL**，将转换逻辑推送到 PostgreSQL。您可以配置源端或完整下推优化，将转换逻辑推送到 PostgreSQL。

有关受支持的函数以及可以推送到 PostgreSQL 数据库的转换，请参阅《*Informatica PowerCenter 10.2 HotFix 2 高级工作流指南*》。

PowerCenter 二进制函数

从版本 10.2 HotFix 2 开始，可以为二进制表达式使用以下函数：

- EBCDIC_ISO88591。将以 EBCDIC 编码的二进制值转换为以 ISO-8859-1 编码的字符串值。
- BINARY_COMPARE。比较两个二进制值，如果相同，则返回 TRUE (1)；如果不同，则返回 FALSE (0)。
- BINARY_CONCAT。将两个或多个二进制值连接在一起，并返回连接后的值。
- BINARY_LENGTH。返回二进制值的长度。
- BINARY_SECTION。返回二进制值的一部分。
- DEC_HEX。解码十六进制编码值，并返回具有数据二进制表示形式的二进制值。
- ENC_HEX。使用十六进制编码将二进制数据编码为字符串数据。
- SHA256。计算输入值的 SHA-256 摘要。

有关自定义函数的详细信息，请参阅《*Informatica PowerCenter 10.2 HotFix 2 转换语言参考*》。

PowerCenter PowerExchange 适配器

本节介绍版本 10.2 HotFix 2 中新增加的 PowerCenter 适配器功能。

PowerExchange for Amazon Redshift

从版本 10.2 HotFix 2 开始，从 Amazon Redshift 读取数据时，可以保留空值。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica PowerExchange for Amazon Redshift 10.2 HotFix 2 用户指南*》。

PowerExchange for Google BigQuery

从版本 10.2 HotFix 2 开始，PowerExchange for Google BigQuery 包含以下新功能：

- 您可以配置自定义 SQL 查询以配置 Google BigQuery 源。
- 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。
- 您可以指定引号字符，用于定义 .csv 文件中文本字符串的边界。您可以配置参数，例如单引号或双引号。
- 在使用 ODBC 连接类型时，您可以配置完整下推优化，将转换逻辑推送到 Google BigQuery。
有关 PowerCenter 集成服务可以推送到 Google BigQuery 的运算符和函数的信息，请参阅《*Informatica PowerCenter 10.2 HotFix 2 高级工作流指南*》。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica PowerExchange for Google BigQuery 10.2 HotFix 2 用户指南*》。

PowerExchange for Kafka

从版本 10.2 HotFix 2 开始，PowerExchange for Kafka 包含以下新功能：

- 您可以导入 Avro 格式的 Kafka 源或目标定义。
- 您可以配置 SSL 配置属性以安全连接到 Kafka。

- 您可以预览 Kafka 源或目标定义的数据。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Kafka 10.2 HotFix 2 用户指南》。

PowerExchange for Microsoft Azure SQL Data Warehouse V3

从版本 10.2 HotFix 2 开始，PowerExchange for Microsoft Azure SQL Data Warehouse V3 包含以下新功能：

- 读取源对象时可以使用自定义查询。
- 可以替代源会话属性中的源对象和源对象架构。在源会话属性中定义的源对象和源对象架构优先。
- 可以替代目标会话属性中的目标对象和目标对象架构。在目标会话属性中定义的目标对象和目标对象架构优先。
- 您可以创建映射，以从更改数据捕获 (CDC) 源读取实时数据或已更改的数据，并将数据加载到 Microsoft Azure SQL Data Warehouse 中。必须选择**数据驱动**作为目标操作才能捕获已更改的数据。当映射失败或在完成会话之前停止时，您可以从中断点恢复提取已更改的数据。
- 使用 ODBC 连接以连接到 Microsoft Azure SQL Data Warehouse 时，可以利用以下下推函数：
 - Date_diff()
 - First()
 - Instr()
 - Last()
 - MD5()
 - ReplaceChr()
 - ReplaceStr()
- 为提高性能，可以在写入目标对象时以 .gzip 格式压缩暂存文件。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Microsoft Azure SQL Data Warehouse V3 10.2 HotFix 2 用户指南》。

PowerExchange for Salesforce

从版本 10.2 HotFix 2 开始，PowerExchange for Salesforce 包含以下新功能：

- 您可以使用版本 43.0、44.0 和 45.0 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。
- 您可以参数化标准连接和 OAuth 连接的 Salesforce 服务 URL。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Salesforce 10.2 HotFix 2 用户指南》。

PowerExchange for Snowflake

从版本 10.2 HotFix 2 开始，PowerExchange for Snowflake 包含以下新功能：

- 您可以将 PowerCenter 集成服务配置为将被拒绝的记录写入被拒绝的文件。如果该文件已存在，PowerCenter 集成服务会将将被拒绝的记录附加到该文件。如果您未指定被拒绝的文件路径，PowerCenter 集成服务不会写入被拒绝的记录。
- 您可以在映射中使用“更新策略”转换，以在 Snowflake 目标中插入、更新、删除或拒绝数据。当您配置“更新策略”转换时，默认情况下，**将源行视为会话属性被标记为数据驱动**。

- 您可以使用 SQL 编辑器在 Snowflake 源和目标会话中创建或编辑 pre-SQL 和 post-SQL 语句。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Snowflake 10.2 HotFix 2 用户指南》。

安全

本节介绍 10.2 HotFix 2 中新增的安全功能。

Kerberos 跨领域身份验证

从版本 10.2 HotFix 2 开始，可以将 Informatica 域配置为使用 Kerberos 跨领域身份验证。Kerberos 跨领域身份验证使属于一个 Kerberos 领域的 Informatica 客户端能够在属于另一个 Kerberos 领域的节点和应用程序服务上进行身份验证。

有关详细信息，请参阅《Informatica 10.2 HotFix 2 安全指南》。

更改 (10.2 HotFix 2)

Analyst 工具

本节介绍版本 10.2 HotFix 2 中对 Analyst 工具的更改。

默认视图

从版本 10.2 Hotfix 2 开始，平面文件和表对象的默认视图为“属性”选项卡。创建或打开平面文件或表数据对象时，会在“属性”选项卡中打开该对象。以前的默认视图为“数据查看器”选项卡。

有关详细信息，请参阅《Informatica 10.2 HotFix 2 Analyst 工具指南》。

结果卡

从版本 10.2 HotFix 2 开始，最多可以导出 100,000 个无效行。为某个度量导出超过 100 个无效行时，数据集成服务会为该结果卡创建一个文件夹，再为该度量创建一个子文件夹，并创建 Microsoft Excel 文件以导出其余无效行。

有关详细信息，请参阅《Informatica 10.2 HotFix 2 Data Discovery 指南》中的“结果卡”一章。

命令行程序

本节介绍版本 10.2 HotFix 2 中对命令的更改。

infasetup 命令

从版本 10.2 HotFix 2 开始，更改了在 infasetup 命令中使用的 -srn 和 -urn 选项的有效值。

如果将某个域配置为使用 Kerberos 身份验证，则 -srn 和 -urn 选项的值无需相同。如果将某个域配置为使用 Kerberos 跨领域身份验证，则可以指定一个字符串，其中包含该域用于对用户进行身份验证的每个 Kerberos 领域的名称，并以逗号分隔。

在以前的版本中，您可将单个 Kerberos 域的名称指定为 -srn 和 -urn 选项的值。

有关详细信息，请参阅《Informatica 10.2 HotFix 2 命令引用》。

Informatica 转换

本节介绍版本 10.2 HotFix 2 中对 Informatica 转换的更改。

地址验证器转换

本节介绍地址验证器转换的更改。

地址验证器转换对地址功能做了以下更新：

所有国家/地区

自版本 10.2 HotFix 2 开始，地址验证器转换使用 5.14.0 版本的 Informatica Address Verification 软件引擎。

以前，该转换使用 5.13.0 版本的 Informatica Address Verification 软件引擎。

Parcel Centroid 和 Rooftop 地理编码

从版本 10.2 HotFix 2 开始，地址验证器转换上的“地理编码数据类型”属性不包括针对 Parcel Centroid 或 Rooftop 地理编码的选项。

有关详细信息，请参阅《*Informatica 10.2 HotFix 2 Developer 转换指南*》和《*Informatica 10.2 HotFix 2 地址验证器端口参考*》。

有关 Informatica Address Verification 软件引擎更新的详细信息，请参阅《*Informatica Address Verification 5.14.0 发行指南*》。

Metadata Manager

SAP Business Warehouse（已弃用）

从版本 10.2 HotFix 2 开始，Informatica 弃用 Metadata Manager 中的 SAP Business Warehouse 资源。

Cognos 版本 8 或更早版本

从版本 10.2 HotFix 2 开始，如果 Cognos 服务器版本高于版本 8，则请从 <MM_Agent_Home>\MetadataManagerAgent\java\CognosRepository\ 位置删除 *axis* 和 *saajjar* 文件。

有关详细信息，请参阅《*Informatica 10.2 HotFix 2 Metadata Manager 管理员指南*》中的“业务情报资源”一章。

PowerCenter

本节介绍版本 10.2 HotFix 2 中的 PowerCenter 更改。

用于 SAP HANA 的下推优化

从版本 10.2 HotFix 2 开始，您必须具有许可证才能在 ODBC 连接中使用 **SAP HANA** ODBC 子类型。如果您没有许可证，则会显示一条警告消息。在版本 10.2 HotFix 2 中，会话不会失败。但是，在将来的版本中，如果您在没有许可证的情况下使用 **SAP HANA** ODBC 子类型连接，Informatica 会开始让会话失败。因此，Informatica 建议您联系 Informatica 全球客户支持来获得许可证。

Informatica PowerExchange 适配器

本节介绍版本 10.2 HotFix 2 中对 Informatica 适配器的更改。

PowerExchange for MongoDB

从版本 10.2 HotFix 2 开始，架构编辑器不再位于以下目录中：

Informatica 安装目录>/clients/tools/mongodb/Tools

要在 Windows 计算机上使用架构编辑器，请应用 Informatica EBF-13871。

有关详细信息，请参阅《*Informatica PowerExchange for MongoDB 10.2 HotFix 2 用户指南*》。

PowerCenter PowerExchange 适配器

本节介绍版本 10.2 HotFix 2 中对 PowerCenter 适配器的更改。

PowerExchange for Google Analytics

从版本 10.2 HotFix 2 开始，PowerExchange for Google Analytics 随 Informatica 服务一起安装。

以前，PowerExchange for Google Analytics 有一个单独的安装程序。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica PowerExchange for Google Analytics 10.2 HotFix 2 用户指南*》。

PowerExchange for Google Cloud Spanner

从版本 10.2 HotFix 2 开始，PowerExchange for Google Cloud Spanner 随 Informatica 服务一起安装。

以前，PowerExchange for Google Cloud Spanner 有一个单独的安装程序。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica PowerExchange for Google Cloud Spanner 10.2 HotFix 2 用户指南*》。

PowerExchange for Google Cloud Storage

从版本 10.2 HotFix 2 开始，PowerExchange for Google Cloud Spanner 随 Informatica 服务一起安装。

以前，PowerExchange for Google Cloud Spanner 有一个单独的安装程序。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica PowerExchange for Google Cloud Spanner 10.2 HotFix 2 用户指南*》。

PowerExchange for Kafka

从版本 10.2 HotFix 2 开始，PowerExchange for Kafka 随 Informatica 服务一起安装。

以前，PowerExchange for Kafka 有一个单独的安装程序。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica PowerExchange for Kafka 10.2 HotFix 2 用户指南*》。

PowerExchange for MongoDB

从版本 10.2 HotFix 2 开始，架构编辑器不再位于以下目录中：

Informatica 安装目录>/clients/tools/mongodb/Tools

要在 Windows 计算机上使用架构编辑器，请应用 Informatica EBF-13871。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica PowerExchange for MongoDB 10.2 HotFix 2 用户指南*》。

PowerExchange for Salesforce

从版本 10.2 HotFix 2 开始，源中的空值在目标中也被视为空值。

以前，源中的空值在目标中也被视为 false。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica PowerExchange for Salesforce 10.2 HotFix 2 用户指南*》。

PowerExchange for Snowflake

从版本 10.2 HotFix 1 开始，PowerExchange for Snowflake 具有以下更改：

- 您可以替代通过 Snowflake 连接导入表时所使用的数据库和架构名称。要替代，请采用以下格式在 Snowflake 连接属性的**其他 JDBC URL 参数字段**中输入数据和架构名称：DB=<DB_name>&Schema=<schema_name>以前，只能在会话属性中为数据库和架构名称指定替代。
- 您可以读取或写入区分大小写的数据或包含特殊字符的数据。不能使用以下特殊字符：@ ~ \ 以前，必须确保源表名称和目标表名称只包含大写字母。
- PowerExchange for Snowflake 使用 Snowflake JDBC 驱动程序版本 3.6.26。以前，PowerExchange for Snowflake 使用 Snowflake JDBC 驱动程序版本 3.6.4。
- 当您运行自定义查询以从 Snowflake 中读取数据时，集成服务将运行该查询并优化导入元数据的性能。以前，当您自定义查询配置为从 Snowflake 中读取数据时，集成服务会将该查询发送到 Snowflake 端点，并且会将要导入的记录数限制为 10 条记录。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Snowflake 10.2 HotFix 2 用户指南》。

PowerExchange for Teradata Parallel Transporter API

从版本 10.2 HotFix 2 开始，可以在 AIX 计算机上运行 Teradata ODBC 映射。当您使用 Teradata 客户端版本 16.20.x 时，必须安装 Teradata ODBC 驱动程序版本 16.20.00.50 或更高版本。

以前，当您使用早于 16.20.00.50 的 Teradata ODBC 驱动程序版本时，无法在 AIX 计算机上运行 Teradata ODBC 映射。

发布任务(10.2 HotFix 2)

Informatica PowerExchange 适配器

本节介绍版本 10.2 HotFix 2 中 Informatica 适配器的发布任务。

PowerExchange for Snowflake

从版本 10.2 HotFix 2 开始，Snowflake JDBC 驱动程序版本 3.6.26 随 Informatica 服务一起安装。10.2 HotFix 1 版本中的现有 Snowflake 映射可能会在升级后失败，因为安装位置现在包含来自版本 10.2 HotFix 1 的 Snowflake JDBC 驱动程序版本 3.6.4 以及来自版本 10.2 HotFix 2 的版本 3.6.26。

要从 Developer Tool 中运行现有的 Snowflake 映射，必须删除 Snowflake JDBC 驱动程序版本 3.6.26 并仅保留数据集成服务计算机中以下目录内的版本 3.6.4：<Informatica 安装目录>\connectors\thirdparty

PowerCenter PowerExchange 适配器

本节介绍版本 10.2 HotFix 2 中 PowerCenter 适配器的发布任务。

PowerExchange for Google BigQuery

从版本 10.2 HotFix 2 开始，PowerExchange for Google BigQuery 插件已更新了**自定义查询**、**SQL 替代**和**引号字符**属性的新元素。早期版本中的现有 Google BigQuery 映射可能会在升级后失败，因为现有插件不包含新元素。

要从 PowerCenter 客户端运行现有 Google BigQuery 映射并启用**自定义查询**、**SQL 替代**和**引号字符**属性，必须在 PowerCenter 存储库中重新注册 bigqueryPlugin.xml 插件。

第 23 章

新功能、更改和发布任务 (10.2 HotFix 1)

本章包括以下主题：

- [新增功能 \(10.2 HotFix 1\), 231](#)
- [更改 \(10.2 HotFix 1\), 241](#)
- [发布任务 \(10.2 HotFix 1\), 247](#)

新增功能 (10.2 HotFix 1)

应用程序服务

本节介绍版本 10.2 HotFix 1 中新增的应用程序服务功能。

模型存储库服务

Git 版本控制系统

从版本 10.2 HotFix 1 开始，可以将模型存储库与 Git 版本控制系统集成在一起。Git 是一个分布式版本控制系统。签出和签入对象时，版本的副本会保存到本地存储库和 Git 服务器。如果 Git 服务器已关闭，则本地存储库会保留该对象的所有版本。

有关详细信息，请参阅《*Informatica 10.2 HotFix 1 应用程序服务指南*》中的“模型存储库服务”一章。

Business Glossary

本节介绍版本 10.2 HotFix 1 中新增的 Business Glossary 功能。

以纯文本形式导入词汇表资产

从 10.2 HotFix 1 开始，您可以以纯文本形式将业务词汇表资产导入 Analyst 工具。

有关导出和导入词汇表资产的详细信息，请参阅《*Informatica 10.2 HotFix 1 Business Glossary 指南*》中的“词汇表管理”一章。

命令行程序

本节介绍版本 10.2 HotFix 1 中新增的命令。

infacmd isp 命令

下表介绍了新的 infacmd isp 命令：

命令	说明
PingDomain	对域、服务、域网关主机或节点执行 Ping 操作。
ListPasswordRules	列出设置复杂密码的规则。
ListWeakPasswordUsers	列出密码不符合密码策略的用户。

有关详细信息，请参阅《Informatica 10.2 HotFix 1 命令参考》中的“infacmd isp 命令参考”一章。

infacmd wfs 命令

下表介绍了新的 infacmd wfs 命令：

命令	说明
pruneOldInstances	从 workflow 数据库中删除 workflow 处理数据。

要删除处理数据，您必须具有该域的"管理服务"特权。

有关详细信息，请参阅《Informatica 10.2.1 命令参考》中的"infacmd wfs 命令参考"一章。

连接

本节介绍版本 10.2 HotFix 1 中新增的连接功能。

与云数据库的连接

从版本 10.2 HotFix 1 开始，您可以在 Informatica Developer 和 PowerCenter 中配置关系连接以连接到云数据库。

您可以配置以下连接：

- 连接到 Oracle 自主数据仓库云的 Oracle 连接版本 18C
- 连接到 Oracle 数据库云服务的 Oracle 连接版本 12C
- 连接到 Azure SQL 数据库的 Microsoft SQL Server 连接
- 连接到 DashDB 的 IBM DB2 连接

有关详细信息，请参阅《Informatica 10.2 HotFix 1 安装和配置指南》。

数据类型

本节介绍 10.2 HotFix 1 中新增的数据类型功能。

Microsoft SQL Server 数据类型

从 10.2 HotFix 1 开始，在使用 Informatica 映射中的 Microsoft SQL Server 连接时，您可以读取或写入日期数据类型。

有关详细信息，请参阅《*Informatica 10.2 HotFix 1 Developer Tool 指南*》中的“数据类型参考”附录。

安装程序

本节介绍版本 10.2 HotFix 1 中的新增安装程序功能。

Docker 实用程序

从版本 10.2 HotFix 1 开始，您可以使用 Informatica PowerCenter Docker 实用程序创建 Informatica 域服务。您可以构建具有基本操作系统和 Informatica 二进制文件的 Informatica Docker 映像，然后运行现有 Docker 映像以在容器中创建 Informatica 域。

有关安装 Informatica PowerCenter Docker 实用程序以创建 Informatica 域服务的详细信息，请参阅 <https://kb.informatica.com/h2l/HowTo%20Library/1/1213-InstallInformaticaUsingDockerUtility-H2L.pdf>。

Informatica 转换

本节介绍版本 10.2 HotFix 1 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 HotFix 1 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。

奥地利、德国和瑞士

从 10.2 HotFix 1 开始，地址验证器转换将在奥地利、德国和瑞士地址中支持大写字符 ß。

转换通过以下方式支持字符 ß：

- 如果将“大小写”属性设置为 UPPER，则转换会为德语字符 ß 返回 ß。如果将“大小写”属性设置为 LOWER，则转换会为德语字符 ß 返回 ß。
- 转换将地址中的 ß 和 S 视为同样有效的字符。在引用数据匹配中，当相同值包含 ß 或 S 时，转换会将其确定为完美匹配。
- 转换将地址中的 ß 和 ss 视为同样有效的字符。在引用数据匹配中，当相同值包含 ß 或 ss 时，转换会将其确定为标准匹配。
- 如果将“首选语种”属性设置为 ASCII_SIMPLIFIED，转换会将字符 ß 返回为 S。
- 如果将“首选语种”属性设置为 ASCII_EXTENDED，转换会将字符 ß 返回为 SS。

玻利维亚

从版本 10.2 HotFix 1 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 HotFix 1 开始，您可以将转换配置为返回短或长形式的元素描述符。

地址验证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 HotFix 1 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 HotFix1 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。

地址验证器转换将哥伦比亚地址验证到门牌号级别。转换不仅可以验证包含房屋所在街道信息的哥伦比亚地址，而且可以验证包含离房屋最近的十字街信息的地址。

例如，转换可以验证以下地址：

AVENIDA 31 65 29 APTO 1626

AVENIDA 31 DIAGONAL 65 29 APTO 1626

请为哥伦比亚地址考虑以下准则：

- 不管地址带不带十字街描述符 DIAGONAL，地址验证器转换都可以验证。
- 不管在十字街号码和门牌号之间有没有短划线符号，转换都可以验证地址。转换在输出地址中的此位置不包括短划线符号。
- 转换可以识别输入地址中十字街号码前的 # 符号。转换将该地址识别为引用数据匹配，而不是更正。

例如，地址验证将以下输入地址识别为引用数据匹配，并从验证后的输出地址中省略 # 符号：

AVENIDA 31 #65 29 APTO 1626

法国

从版本 10.2 HotFix 1 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。

印度

从版本 10.2 HotFix 1 开始，地址验证器转换会将印度的地址验证到门牌号级别。

秘鲁

从版本 10.2 HotFix 1 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。

南非

从版本 10.2 HotFix 1 开始，地址验证器转换改进了对南非地址中的投递服务描述符的解析和验证。

转换通过以下方式改进了对投递服务描述符的解析和验证：

- Address Verification 将"专用邮袋"、"团体信箱"、"邮政信箱"和 Postnet Suite 识别为不同的投递服务类型。Address Verification 不会将一个投递服务描述符标准化为其他投递服务描述符。例如，Address Verification 不会将 Postnet Suite 标准化为"邮政信箱"。
- Address Verification 会将 Postnet Box 解析为非标准投递服务描述符，并将 Postnet Box 更正为有效的描述符 Postnet Suite。
- Address Verification 不会将子建筑物描述符 Flat 标准化为 Fl。

韩国

从版本 10.2 HotFix 1 开始，地址验证器转换为韩国引入了以下功能和增强功能：

- 韩国地址引用数据包含建筑物信息。转换可以读取、验证和更正韩国地址中的建筑物信息。
- 转换将返回旧地址所代表的房屋的所有当前地址。旧地址可能代表单个当前地址，或者可能代表多个地址，例如，如果多位居民居住在该房屋地点。

要返回当前地址，请先查找旧房屋的地址 ID。以地址代码查找模式提交最后一个字符为 A 的地址 ID 时，转换将返回与该地址 ID 匹配的所有当前地址。

注意：地址验证器转换使用“最大结果数”属性来确定为输入的地址 ID 返回的最大地址个数。“计数溢出”属性指示数据库是否包含该地址 ID 的其他地址。

瑞典

从版本 10.2 HotFix 1 开始，地址验证器转换改进了瑞典地址中的街道名称验证。

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

泰国

从版本 10.2 HotFix 1 开始，地址验证器转换为泰国引入了以下功能和增强功能：

对泰国地址进行的改进

转换改进了对采用拉丁语种的泰国地址的解析和验证。

此外，转换在验证地址时会验证到门牌号级别。

对泰国地址的本地支持

地址验证器转换能以本地泰语和拉丁语种读取和写入泰国地址。Informatica 更新了泰国地址的引用数据，并添加了采用本地泰语语种的引用数据。

Informatica 为每种语种的泰国地址提供单独的引用数据库。要验证采用本地泰语语种的地址，请安装本地泰语数据库。要验证采用拉丁语种的地址，请安装拉丁语数据库。

注意: 如果要验证泰国地址，请勿同时安装这两种类型的数据库。接受“首选语种”属性的默认选项。

阿拉伯联合酋长国

从版本 10.2 HotFix 1 开始，地址验证器转换将验证阿拉伯联合酋长国地址中的街道名称。要验证阿拉伯联合酋长国地址中的街道名称，请安装当前的阿拉伯联合酋长国地址引用数据库。

英国

从版本 10.2 HotFix 1 开始，地址验证器转换可以返回英国领地名称。

转换在 Country_2 元素中返回领地名称，在 Country_1 元素中返回国家/地址名称。您可以为输出地址配置这两个元素；或者如果是在英国境内投递邮件，则可以省略 Country_1 元素。在信封或标签上，领地名称显示在英国地址中的邮政编码上方。

要返回领地名称，请安装当前的英国地址引用数据。

美国

从版本 10.2 HotFix 1 开始，地址验证器转换最多可以识别美国地址中的三个子建筑物级别。

按照美国邮政署的要求，转换将单个子建筑物元素中的信息与引用数据匹配。如果 Sub-building_1 中的信息不匹配，则转换会比较 Sub-building_2 中的信息。如果 Sub-building_2 中的信息不匹配，则转换会比较 Sub-building_3 中的信息。转换会将输入地址中不匹配的子建筑物信息复制到输出地址。

有关 Informatica 嵌入到版本 10.2 HotFix 1 的地址验证软件引擎版本的功能和操作的全面信息，请参阅《*Informatica Address Verification 5.13.0 开发人员指南*》。

Metadata Manager

本节介绍版本 10.2 HotFix 1 中新增的 Metadata Manager 功能。

Metadata Manager 的 SAML 身份验证

从版本 10.2 HotFix 1 开始，Metadata Manager 支持基于安全断言标记语言 (SAML) 的单点登录。基于 SAML 的单点登录根据存储在 Microsoft Active Directory 中的帐户凭据对用户进行身份验证。可以从 Active Directory 将帐户导入到 Informatica 域中的安全域。

有关信息，请参阅《*Informatica 10.2 HotFix 1 安全指南*》中的“Informatica Web 应用程序的 SAML 身份验证”一章。

在 Metadata Manager 存储库备份或还原操作期间跳过沿袭

从版本 10.2 HotFix 1 开始，在 backupRepository 和 restoreRepository 命令中使用 [`<-sl|--skipLineage>` skipLineage] 选项，以在 Metadata Manager 存储库备份和还原操作期间跳过沿袭。

有关信息，请参阅《*Informatica 10.2 HotFix 1 Metadata Manager 命令参考指南*》。

PowerCenter

本节介绍版本 10.2 HotFix 1 中新增的 PowerCenter 功能。

用于 SAP HANA 的下推优化

从版本 10.2 HotFix 1 开始，如连接类型为 ODBC，则可选择 ODBC 提供程序子类型作为 **SAP HANA**，将转换逻辑推送到 SAP HANA。您可以配置源侧、目标侧或完整下推优化，将转换逻辑推送到 SAP HANA。

有关详细信息，请参阅《*Informatica PowerCenter 10.2 HotFix 1 高级工作流指南*》。

下推优化用于 Snowflake

从版本 10.2 HotFix 1 开始，如连接类型为 ODBC，则可配置源侧或完整下推优化，将转换逻辑推送到 Snowflake。

有关信息，请参阅《*Informatica PowerCenter 10.2 HotFix 1 高级工作流指南*》。

Informatica PowerExchange 适配器

本节介绍版本 10.2 HotFix 1 中新增的 Informatica 适配器功能。

PowerExchange for Microsoft Azure Blob 存储

从版本 10.2 HotFix 1 开始，PowerExchange for Microsoft Azure Blob 存储包含以下功能：

- 可以在 Spark 引擎上运行映射。
- 您可以使用文件端口存储数据集成服务在运行时从中读取数据的文件名。
- 您可以在本地环境中运行映射时读取和写入 .csv、Avro 和 Parquet 文件。
- 您可以在本地环境中运行映射时读取目录。
- 在本地环境中运行映射时，可以生成或跳过标题行。在 Spark 引擎上，默认情况下会创建标题行。
- 可以追加现有 blob。在本地环境中，追加操作仅适用于 .csv 文件。
- 可以覆盖 blob 或容器名称。在“Blob 容器覆盖”字段中，使用绝对路径指定容器名称或根容器中的子文件夹。
- 在本地环境中，您可以读取和写入用 gzip 格式压缩的 .csv 文件。

有关详细信息，请参阅《*Informatica PowerExchange for Microsoft Azure Blob 存储 10.2 HotFix 1 用户指南*》。

PowerExchange for Microsoft Azure SQL 数据仓库

从版本 10.2 HotFix 1 开始，PowerExchange for Microsoft Azure SQL 数据仓库包含以下功能：

- 从 Microsoft Azure SQL 数据仓库对象读取数据时，可以配置键范围分区。
- 从 Microsoft Azure SQL 数据仓库对象读取数据时，可以替代 SQL 查询并定义约束。
- 可为映射中的源对象和目标对象配置 pre-SQL 和 post-SQL 查询。
- 可以为源数据对象操作配置本地表达式筛选器。
- 可以对 Microsoft Azure SQL 数据仓库表执行更新、插入和删除操作。
- 您可以在本地环境中配置未缓存的查找操作。

有关详细信息，请参阅《*Informatica PowerExchange for Microsoft Azure SQL 数据仓库 10.2 HotFix 1 用户指南*》。

PowerExchange for Netezza

从版本 10.2 HotFix 1 开始，您可以配置动态映射，以便在运行时根据定义的参数和规则更改 Netezza 源和目标。

当您配置动态映射时，还可以在运行时创建或替换目标。您可以在 Netezza 数据对象写入操作的高级属性中选择 **在运行时创建或替换表** 选项。

PowerExchange for Teradata Parallel Transporter API

从版本 10.2 HotFix 1 开始，您可以配置动态映射，以便在运行时根据定义的参数和规则更改 Teradata 源和目标。

当您配置动态映射时，还可以在运行时创建或替换 Teradata 目标。您可以在 Teradata 数据对象写入操作的高级属性中选择 **在运行时创建或替换表** 选项。

PowerCenter PowerExchange 适配器

本节介绍版本 10.2 HotFix 1 中新增的 PowerCenter 适配器功能。

PowerExchange for Amazon Redshift

从版本 10.2 HotFix 1 开始，PowerExchange for Amazon Redshift 包含以下新功能：

- 除了现有区域，还可以从 AWS GovCloud 区域读取数据或向其中写入数据。
- 您可以指定对象的部分大小，以便分多个部分从 Amazon S3 下载该对象。
- 从 Amazon Redshift 提取文件时，您可以使用 AWS 托管的加密密钥或 AWS KMS 托管的客户主密钥加密数据来进行服务器端加密。
- 您可以提供文件数，以便为每一个批处理计算暂存文件数。如果您不提供文件数，PowerExchange for Amazon Redshift 将计算暂存文件数。
- 将数据写入目标之前，您可以使用复制命令中的 TRUNCATECOLUMNS 选项截断 VARCHAR 和 CHAR 数据类型列的数据。
- PowerExchange for Amazon Redshift 支持版本 11 和 12 的 SuSe Linux Enterprise Server 操作系统。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Amazon Redshift 10.2 HotFix 1 用户指南》。

PowerExchange for Amazon S3

从版本 10.2 HotFix 1 开始，PowerExchange for Amazon S3 包含以下新功能：

- 除了现有区域，还可以从 AWS GovCloud 区域读取数据或向其中写入数据。
- 您可以指定从 Amazon S3 读取数据时要用做表头的行。您可以在源会话属性下的**标题行号**属性中指定行号。
- 您可以指定希望 PowerCenter 集成服务从其读取数据的行号。您可以配置源会话属性下的**数据读取行号**属性。
- 您可以在文件名中指定星号 (*) 通配符以提取 Amazon S3 存储桶中的文件。您可以指定星号 (*) 通配符以提取所有文件或仅提取与名称模式相匹配的文件。
- 您可以向存储在 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。您可以输入键值对，或者指定包含键值对的绝对文件路径。
- 您可以指定对象的部分大小，以便分多个部分从 Amazon S3 下载该对象。
- 您可以为 Amazon S3 源配置分区。从 Amazon S3 源读取数据时，分区可以优化运行时的映射性能。

- PowerExchange for Amazon S3 支持版本 11 和 12 的 SuSe Linux Enterprise Server 操作系统。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Amazon S3 10.2 HotFix 1 用户指南》。

PowerExchange for Cassandra

从版本 10.2 HotFix 1 开始，Informatica Cassandra ODBC 驱动程序支持异步写入。

要在 Linux 操作系统上启用异步写入，必须在 odbc.ini 文件中添加 **EnableAsynchronousWrites** 项名称并将值设置为 1。

要在 Windows 操作系统上启用异步写入，必须在 Windows 注册表中为 Cassandra ODBC 数据源名称添加 **EnableAsynchronousWrites** 属性并将值设置为 1。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Cassandra 10.2 HotFix 1 用户指南》。

PowerExchange for Microsoft Dynamics CRM

从版本 10.2 HotFix 1 开始，PowerExchange for Microsoft Dynamics CRM 包含以下新功能：

- 您可以选择**发现服务**或**组织服务**作为在 Microsoft Dynamics CRM 运行时连接中进行护照身份验证的服务类型。
- 您可以在更新、插入和删除操作中配置替代键。
- 您可以为“Lookup”、“Customer”、“Owner”和“PartyList”数据类型指定替代键作为引用。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Microsoft Dynamics CRM 10.2 HotFix 1 用户指南》。

PowerExchange for Salesforce

从版本 10.2 HotFix 1 开始，PowerExchange for Salesforce 包含以下新功能：

- 您可以使用版本 42.0 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。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Salesforce 10.2 HotFix 1 用户指南》。

PowerExchange for SAP NetWeaver

从版本 10.2 HotFix 1 开始，PowerExchange for SAP NetWeaver 包含以下新功能：

对业务内容集成映射的连接恢复支持

您可以在业务内容集成映射的侦听器会话中配置以下连接恢复参数：

- **连接恢复的重试次数。**定义 PowerCenter 集成服务在与 SAP 连接不成功时必须尝试的连接重试次数。
- **各次连接恢复尝试之间的延迟。**定义各次连接重试之间以秒为单位的时间间隔。

新 SAP 数据类型支持

您可以根据使用的集成方法使用以下新 SAP 数据类型：

数据类型	使用 ABAP 程序（Table Reader 和 Table Writer）进行数据集成	使用 BAPI/RFC 功能进行数据集成	使用 ALE 进行 IDoc 集成
INT8	支持	不受支持	不受支持
DF16_DEC	支持	支持	不受支持
DF34_DEC	支持	支持	不受支持
DF16_DEC	支持	支持	不受支持
DF34_DEC	支持	支持	不受支持
DF16_RAW	支持	支持	不受支持
DF34_RAW	支持	支持	不受支持
RAWSTRING	支持	不受支持	支持
STRING	支持	支持	支持
SSTRING	支持	支持	不受支持

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for SAP NetWeaver 10.2 HotFix 1 用户指南》。

PowerExchange for Snowflake

从版本 10.2 HotFix 1 开始，PowerExchange for Snowflake 包含以下新功能：

- 您可以配置自定义 SQL 查询以配置 Snowflake 源。
- 当您在 Snowflake 暂存数据库中创建临时表时，可以替代数据库和架构名称。您必须添加希望在 Snowflake 连接的 JDBC URL 参数中替代的数据库和架构名称。
- 您可以通过在 Snowflake 会话属性中指定表名称来替代导入的 Snowflake 源表名称。还可以配置 SQL 替代，以替代用于从 Snowflake 源提取数据的默认 SQL 查询。
- 您可以通过在 Snowflake 目标会话属性中指定表名称来替代 Snowflake 目标表名称。
- 在使用 ODBC 连接类型时，您可以配置源侧或完整下推优化，将转换逻辑推送到 Snowflake。有关 PowerCenter 集成服务可以推送到 Snowflake 的运算符和函数的信息，请参阅《Informatica PowerCenter 10.2 HotFix 1 高级工作流指南》。
- 您可以通过指定联接条件来联接多个 Snowflake 源表。
- 您可以为映射中的源限定符配置未连接的查找转换。
- 您可以为 Snowflake 会话配置直通分区。添加分区数后，可以为每一个分区指定 SQL 替代或筛选器替代条件。
- 您可以在设计时或运行时配置 HTTP 代理服务器身份验证设置，以读取 Snowflake 中的数据或将数据写入其中。
- 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。

- 您可以从为 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《适用于 PowerCenter 的 Informatica PowerExchange for Snowflake 10.2 HotFix 1 用户指南》。

PowerExchange for Teradata Parallel Transporter API

从版本 10.2 HotFix 1 开始，当您使用 Load 操作符时，可以设置 Teradata PT API 用于写入数据的最大缓冲区大小（以 KB 为单位）。您可以在 Teradata 目标会话属性中设置最大缓冲区大小值。

安全

本节介绍版本 10.2 HotFix 1 中新增的安全功能。

安全 - 密码复杂性

从版本 10.2 HotFix 1 开始，您可以启用密码复杂性以验证密码强度。默认情况下，禁用此选项。

有关详细信息，请参阅《Informatica 10.2 HotFix 1 安全指南》中的“Informatica Administrator 中的安全管理”一章。

更改 (10.2 HotFix 1)

支持更改

本节介绍 10.2 HotFix 1 中的支持更改。

Big Data Management Hadoop 发行版

下表列出了 Big Data Management 10.2 HotFix 1 中支持的 Hadoop 发行版版本和更改：

Hadoop 发行版	支持的发行版版本	10.2 HotFix 1 中的更改
Amazon EMR	5.8	终止了对版本 5.4 的支持。
Azure HDInsight	3.5.x 3.6.x	无更改。
Cloudera CDH	5.10.x 5.11.x 5.12.x 5.13.x	终止了对版本 5.9 的支持。
Hortonworks HDP	2.5x 2.6x	终止了对版本 2.4 的支持。

Hadoop 发行版	支持的发行版版本	10.2 HotFix 1 中的更改
IBM BigInsights	不受支持。	终止了对 IBM BigInsights 的支持。
MapR	5.2 MEP 3.0.x	终止了对版本 5.2 MEP 2.0.x 的支持。

Informatica 大数据产品支持各种 Hadoop 发行版。在每个版本中，Informatica 会增加、延期和终止对 Hadoop 发行版版本的支持。Informatica 可能会在将来的版本中恢复对延续版本的支持。

要查看支持的版本的最新列表，请参阅 Informatica Network 上的"产品可用性列表"，网址为：
<https://network.informatica.com/community/informatica-network/product-availability-matrices>。

Intelligent StreamingHadoop 分布

下表列出了 Intelligent Streaming 10.2 HotFix 1 中支持的 Hadoop 发行版版本和更改：

发行版	支持的版本	自 10.2 以来的更改
Amazon EMR	5.8	终止了对版本 5.4 的支持。
Cloudera CDH	5.10 5.11 5.12 5.13	无更改。
Hortonworks HDP	2.5.x 2.6.x	无更改。
MapR	5.2 MEP 3.0	终止了对版本 5.2 MEP 2.0 的支持。

要查看支持的版本的最新列表，请参阅 Informatica Network 上的“产品可用性列表”，网址为：<https://network.informatica.com/community/informatica-network/product-availability-matrices>。

应用程序服务

本节介绍版本 10.2 HotFix 1 中对应用程序服务的更改。

模型存储库服务

监视模型存储库服务

从版本 10.2 HotFix 1 开始，可以将模型存储库服务配置为监视模型存储库服务，以监视 Ad Hoc 作业、应用程序、逻辑数据对象、SQL 数据服务、Web 服务和工作流的统计信息。配置监视模型存储库和模型存储库时应使用不同的数据库用户帐户。

在以前的版本中，可以使用模型存储库服务将设计时对象和运行时对象存储在模型存储库中。

有关详细信息，请参阅《Informatica 10.2 HotFix 1 应用程序服务指南》中的“模型存储库服务”一章。

Big Data Management

本节介绍版本 10.2 HotFix 1 中对 Big Data Management 进行的更改。

Hive 引擎上的精度和小数位数

从版本 10.2 HotFix 1 开始，如果以下条件属实，则用户定义的在 Hive 引擎上执行乘法的函数的输出具有最大小数位数 6：

- 精度和小数位数之差大于等于 32。
- 结果精度大于 38。

以前，小数位数可能低至 0。

有关详细信息，请参阅《*Informatica Big Data Management 10.2 HotFix 1 用户指南*》中的“Hadoop 环境中的映射”一章。

Business Glossary

本节介绍版本 10.2 HotFix 1 中 Business Glossary 的更改。

警告和通知

从 10.2 HotFix 1 开始，除了发送人及提交供审查的资产的名称，Analyst 工具还将显示词汇表名称。以前，Analyst 工具仅显示发送人的名称及提交供审查的资产的名称。

有关详细信息，请参阅《*Informatica 10.2 HotFix 1 Business Glossary 指南*》中的“查找词汇表内容”一章。

文档

本节介绍版本 10.2 HotFix 1 中 Informatica 文档中的指南更改。

Informatica PowerCenter 10.2 HotFix 1 存储库指南

从版本 10.2 HotFix 1 开始，《*Informatica PowerCenter 10.2 HotFix 1 存储库指南*》包括附录“MX 视图参考”和“PowerCenter 报告参考”。《*Informatica PowerCenter 10.2 HotFix 1 PowerCenter 报告指南*》和《*Informatica PowerCenter 10.2 HotFix 1 PowerCenter 报告使用指南*》已移至《*Informatica PowerCenter 10.2 HotFix 1 存储库指南*》，作为其中的一个附录。

有关详细信息，请参阅《*Informatica PowerCenter 10.2 HotFix 1 存储库指南*》。

Informatica Development Platform

本节介绍版本 10.2 HotFix 1 中对 Informatica Development Platform 的更改。

Informatica 连接器工具包

从版本 10.2 HotFix 1 开始，默认可以使用 Informatica 连接器工具包为 Informatica 智能云服务构建连接器。如果您要发布使用 10.2 HotFix 1 之前的 Informatica 连接器工具包版本开发的连接器，请设置 YES 作为 CTK_ADAPTER 系统环境变量的值。例如，CTK_ADAPTER =YES。

以前，您可以使用 Informatica 连接器工具包为 Informatica 云服务构建连接器。

有关详细信息，请参阅《*Informatica Development Platform 10.2 HotFix 1 用户指南*》。

Informatica 转换

本节介绍版本 10.2 HotFix 1 中对 Informatica 转换的更改。

地址验证器转换

本节介绍地址验证器转换的更改。

地址验证器转换对地址功能做了以下更新：

所有国家/地区

从版本 10.2 HotFix 1 开始，地址验证器转换使用 5.13.0 版本的 Informatica Address Verification 软件引擎。该引擎支持 Informatica 在版本 10.2 HotFix 1 中向地址验证器转换添加的功能。

以前，该转换使用 5.11.0 版本的 Informatica Address Verification 软件引擎。

Parcel Centroid 和 Rooftop 地理编码

从 2018 年 6 月开始，Informatica 停止更新包含 Parcel Centroid 和 Rooftop 地理编码数据的引用数据文件，并且 Informatica 停止向第一次使用的用户提供这些文件。Parcel Centroid 和 Rooftop 地理编码的当前引用数据文件将保持工作状态。

有关详细信息，请参阅《*Informatica 10.2 HotFix 1 Developer 转换指南*》和《*Informatica 10.2 HotFix 1 地址验证器端口参考*》。

有关从版本 5.11.0 到版本 5.13.0 中的 Informatica Address Verification 软件引擎的更新的全面信息，请参阅《*Informatica Address Verification 5.13.0 发行指南*》。

PowerCenter

本节介绍版本 10.2 HotFix 1 中对 PowerCenter 的更改。

Microsoft Analyst for Excel

从版本 10.2 HotFix 1 开始，Informatica 支持 Mapping Analyst for Excel 与 Microsoft Excel 2016。Mapping Analyst for Excel 包括一个 Excel 加载项，您可以使用它来配置 Microsoft Excel 2016 中的映射规范。

以前，Informatica 支持的是 Mapping Analyst for Excel 与 Microsoft Excel 2007 和 Microsoft Excel 2010。

有关安装 Microsoft Excel 2016 加载项的详细信息，请参阅《*Informatica PowerCenter 10.2 HotFix 1 Mapping Analyst for Excel 指南*》。

PowerCenter PowerExchange 适配器

本节介绍版本 10.2 HotFix 1 中对 PowerCenter 适配器的更改。

PowerExchange for Amazon Redshift

从版本 10.2 HotFix 1 开始，PowerExchange for Amazon Redshift 具有以下更改：

- 您可以在目标会话属性下的**每个批处理的文件数**字段中提供文件数，以便为每一个批处理计算暂存文件数。以前，按照您在连接属性下的**群集节点类型**和**群集中的节点数**字段中提供的值计算每个批处理的暂存文件数。
- 会话日志包含有关以下操作分别所需时间的信息：将数据上载到本地暂存区域、从本地暂存区域将数据上载到 Amazon S3，然后通过发出复制命令将数据上载到 Amazon Redshift 目标。以前，会话日志仅包含将数据从源写入目标所需的总时间。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica PowerExchange for Amazon Redshift 10.2 HotFix 1 用户指南*》。

PowerExchange for Cassandra

从版本 10.2 HotFix 1 开始，PowerExchange for Cassandra 具有以下更改：

- Informatica PowerExchange for Cassandra ODBC 驱动程序文件的名称和目录已更改。

下表根据 Linux 和 Windows 操作系统列出了 Cassandra ODBC 驱动程序文件名称与目录：

操作系统	Cassandra ODBC 驱动程序文件名称	文件目录
Linux	libcassandraodbc_sb64.so	<Informatica 安装目录>\tools\cassandra\lib\libcassandraodbc_sb64.so
Windows	CassandraODBC_sb64.dll	<Informatica 安装目录>\tools\cassandra\lib\CassandraODBC_sb64.dll

在 Linux 操作系统上，必须为 odbc.ini 文件中的现有 Cassandra 数据源将 Driver 属性的值更新为 <Informatica 安装目录>\tools\cassandra\lib\libcassandraodbc_sb64.so。

在 Windows 上，必须在 Windows 注册表中为现有 Cassandra 数据源名称更新以下属性：

Driver=<Informatica installation directory>\tools\cassandra\lib\CassandraODBC_sb64.dll
Setup=<Informatica installation directory>\tools\cassandra\lib\CassandraODBC_sb64.dll

- “负载平衡策略”选项的新项名称为 **LoadBalancingPolicy**。
以前，“负载平衡策略”的项名称为 **COLoadBalancingPolicy**。
- 以下 Cassandra ODBC 驱动程序属性的默认值已更改：

驱动程序属性名称	项名称	新的默认值
并发请求数	NumConcurrentRequests	100
插入查询线程	NumInsertQueryThreads	2
每个插入线程迭代次数	NumIterationsPerInsertThread	50

有关详细信息，请参阅《Informatica PowerExchange for Cassandra 10.2 HotFix 1 用户指南》。

PowerExchange for Google BigQuery

从版本 10.2 HotFix 1 开始，PowerExchange for Google BigQuery 随 Informatica 服务一起安装。

以前，PowerExchange for Google BigQuery 有一个单独的安装程序。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Google BigQuery 10.2 HotFix 1 用户指南》。

PowerExchange for Salesforce

从版本 10.2 HotFix 1 开始，您在 Salesforce 会话期间遇到的错误消息可能会因为 Salesforce API 更改而有所不同。

例如，当您重新连接到 Salesforce，将显示以下错误消息。

[ERROR] 由于发生错误 [返回未知格式的服务器错误]，请重新尝试 Salesforce 请求 [getBatchInfo]。

以前，对于相同的场景显示以下错误消息：

[ERROR] 由于发生错误 [输入流不能为空]，请重新尝试 Salesforce 请求 [getBatchInfo]。

PowerExchange for Salesforce Analytics

从版本 10.2 HotFix 1 开始，PowerExchange for Salesforce Analytics 随 Informatica 服务一起安装。

以前，PowerExchange for Salesforce Analytics 有单独的安装程序。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Salesforce Analytics 10.2 HotFix 1 用户指南》。

PowerExchange for Snowflake

从版本 10.2 HotFix 1 开始，PowerExchange for Snowflake 随 Informatica 服务一起安装。

以前，PowerExchange for Snowflake 有一个单独的安装程序。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Snowflake 10.2 HotFix 1 用户指南》。

引用数据

本节介绍版本 10.2 HotFix 1 中对引用数据操作的更改。

内容安装程序

从 2018 年春季开始，Informatica 不再为加速器文件和引用数据文件提供内容安装程序实用程序。要将加速器文件或引用数据文件添加到 Informatica 安装中，请提取文件并将其复制到安装中的适当目录。

以前，可以使用内容安装程序来提取文件并将其复制到 Informatica 目录。

有关详细信息，请参阅《Informatica 10.2 HotFix 1 内容指南》。

PowerCenter 中的属性文件

从版本 10.2 HotFix 1 开始，升级过程和 HotFix 安装过程将保持早期 PowerCenter 版本中的引用数据属性的内容。在安装 HotFix 或升级到版本 10.2 HotFix 1 后，您不需要编辑任何引用数据属性文件。

PowerCenter 从以下属性文件中读取引用数据的配置信息：

- AD50.cfg：包含地址引用数据的属性。
- CLASSIFIER.properties：包含分类器模型的属性。
- IDQTx.cfg：包含标识社群的属性。
- NER.properties：包含概率模型的属性。

HotFix 安装或升级过程将属性文件的备份版本写入到 Informatica 10.2 HotFix 1 目录。备份文件是默认版本的文件，不包含您在早期安装中设置的任何值。每个备份文件名均以扩展名 .bak 结尾。

以前，升级操作使用扩展名 .bak 重命名它找到的任何引用数据属性文件。升级操作还会创建它重命名的任何属性文件的默认版本。

注意：以前，如果您为某个安装安装了 HotFix，为了让 Informatica 目录结构不发生更改，安装过程会保留 AD50.cfg 文件。HotFix 安装会向它找到的每一个引用数据属性文件添加扩展名 .bak，并为每个文件创建一个默认版本。

有关详细信息，请参阅《Informatica 10.2 HotFix 1 内容指南》。

发布任务 (10.2 HotFix 1)

Informatica PowerExchange 适配器

本节介绍版本 10.2 HotFix 1 中 Informatica 适配器的发布任务。

PowerExchange for Netezza

从版本 10.2 HotFix 1 开始，如果您希望为 Netezza 对象运行动态映射，则必须在以下位置添加 Netezza JDBC jar: <Informatica 安装目录>/externaljdbcjars。

有关详细信息，请参阅《*Informatica 10.2 HotFix 1 PowerExchange for Netezza 用户指南*》。

PowerExchange for Teradata Parallel Transporter API

从版本 10.2 HotFix 1 开始，如果您希望为 Teradata 对象运行动态映射，则必须在以下位置添加 Teradata JDBC jar: <Informatica 安装目录>/externaljdbcjars。

有关详细信息，请参阅《*Informatica 10.2 HotFix 1 PowerExchange for Teradata Parallel Transporter API 用户指南*》。

PowerCenter PowerExchange 适配器

本节介绍版本 10.2 HotFix 1 中 PowerCenter 适配器的发布任务。

PowerExchange for Amazon Redshift

从版本 10.2 HotFix 1 开始，PowerExchange for Amazon Redshift 包含以下发布任务：

- **群集节点类型**和**群集中的节点数**字段在连接属性中不可用。升级后，PowerExchange for Amazon Redshift 将计算暂存文件数，并忽略您在现有映射的上一个版本中指定的值。
您可以在目标会话属性下的**每个批处理的文件数**字段中指定文件数，以便为每一个批处理计算暂存文件数。
- 将 AWS SDK for Java 更新到版本 1.11.250。
- 将以下第三方 jar 更新到最新版本 2.9.5：
 - jackson-annotations
 - jackson-databind
 - jackson-core

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica 10.2 HotFix 1 PowerExchange for Amazon Redshift 用户指南*》。

PowerExchange for Amazon S3

从版本 10.2 HotFix 1 开始，会将 AWS SDK for Java 更新到版本 1.11.250。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica 10.2 HotFix 1 PowerExchange for Amazon S3 用户指南*》。

PowerExchange for Teradata Parallel Transporter API

从版本 10.2 HotFix 1 开始，您可以定义 Teradata PT API 用于写入数据的最大缓冲区大小（以 KB 为单位）。

当您从早期版本升级时，您必须使用 PowerCenter 存储库服务重新注册 TeradataPT.xml 插件，以启用最大缓冲区大小属性。注册后，可以在 Teradata 目标会话属性中定义最大缓冲区大小。

有关配置最大缓冲区大小的详细信息，请参阅《适用于 PowerCenter 的 Informatica 10.2 HotFix 1 PowerExchange for Teradata Parallel Transporter API 用户指南》。

第 24 章

新产品 (10.2)

本章包括以下主题：

- [PowerExchange 适配器, 249](#)

PowerExchange 适配器

Informatica PowerExchange 适配器

本节介绍版本 10.2 中新的 Informatica 适配器。

[PowerExchange for Microsoft Azure Data Lake Store](#)

从版本 10.2 开始，可以创建一个 Microsoft Azure 数据湖存储连接，以指定要包含在数据对象中的 Microsoft Azure 数据湖存储源和目标的位置。您可以在数据对象读取和写入操作中使用 Microsoft Azure 数据湖存储连接。可以在本地环境中或在 Hadoop 环境中的 Blaze 引擎上验证并运行映射。

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

第 25 章

新增功能 (10.2)

本章包括以下主题：

- [应用程序服务, 250](#)
- [大数据, 251](#)
- [命令行程序, 254](#)
- [数据类型, 262](#)
- [文档, 262](#)
- [Enterprise Information Catalog, 263](#)
- [Informatica Analyst, 266](#)
- [Intelligent Data Lake, 267](#)
- [Informatica Developer, 268](#)
- [Informatica 安装, 269](#)
- [Intelligent Streaming, 269](#)
- [Metadata Manager, 270](#)
- [PowerCenter, 271](#)
- [PowerExchange 适配器, 271](#)
- [规则规范, 275](#)
- [安全, 275](#)
- [转换语言, 276](#)
- [转换, 277](#)
- [工作流, 280](#)

应用程序服务

本节介绍版本 10.2 中新增的应用程序服务功能。

模型存储库服务

本节介绍版本 10.2 中新增的模型存储库服务功能。

从以前的版本导入对象

从版本 10.2 开始，可以使用 `infacmd` 将从 Informatica 10.1 或 10.1.1 模型存储库导出的对象升级到当前的元数据格式，然后将升级后的对象导入当前的 Informatica 版本。

有关详细信息，请参阅《*Informatica 10.2 Developer tool 指南*》中的“对象导入和导出”一章以及《*Informatica 10.2 命令参考*》中的“`infacmd mrs` 命令参考”一章。

大数据

本节介绍版本 10.2 中新增的大数据功能。

Big Data Management 安装

从版本 10.2 开始，数据集成服务会自动在群集上安装 Big Data Management 二进制文件。

运行映射时，数据集成服务会检查群集上是否具有二进制文件。如果这些文件不存在或未同步，数据集成服务会准备文件以进行传输。随后，它会通过 HDFS 上的 Informatica Hadoop 暂存目录将文件传输到分布式缓存。默认情况下，该暂存目录为 `/tmp`。此过程取代了在 Hadoop 群集上安装分发包的要求。

有关详细信息，请参阅《*Informatica Big Data Management 10.2 Hadoop 集成指南*》。

群集配置

群集配置是含有 Hadoop 群集配置信息的域中的一个对象。使用群集配置有助于数据集成服务将映射逻辑推送到 Hadoop 环境。

创建群集配置时，将导入配置站点文件中包含的群集配置属性。您可以从群集直接导入或从群集配置存档文件导入这些属性。此外还可以创建与群集配置关联的连接。

以前，为了让 Informatica 域能够与群集通信，需要运行 Hadoop Configuration Manager 实用工具来配置连接和其他信息。

有关群集配置的详细信息，请参阅《*Informatica Big Data Management 10.2 管理员指南*》中的“群集配置”一章。

处理层次结构数据

从版本 10.2 开始，可以在 Spark 引擎上运行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，Spark 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。

开发具有复杂端口、运算符和函数的映射来执行以下任务：

- 生成和修改层次结构数据。
- 将关系数据转换为层次结构数据。
- 将层次结构数据转换为关系数据。
- 将数据从一个复杂文件格式转换为另一个。

处理层次结构数据时，可以使用层次结构转换向导来简化映射开发任务。在以下情况下使用这些向导：

- 从一个或多个端口生成类型为结构的层次结构数据。
- 从两个转换中的端口生成类型为嵌套结构的层次结构数据。

- 从复杂端口中的层次结构数据中提取元素。
- 将复杂端口中的层次结构数据平展开来。

有关详细信息，请参阅《*Informatica Big Data Management 10.2 用户指南*》中的“在 Spark 引擎上处理层次结构数据”一章。

Spark 引擎上的有状态计算

从版本 10.2 开始，可以使用表达式转换中的窗口函数在 Spark 引擎上执行有状态计算。窗口函数会对一组行执行操作，并为每个输入行计算单个返回值。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。
- 计算一组行的累计总和。
- 计算一组行的累计平均值。

有关详细信息，请参阅《*Big Data Management 10.2 用户指南*》中的“Spark 引擎上的有状态计算”一章。

数据集成服务队列

从版本 10.2 开始，如果同时部署多个映射作业或工作流映射任务，数据集成服务将在持久队列中对作业进行排队，并在资源可用时运行作业。您可以在 Administrator 工具的“监视”选项卡上查看映射作业的当前状态。

默认情况下所有队列均为持久队列。如果数据集成服务节点意外关闭，则当数据集成服务进行故障转移时，队列不会进行故障转移。队列会保留在数据集成服务计算机上，并且数据集成服务会在重新启动后恢复处理队列。

默认情况下，每个队列一次可以容纳 1 万个作业。队列已满时，数据集成服务将拒绝作业请求并将其标记为失败。数据集成服务开始运行队列中的作业时，您可以部署其他作业。

有关详细信息，请参阅《*Informatica Big Data Management 10.2 管理员指南*》中的“队列”一章。

Blaze 作业监视

从版本 10.2 开始，可以配置主机和端口号，以启动 Hadoop 连接属性中指定的 Blaze 作业监视应用程序。默认值为 <主机名>:9080。如果未配置主机名，Blaze 引擎会根据字母顺序，使用群集中的第一个节点。

有关详细信息，请参阅《*Big Data Management 10.2 用户指南*》中的“连接”一章。

适用于 Hadoop 集成的数据集成服务属性

从版本 10.2 开始，数据集成服务添加了将域与 Hadoop 环境集成所需的属性。

下表介绍了新的属性：

属性	说明
Hadoop 暂存目录	用来存储数据集成服务推送的 Informatica Hadoop 二进制文件以及处理期间的临时文件的 HDFS 目录。默认值为 /tmp。
Hadoop 暂存用户	数据集成服务用户为空时为必填项。在 Hadoop 暂存目录上执行操作的 HDFS 用户。用户需要对 Hadoop 暂存目录具有写入权限。默认为数据集成服务用户。
自定义 Hadoop OS 路径	<p>与 Hadoop 操作系统兼容的 Informatica hadoop 二进制文件的本地路径。当 Hadoop 群集和数据集成服务位于不同的受支持操作系统上时为必填项。</p> <p>在托管数据集成服务的计算机上，为 Hadoop 群集下载并提取 Informatica 二进制文件。数据集成服务使用此目录中的二进制文件将域与 Hadoop 群集集成。</p> <p>数据集成服务可以同步下列操作系统：</p> <ul style="list-style-type: none">- SUSE 11 和 Redhat 6.5 <p>所做更改将在再次应用数据集成服务后生效。</p>

群集集成发生更改后，以下属性将从数据集成服务中删除：

- Hadoop 中的 Informatica Home 目录
- Hadoop 分布目录

有关详细信息，请参阅《*Informatica 10.2 Hadoop 集成指南*》。

Sqoop

从版本 10.2 开始，如果使用 Sqoop 数据对象，则可以使用以下专用 Sqoop 连接器在 Spark 引擎上运行映射：

- 由 Teradata 提供技术支持的 Cloudera 连接器
- Hortonworks Connector for Teradata

这些专用连接器使用本地协议来连接到 Teradata 数据库。

有关详细信息，请参阅《*Informatica Big Data Management 10.2 用户指南*》。

在 Amazon EMR 群集中自动扩展

从版本 10.2 开始，Big Data Management 新增了对 Spark 映射的支持，可以利用在 Amazon EMR 群集中自动扩展的优势。

利用自动扩展功能，EMR 群集管理员可构建基于阈值的规则，并利用这些规则来增加和减少群集任务和核心节点。Big Data Management 对在支持自动扩展的 EMR 群集上支持 Spark 映射进行了认证。

Blaze 引擎上的转换支持

从版本 10.2 开始，Blaze 引擎还支持以下转换

- 更新策略。支持在所有列属于带桶的 ORC 的目标。

有关详细信息，请参阅《*Informatica Big Data Management 10.2 用户指南*》中的“Hadoop 环境中的映射对象”一章。

Blaze 引擎的 Hive 功能

从版本 10.2 开始，在 Blaze 引擎上运行的映射可以读取和写入到带桶且已排序的目标。

有关如何为 Blaze 引擎配置映射的信息，请参阅《*Informatica Big Data Management 10.2 用户指南*》中的“Hadoop 环境中的映射”一章。

Spark 引擎上的转换支持

从版本 10.2 开始，Spark 引擎有限制地支持以下转换：

- 规范器
- 等级
- 更新策略

从版本 10.2 开始，Spark 引擎还支持以下转换：

- 查找支持筛选器、汇总器、路由器、表达式和更新策略转换中的未连接查找。

有关详细信息，请参阅《*Informatica Big Data Management 10.2 用户指南*》中的“Hadoop 环境中的映射对象”一章。

Spark 引擎的 Hive 功能

从版本 10.2 开始，在 Spark 引擎上运行的映射支持以下功能：

- 在 Amazon S3 桶中读取和写入至 Hive 资源
- 读取和写入至事务性 Hive 表
- 读取和写入至使用细粒度 SQL 授权保护的 Hive 表列

有关如何为 Spark 引擎配置映射的信息，请参阅《*Informatica Big Data Management 10.2 用户指南*》中的“Hadoop 环境中的映射”一章。

命令行程序

本节介绍版本 10.2 中的新命令。

infacmd cluster 命令

cluster 是一个新的 infacmd 插件，用来对群集配置执行操作。

下表介绍了新的 infacmd cluster 命令：

命令	说明
clearConfigurationProperties	在群集配置集中清除覆盖的属性值。
createConfiguration	从 XML 文件或远程群集管理器创建新的群集配置。
deleteConfiguration	从域中删除群集配置。

命令	说明
exportConfiguration	将群集配置导出到压缩文件或合并 XML 文件。
listAssociatedConnections	按类型列出与指定群集配置相关联的连接。
listConfigurationGroupPermissions	列出组对群集配置的权限。
listConfigurationSets	列出群集配置中的配置集。
listConfigurationProperties	列出群集配置集中的配置属性。
listConfigurations	列出群集配置名称。
listConfigurationUserPermissions	列出用户对群集配置的权限。
refreshConfiguration	从 XML 文件或远程群集管理器刷新群集配置。
setConfigurationPermissions	删除先前的权限后，将群集配置的权限设置为某个用户或某个组。
setConfigurationProperties	在群集配置集中设置覆盖的属性值。

有关详细信息，请参阅《*Informatica 10.2 命令参考*》中的“infacmd cluster 命令参考”一章。

infacmd dis 选项

下表介绍了 infacmd UpdateServiceOptions 新增的数据集成服务选项：

命令	说明
ExecutionOptions.MaxHadoopBatchExecutionPoolSize	可以并发运行的已部署 Hadoop 作业的最大数量。
ExecutionOptions.MaxNativeBatchExecutionPoolSize	每个数据集成服务进程可以并发运行的已部署本机作业的最大数量。
ExecutionOptions.MaxOnDemandExecutionPoolSize	可以并发运行的按需作业的最大数量。作业包括数据预览、剖析作业、REST 和 SQL 查询、Web 服务请求以及从 Developer tool 运行的映射。
WorkflowOrchestrationServiceOptions.MaxWorkerThreads	在工作流中的一对非独占网关之间运行并行任务时，数据集成服务可使用的最大线程数。默认值为 10。 如果非独占网关之间的任务数大于该最大值，数据集成服务将根据指定的值分批运行任务。

有关详细信息，请参阅《*Informatica 10.2 命令参考*》中的“infacmd dis 命令参考”一章。

infacmd ipc 命令

下表介绍了 infacmd ipc 命令新增的一个选项：

命令	说明
genReuseReportFromPC	包含以下新选项： -BlockSize：可选。要对其运行 infacmd ipc genReuseReportFromPC 命令的映射数。

有关详细信息，请参阅《Informatica 10.2 命令参考》中的“infacmd ipc 命令参考”一章。

infacmd isp 命令

下表介绍了对 infacmd isp 命令的更改：

命令	说明
createConnection	定义连接和连接选项。 添加、更改和删除了 Hadoop 连接选项。请参阅 infacmd isp createConnection。
getDomainSamlConfig	已从 getSamlConfig 更名。 返回为安全断言标记语言 (SAML) 身份验证设置的 cst 选项的值。指定 Active Directory Federation Services (AD FS) 主机系统时钟与主网关节点上系统时钟之间允许的时间差。
getUserActivityLog	返回用户活动日志数据，其中现在包括来自 Informatica 客户端的成功和不成功的用户登录尝试。 用户活动数据包括来自 Informatica 客户端的每个登录尝试的下列属性： - 应用程序名称 - 应用程序版本 - 应用程序主机的主机名或 IP 地址 如果客户端对登录请求设置了自定义属性，则数据包括这些自定义属性。
listConnections	按类型列出连接名称。您可以按所有连接类型列出或者按一个连接类型筛选结果。 该命令现在可使用 -ct 选项。使用 -ct 选项来筛选连接类型。
purgeLog	清除有关许可证使用量的日志事件和数据库记录。 -lu 选项现在已过时。
SwitchToGatewayNode	配置 SAML 身份验证时可使用以下新增选项： - asca。将身份提供程序断言签名证书导入用于 SAML 身份验证的信任库文件时指定的别名。 - saml。在 Informatica 域中启用或禁用 SAML 身份验证。 - std。包含在域内的网关节点上使用 SAML 身份验证所需的自定义信任库文件的目录。 - stp。用于 SAML 身份验证的自定义信任库密码。

有关详细信息，请参阅《Informatica 10.2 命令参考》中的“infacmd isp 命令参考”一章。

infacmd isp createConnection

本节列出了版本 10.2 中为 infacmd isp createConnection 属性新增、更改和删除的 Hadoop 连接选项。

Hadoop 连接选项

下表介绍了版本 10.2 中新增的 Hadoop 连接选项：

选项	说明
clusterConfigId	与 Hadoop 群集关联的群集配置 ID。
blazeJobMonitorURL	Blaze 作业监视器的主机名和端口号。
rejDirOnHadoop	启用 hadoopRejDir。用于指定运行映射时将拒绝文件移动到的位置。
hadoopRejDir	运行映射时，数据集成服务将拒绝文件移动到的远程目录。使用 rejDirOnHadoop 启用拒绝目录。
sparkEventLogDir	Spark 引擎用来记录事件的目录的可选 HDFS 文件路径。
sparkYarnQueueName	Spark 引擎用来指定群集上的可用资源的 YARN 调度程序队列名称。

下表介绍了版本 10.2 中重命名的 Hadoop 连接选项：

当前名称	以前名称	说明
blazeYarnQueueName	cadiAppYarnQueueName	由在群集上指定可用资源的 Blaze 引擎使用的 YARN 计划程序队列名称。该名称区分大小写。
blazeExecutionParameterList	cadiExecutionParameterList	对 Blaze 引擎唯一的自定义属性。
blazeMaxPort	cadiMaxPort	Blaze 引擎的端口号范围的最大值。
blazeMinPort	cadiMinPort	Blaze 引擎的端口号范围的最小值。
blazeUserName	cadiUserName	Blaze 服务和 Blaze 服务日志的所有者。
blazeStagingDirectory	cadiWorkingDirectory	Blaze 引擎用于存储临时文件的目录的 HDFS 文件路径。
hiveStagingDatabaseName	databaseName	Hive 暂存表的命名空间。
impersonationUserName	hiveUserName	Hadoop 模拟用户：数据集成服务为在 Hadoop 环境中运行映射而模拟的用户名。
sparkStagingDirectory	SparkHDFSStagingDir	Spark 引擎用于存储运行作业的临时文件的目录的 HDFS 文件路径。

下表介绍了从 UI 中删除并导入到群集配置中的 Hadoop 连接选项：

选项	说明
RMAddress	Hadoop 中提交资源的请求或生成 YARN 应用程序的服务。 作为属性 <code>yarn.resourcemanager.address</code> 导入到群集配置中。
defaultFSURI	用于访问默认的 Hadoop 分布式文件系统的 URI。 作为属性 <code>fs.defaultFS</code> 或 <code>fs.default.name</code> 导入到群集配置中。

下表介绍了版本 10.2 中弃用，进而在 UI 中不再可用的 Hadoop 连接选项：

选项	说明
metastoreDatabaseDriver*	JDBC 数据存储的驱动程序类名称。
metastoreDatabasePassword*	元存储用户名的密码。
metastoreDatabaseURI*	用于在本地元存储设置下访问数据存储的 JDBC 连接 URI。
metastoreDatabaseUserName*	元存储数据库用户名。
metastoreMode*	连接到远程元存储还是本地元存储的控制机制。
remoteMetastoreURI*	用于在远程元存储设置下访问元数据的元存储 URI。 此属性作为属性 <code>hive.metastore.uris</code> 导入到群集配置中。
jobMonitoringURL	MapReduce JobHistory 服务器的 URL。
* 这些属性在版本 10.2 中已弃用。升级到 10.2 后，在以前的版本中设置的属性值将保存在存储库中，但它们不会出现在连接属性中。	

以下属性已删除。即便它们出现在连接字符串中也不起作用：

- `hadoopClusterInfoExecutionParametersList`
- `passThroughSecurityEnabled`
- `hiverserver2Enabled`
- `hiveInfoExecutionParametersList`
- `cadiPassword`
- `sparkMaster`
- `sparkDeployMode`

HBase 连接

下表介绍了从连接中删除并导入到群集配置中的 HBase 连接选项：

属性	说明
ZOOKEEPERHOSTS	托管 ZooKeeper 服务器的计算机的名称。
ZOOKEEPERPORT	托管 ZooKeeper 服务器的计算机的端口号。
ISKERBEROSENABLED	启用 Informatica 域可与使用 Kerberos 身份验证的 HBase 主服务器或区域服务器进行通信。
hbaseMasterPrincipal	HBase 主服务器的服务主体名称 (SPN)。
hbaseRegionServerPrincipal	HBase 区域服务器的服务主体名称 (SPN)。

Hive 连接

下表介绍了从连接中删除并导入到群集配置中的 Hive 连接选项：

属性	说明
defaultFSURI	用于访问默认的 Hadoop 分布式文件系统的 URI。
jobTrackerURI	Hadoop 中将 MapReduce 任务提交到群集中特定节点的服务。
hiveWarehouseDirectoryOnHDFS	仓库（此仓库对群集来说属于本地仓库）的默认数据库的绝对 HDFS 文件路径。
metastoreExecutionMode	连接到远程元存储还是本地元存储的控制机制。
metastoreDatabaseURI	用于在本地元存储设置下访问数据存储的 JDBC 连接 URI。
metastoreDatabaseDriver	JDBC 数据存储的驱动程序类名称。
metastoreDatabaseUserName	元存储数据库用户名。
metastoreDatabasePassword	元存储用户名的密码。
remoteMetastoreURI	用于在远程元存储设置下访问元数据的元存储 URI。 此属性作为属性 hive.metastore.uris 导入到群集配置中。

适用于 MapR-DB 的 HBase 连接选项

ISKERBEROSENABLED 连接选项已过时并导入到群集配置中。

infacmd mrs 命令

下表介绍了新的 infacmd mrs 命令：

命令	说明
manageGroupPermissionOnProject	管理组对多个项目的权限。
manageUserPermissionOnProject	管理用户对多个项目的权限。
upgradeExportedObjects	将从以前版本 Informatica 导出到 .xml 文件的对象升级到当前元数据格式。该命令会生成一个包含已升级对象的 .xml 文件。

有关详细信息，请参阅《Informatica 10.2 命令参考》中的“infacmd mrs 命令参考”一章。

infacmd ms 命令

下表介绍了新的 infacmd ms 命令：

命令	说明
GetMappingStatus	按作业 ID 获取映射作业的当前状态。

有关详细信息，请参阅《Informatica 10.2 命令参考》中的“infacmd ms 命令参考”一章。

infacmd wfs 命令

下表介绍了新的 infacmd wfs 命令：

命令	说明
completeTask	完成指定的人工任务实例。
delegateTask	将人工任务实例的所有权分配给用户或组。
listTasks	列出符合指定的筛选条件的人工任务实例。
releaseTask	从当前所有者释放一个人工任务实例，并将任务实例的所有权返回给 workflow 配置标识的业务管理员。
startTask	将人工任务实例的状态更改为 IN_PROGRESS。

有关详细信息，请参阅《Informatica 10.2 命令参考》中的“infacmd wfs 命令参考”一章。

infasetup 命令

下表介绍了对 infasetup 命令的更改：

命令	说明
DefineDomain	配置安全断言标记语言 (SAML) 身份验证时可使用以下新增选项： <ul style="list-style-type: none">- asca。将身份提供程序断言签名证书导入用于 SAML 身份验证的信任库文件时指定的别名。- cst。Active Directory Federation Services (AD FS) 主机系统时钟与主网关节点上系统时钟之间允许的时间差。- std。包含在域内的网关节点上使用 SAML 身份验证所需的自定义信任库文件的目录。- stp。用于 SAML 身份验证的自定义信任库密码。
DefineGatewayNode	配置 SAML 身份验证时可使用以下新增选项： <ul style="list-style-type: none">- asca。将身份提供程序断言签名证书导入用于 SAML 身份验证的信任库文件时指定的别名。- saml。在 Informatica 域中启用或禁用 SAML 身份验证。- std。包含在域内的网关节点上使用 SAML 身份验证所需的自定义信任库文件的目录。- stp。用于 SAML 身份验证的自定义信任库密码。
UpdateDomainSamlConfig	已从 UpdateSamlConfig 更名。 配置 SAML 身份验证时可使用以下新增选项： <ul style="list-style-type: none">- cst。AD FS 主机系统时钟与主网关节点上系统时钟之间允许的时间差。
UpdateGatewayNode	配置 SAML 身份验证时可使用以下新增选项。 <ul style="list-style-type: none">- asca。将身份提供程序断言签名证书导入用于 SAML 身份验证的信任库文件时指定的别名。- saml。在 Informatica 域中启用或禁用 SAML 身份验证。- std。包含在域内的网关节点上使用 SAML 身份验证所需的自定义信任库文件的目录。- stp。用于 SAML 身份验证的自定义信任库密码。

有关详细信息，请参阅《Informatica 10.2 命令参考》中的“infasetup 命令参考”一章。

pmrep 命令

下表介绍了新增的 pmrep 命令：

命令	说明
CreateQuery	在存储库中创建查询。
DeleteQuery	从存储库中删除查询。

下表介绍了对 pmrep 命令的更新：

命令	说明
CreateConnection	包含以下更新后的选项： -w。允许在密码选项中使用一个参数。
ListObjectDependencies	包含以下更新后的选项： -o。对象类型列表包括查询和 deploymentgroup。
UpdateConnection	包含以下更新的选项： -w。允许在密码选项中使用一个参数。 -x。如果在密码中使用参数，则禁止使用密码参数。

有关详细信息，请参阅《Informatica 10.2 命令参考》中的“pmrep 命令参考”一章。

数据类型

本节介绍版本 10.2 中新增的数据类型功能。

Informatica 数据类型

本节介绍 Developer tool 中的新数据类型。

复杂数据类型

从版本 10.2 开始，某些转换支持在 Spark 引擎上运行的映射中的复杂数据类型。

下表介绍了可以在转换中使用的复杂数据类型：

复杂数据类型	说明
array	包含元素的有序集合。数组中的所有元素必须具有相同的数据类型。元素可以是原始或复杂数据类型。
map	包含键-值对的无序集合。键部分必须是原始数据类型。值部分可以是原始或复杂数据类型。
struct	包含不同数据类型的元素的集合。元素可以是原始或复杂数据类型。

有关详细信息，请参阅《Informatica Big Data Management 10.2 用户指南》中的“数据类型参考”附录。

文档

本节介绍版本 10.2 中指南的新增和更新内容。

Informatica 文档包含以下更改：

Informatica Big Data Management 安全指南

从版本 10.2 开始, 《*Informatica Big Data Management 安全指南*》更名为 《*Informatica Big Data Management 管理员指南*》。它包含 Big Data Management 的安全信息及新增的管理员任务。

有关详细信息, 请参阅 《*Informatica Big Data Management 10.2 管理员指南*》。

Informatica Big Data Management 安装和升级指南

从版本 10.2 开始, 《*Informatica Big Data Management 安装和升级指南*》更名为 《*Informatica Big Data Management Hadoop 集成指南*》。从版本 10.2 开始, 数据集成服务可以自动将 Big Data Management 二进制文件安装到 Hadoop 群集, 以便将域与群集集成。该指南中的集成任务不包括分发包安装操作。

有关详细信息, 请参阅 《*Informatica Big Data Management 10.2 Hadoop 集成指南*》。

Informatica Catalog 管理员指南

从版本 10.2 开始, 《*Informatica Live Data Map 管理员指南*》更名为 《*Informatica Catalog 管理员指南*》。

有关详细信息, 请参阅 《*Informatica Catalog Administrator Guide 10.2*》。

适用于 Enterprise Information Catalog 的 Informatica Administrator 参考

从版本 10.2 开始, 《*适用于 Live Data Map 的 Informatica Administrator 参考*》更名为 《*适用于 Enterprise Information Catalog 的 Informatica Administrator 参考*》。

有关详细信息, 请参阅 《*Informatica Administrator Reference for Enterprise Information Catalog 10.2*》。

Informatica Enterprise Information Catalog 自定义元数据集成指南

从版本 10.2 开始, 可以使用 Enterprise Information Catalog 将自定义元数据获取到目录中。有关详细信息, 请参阅 《*Informatica Enterprise Information Catalog 10.2 自定义元数据集成指南*》这本新指南。

Informatica Enterprise Information Catalog 安装和配置指南

从版本 10.2 开始, 《*Informatica Live Data Map 安装和配置指南*》更名为 《*Informatica Enterprise Information Catalog 安装和配置指南*》。

有关详细信息, 请参阅 《*Informatica Enterprise Information Catalog 10.2 安装和配置指南*》。

Informatica Enterprise Information Catalog REST API 参考

从版本 10.2 开始, 可以使用 Enterprise Information Catalog 开放的 REST API。有关详细信息, 请参阅 《*Informatica Enterprise Information Catalog 10.2 REST API 参考*》这本新指南。

Informatica Enterprise Information Catalog 升级指南

从版本 10.2 开始, 《*从 <x> 版本升级 Informatica Live Data Map*》重命名为 《*从 10.1、10.1.1、10.1.1 HF1 和 10.1.1 Update 2 升级 Informatica Enterprise Information Catalog*》。

有关详细信息, 请参阅 《*从版本 10.1、10.1.1、10.1.1 HF1 和 10.1.1 Update 2 升级 Informatica Enterprise Information Catalog 指南*》。

Enterprise Information Catalog

本节介绍版本 10.2 中新增的 Enterprise Information Catalog 功能。

新数据源

从版本 10.2 开始，Informatica Enterprise Information Catalog 允许从新数据源提取元数据。

您可以在 Informatica Catalog Administrator 中创建资源，以便从以下数据源中提取元数据：
Apache Atlas

Hadoop 元数据框架。

Azure Microsoft SQL 数据仓库

用来处理大量数据的基于云的关系数据库。

Azure Microsoft SQL Server

受管云数据库。

Azure WASB 文件系统

用于将数据加载到 Azure blob 的 Windows Azure 存储 blob 接口。

Erwin

数据建模工具。

Informatica Axon

企业数据治理解决方案。

有关新资源的详细信息，请参阅《*Informatica Catalog Administrator Guide 10.2*》。

自定义扫描器框架

从版本 10.2 开始，可以将自定义元数据获取到目录中。

自定义元数据是您定义的元数据。您可以定义自定义模型，创建自定义资源类型，并创建自定义资源，以便从自定义数据源获取自定义元数据。您可以使用自定义元数据集成从 Enterprise Information Catalog 并未提供模型的自定义数据源提取和获取元数据。

有关自定义元数据集成的详细信息，请参阅《*Informatica Enterprise Information Catalog 10.2 自定义元数据集成指南*》。

REST API

从版本 10.2 开始，可以使用 Informatica Enterprise Information Catalog REST API 来访问和配置与数据源关联的对象和模型相关的功能。

使用 REST API 可以检索与数据源关联的对象和模型相关的信息。此外，还可以创建、更新或删除与模型和对象相关的实体，例如属性、关联和类。

有关非结构化文件源的详细信息，请参阅《*Informatica Enterprise Information Catalog 10.2 REST API 参考*》。

复合数据域

从版本 10.2 开始，可以创建复合数据域。复合数据域是数据域或可以使用规则链接的其他复合数据域的集合。使用复合数据域可以在一个数据源的多个架构中搜索所需的有关某个实体的详细信息。

在 Catalog Administrator 中为资源创建并启用数据域发现后，可以在“资产详细信息”视图中查看表资产的复合数据域。此外还可以在“资产详细信息”视图中搜索复合数据域并查看复合数据域的详细信息。

有关复合数据域的详细信息，请参阅《*Informatica Enterprise Information Catalog 10.2 用户指南*》中的“查看资产”一章，并参阅《*Informatica Catalog Administrator Guide 10.2*》中的“管理复合数据域”一章。

数据域

本节介绍与 Enterprise Information Catalog 中的数据域有关的新功能。

定义数据域

从版本 10.2 开始，可以在创建数据域时配置以下附加选项：

- 使用引用表、规则和正则表达式创建数据规则或列规则。
- 为数据域匹配使用最小遵从性百分比或最小遵从行数。
- 使用自动接受选项可在当数据域匹配超过配置的自动接受百分比时，在 Enterprise Information Catalog 中自动接受数据域。

有关 Catalog Administrator 中数据域的详细信息，请参阅《*Informatica Catalog Administrator Guide 10.2*》中的“管理数据域”一章。

配置数据域

从版本 10.2 开始，可以在创建或编辑源时，为数据域匹配使用预定义的值或者输入遵从性值。

有关数据域和资源的详细信息，请参阅《*Informatica Catalog Administrator Guide 10.2*》中的“管理资源”一章。

数据域权限

从版本 10.2 开始，可以在 Informatica Administrator 中配置用来在 Catalog Administrator 中查看、创建、编辑或删除数据域或数据域组的**域管理：管理 - 查看域和域组**和**域管理：管理 - 编辑域和域组**特权。

有关特权的详细信息，请参阅《*Informatica Administrator Reference for Enterprise Information Catalog 10.2*》中的“特权和角色”一章。

数据域内容管理

从版本 10.2 开始，如果数据域匹配百分比超过在 Catalog Administrator 中配置的自动接受百分比，Enterprise Information Catalog 会自动接受数据域。

有关数据域内容管理的详细信息，请参阅《*Informatica Enterprise Information Catalog 10.2 用户指南*》中的“查看资产”一章。

导出和导入自定义属性

从版本 10.2 开始，可以将资源中配置的自定义属性导出到 CSV 文件中，并将 CSV 文件重新导入 Enterprise Information Catalog。您可以使用导出的 CSV 文件同时将自定义属性值分配给多个资产。

有关自定义属性导出和导入的详细信息，请参阅《*Informatica Enterprise Information Catalog 10.2 用户指南*》中的“查看资产”一章。

将富文本作为自定义属性值

从版本 10.2 开始，可以编辑自定义属性以将多个富文本字符串指定为属性值。

有关为资产分配自定义属性值的详细信息，请参阅《*Informatica Enterprise Information Catalog 10.2 用户指南*》中的“查看资产”一章。

转换逻辑

从版本 10.2 开始，可以在沿袭和影响视图中查看资产的转换逻辑。沿袭和影响视图会显示包含转换的资产的转换逻辑。转换视图会显示数据结构（如表和列）的转换逻辑。该视图还会显示各种类型的转换，如筛选器、联接器、查找、表达式、排序器、联合和聚合。

有关转换逻辑的详细信息，请参阅《*Informatica Enterprise Information Catalog 10.2 用户指南*》中的“查看沿袭和影响”一章。

非结构化文件类型

从版本 10.2 开始，可以为数据源中的所有行在非结构化文件类型和扩展非结构化格式上运行**数据域发现**配置文件或**列配置文件**和**数据域发现**配置文件。非结构化文件类型包括压缩文件、电子邮件格式、网页文件、Microsoft Excel、Microsoft PowerPoint、Microsoft Word 和 PDF。扩展非结构化格式包括 mp3、mp4、bmp 和 jpg。

有关非结构化文件类型的详细信息，请参阅《*Informatica Catalog Administrator Guide 10.2*》中的“管理资源”一章。

值频率

配置和查看值频率

从版本 10.2 开始，可以在 Catalog Administrator 中启用值频率以及列数据相似性，从而计算数据源中值的频率。在 Catalog Administrator 中对数据源运行值频率后，可以在**资产详细信息**视图中查看视图列、表列、CSV 字段、XML 文件字段和 JSON 文件数据资产的值频率。

有关配置值频率的详细信息，请参阅《*Informatica Catalog Administrator Guide 10.2*》中的“Catalog Administrator 概念”一章。要查看数据资产的值频率，请参阅《*Informatica Enterprise Information Catalog 10.2 用户指南*》中的“查看资产”一章。

在 Enterprise Information Catalog 中查看值频率的特权

从版本 10.2 开始，需要具备以下权限和特权才能查看数据资产的值频率：

- 数据资产的读取权限。
- **数据特权：查看数据**特权。
- **数据特权：查看敏感数据**特权。

有关权限和特权的详细信息，请参阅《*Informatica Administrator Reference for Enterprise Information Catalog 10.2*》中的“权限概述”和“特权和角色概述”一章。

对 Azure HDInsight 部署的支持

从版本 10.2 开始，可以在 Azure HDInsight Hadoop 发行版上部署 Enterprise Information Catalog。

有关详细信息，请参阅《*Informatica Enterprise Information Catalog 10.2 安装和配置指南*》中的“创建应用程序服务”一章。

Informatica Analyst

本节介绍版本 10.2 中新增的 Analyst 工具功能。

配置文件

本节介绍为配置文件和结果卡新增的功能。

规则规范

从版本 10.2 开始，可以在 Analyst 工具中配置规则规范，并在列配置文件中使用时规则规范。

有关在列配置文件中使用时规则规范的详细信息，请参阅《*Informatica 10.2 数据发现指南*》中的“Informatica Analyst 中的规则”一章。

Intelligent Data Lake

本节介绍版本 10.2 中新增的 Intelligent Data Lake 功能。

使用可视化及 Apache Zeppelin 验证和评估数据

从版本 10.2 开始，在发布数据后，可以直观地验证数据，以确保数据内容和质量都适合进行分析。然后，可以选择修复脚本，从而支持“准备-发布-验证”这一迭代流程。

智能数据湖使用 Apache Zeppelin 以可视化笔记本（其中包含图形和图表）的形式查看工作表。有关 Apache Zeppelin 的详细信息，请参阅 Apache Zeppelin 文档。使用 Zeppelin 功能对数据进行可视化处理时，可以查看不同列之间的关系并创建多个图表和图形。

发布数据资产后第一次打开可视化笔记本时，智能数据湖使用 CLAIRE 引擎以用户创建的数字列的直方图形式创建智能可视化建议。

有关可视化笔记本的详细信息，请参阅《*Informatica Intelligent Data Lake 10.2 用户指南*》中的“使用 Apache Zeppelin 可视化功能验证和评估数据”一章。

在数据预览期间使用筛选器评估数据

从版本 10.2 开始，可以在数据预览过程中筛选数据，从而更好地评估数据资产。您可以为多个字段添加筛选器，并组合应用此类筛选器。筛选条件取决于数据类型。而且，您可以查看在对字符串值剖析期间发现的列值频率（如果可用）。

有关详细信息，请参阅《*Informatica Intelligent Data Lake 10.2 用户指南*》中的“发现数据”一章。

增强的脚本面板布局

从版本 10.2 开始，可以在准备数据的过程中看到一个专门用来配置脚本步骤的面板。脚本步骤简明扼要，而且与彩色代码相对应，以指示函数名称、涉及的列和输入源。您可以编辑这些步骤或将其删除。您还可以返回到脚本中的特定步骤，并查看数据状态。您可以从源刷新脚本。您还可以看到一个单独的成分面板，其中会显示此工作表使用的源。

有关详细信息，请参阅《*Informatica Intelligent Data Lake 10.2 用户指南*》中的“准备数据”一章。

应用数据质量规则

从版本 10.2 开始，在准备数据的同时，可以使用在交互式数据准备过程中可用的预构建规则。这些规则是使用 Informatica Developer 或 Informatica Analyst 工具创建的。如果拥有 Big Data Quality 许可证，则还可使用供 Intelligent Data Lake 用户使用的数千个预构建规则。使用预构建规则可有效促进业务部门和 IT 部门之间的协作，还可提高规则和知识的可重用性、使用一致性和可扩展性。

有关详细信息，请参阅《*Informatica Intelligent Data Lake 10.2 用户指南*》中的“准备数据”一章。

在数据预览和工作表视图中查看数据资产的业务术语

从版本 10.2 开始，可以在数据预览以及数据准备期间查看与数据资产列相关的业务术语。

有关详细信息，请参阅《*Informatica Intelligent Data Lake 10.2 用户指南*》中的“发现数据”一章。

为带分隔符的文件准备数据

从版本 10.2 开始，您能够以数据分析师身份，对已在湖中的带分隔符 HDFS 文件进行清理、转换、合并、聚合和执行其他操作。您可以先预览这些文件，然后再将它们添加到项目中。接着，您可以配置这些资产的采样设置，并对其执行数据准备操作。

有关详细信息，请参阅《*Informatica Intelligent Data Lake 10.2 用户指南*》中的“准备数据”一章。

在联接工作表中编辑联接

从版本 10.2 开始，可以编辑现有联接工作表的联接条件，如联接键、联接类型（如内连接和外联接）。

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

编辑数据准备的采样设置

从版本 10.2 开始，可以在准备数据资产期间编辑采样设置。您可以更改为采样选择的列，编辑选定的筛选器，以及更改采样条件。

有关详细信息，请参阅《*Informatica Intelligent Data Lake 10.2 用户指南*》中的“准备数据”一章。

数据湖中支持多个 Enterprise Information Catalog 资源

从版本 10.2 开始，可以配置多个 Enterprise Information Catalog 资源，以便用户可以使用湖中的所有类型资产和所有适用的 Hive 架构。

为数据准备服务存储库使用 Oracle

从版本 10.2 开始，现在可以为数据准备服务存储库使用 Oracle 11gR2 和 12c。

提高了数据准备服务的可扩展性

从版本 10.2 开始，可以通过使用具有多个数据准备服务节点的数据准备服务网格来确保横向可扩展性。更高的可扩展性有助于在数据量增加以及用户数增加时支持高性能、交互式数据准备。

Informatica Developer

本节介绍版本 10.2 中新增的 Developer tool 功能。

非关系数据对象

从版本 10.2 开始，可以一次导入多个非关系数据对象。

有关详细信息，请参阅《*Informatica 10.2 Developer tool 指南*》中的“物理数据对象”一章。

配置文件

本节介绍为配置文件和结果卡新增的功能。

规则规范

从版本 10.2 开始，可以在 Developer tool 中创建列配置文件时使用规则规范。要使用规则规范，请从规则规范生成 Mapplet，并将 Mapplet 作为规则进行验证。

有关在列配置文件中使用时规则规范的详细信息，请参阅《*Informatica 10.2 数据发现指南*》中的“Informatica Developer 中的规则”一章。

Informatica 安装

本节介绍版本 10.2 中新增的安装功能。

Informatica Upgrade Advisor

从版本 10.2 开始，可以运行 Informatica Upgrade Advisor 来验证服务，并在执行升级之前检查域中过时的服务、支持的数据库和支持的操作系统。

有关升级顾问的详细信息，请参阅《*Informatica 升级指南*》。

Intelligent Streaming

本节介绍版本 10.2 中新增的 Intelligent Streaming 功能。

CSV 格式

从版本 10.2 开始，流式映射可以读取和写入 CSV 格式的数据。

有关 CSV 格式的详细信息，请参阅《*Informatica Intelligent Streaming 10.2 用户指南*》中的“流式映射中的源和目标”一章。

数据类型

从版本 10.2 开始，流式映射可以读取、处理和写入分层数据。您可以使用数组、结构和映射复杂数据类型来处理分层数据。

有关详细信息，请参阅《*Informatica Intelligent Streaming 10.2 用户指南*》中的“流式映射中的源和目标”一章。

连接

从版本 10.2 开始，可以在流映射中使用以下新的消息连接：

- AmazonKinesis。访问作为源的 Amazon Kinesis 流或作为目标的 Amazon Kinesis Firehose。您可以在 Developer tool 中或通过 infacmd 创建和管理 AmazonKinesis 连接。
- MapRStreams。访问作为目标的 MapRStreams。您可以在 Developer tool 中或通过 infacmd 创建和管理 MapRStreams 连接。

有关详细信息，请参阅《*Informatica Intelligent Streaming 10.2 用户指南*》中的“连接”一章。

传递映射

从版本 10.2 开始，可以在流映射中直接将有效负载格式从源传递至目标。

您可以用二进制格式对列进行投影，从而将有效负载按其原始格式从源传递到目标，或传递不受支持的有效负载格式。

有关详细信息，请参阅《*Informatica Intelligent Streaming 10.2 用户指南*》中的“流式映射中的源和目标”一章。

源和目标

从版本 10.2 开始，可以创建以下新的物理数据对象：

- AmazonKinesis。表示 Amazon Kinesis 流或 Amazon Kinesis Firehose 交付流中的数据。
- MapRStreams。表示 MapR 流中的数据。

有关详细信息，请参阅《*Informatica Intelligent Streaming 10.2 用户指南*》中的“流式映射中的源和目标”一章。

转换支持

从版本 10.2 开始，可以在流映射中使用具有限制的等级转换。

有关详细信息，请参阅《*Informatica Intelligent Streaming 10.2 用户指南*》中的“智能流映射”一章。

Metadata Manager

本节介绍版本 10.2 中新增的 Metadata Manager 功能。

Cloudera 导航器

从版本 10.2 开始，可以提供信任库文件信息，以便与 Cloudera Navigator 资源建立安全连接。创建或编辑 Cloudera Navigator 资源时，请输入 Cloudera Navigator SSL 实例的信任库文件的路径和文件名以及该信任库文件的密码。

有关创建 Cloudera Navigator 资源的详细信息，请参阅《*Informatica Metadata Manager 10.2 管理员指南*》中的“数据库管理资源”一章。

PowerCenter

本节介绍版本 10.2 中新增的 PowerCenter 功能。

审核日志

从版本 10.2 开始，将 .xml 文件导入到 PowerCenter 存储库时可以生成审核日志。导入一个或多个存储库对象时，可以生成审核日志。可以在 Administrator 工具的 PowerCenter 存储库服务属性中启用"安全审核跟踪"配置选项，以便在将 .xml 文件导入到 PowerCenter 存储库时生成审核日志。用户活动日志会捕获所有审核消息。

审核日志包含有关该文件的以下信息：文件名和文件大小、导入的对象数和导入操作所用的时间。

有关详细信息，请参阅《*Informatica 10.2 命令参考*》中的"pmrep 命令参考"一章、《*Informatica 10.2 应用程序服务指南*》以及《*Informatica 10.2 管理员指南*》。

用于 SAP HANA 目标的批量更新插入

从版本 10.2 开始，当您将数据更新插入到 SAP HANA 目标中时，可以配置 EnableArrayUpsert 自定义属性，以便批量更新插入数据并提高会话性能。可以在会话级别或 PowerCenter 集成服务级别配置 EnableArrayUpsert 自定义属性，将其值设置为"yes"。

有关详细信息，请参阅《*Informatica 10.2 PowerCenter Designer 指南*》中的"处理目标"一章。

对象查询

从版本 10.2 开始，可以使用 *pmrep* 命令创建和删除对象查询。

有关详细信息，请参阅《*Informatica 10.2 命令参考*》中的"pmrep 命令参考"一章。

在密码中使用参数

从版本 10.2 开始，可以使用 *pmrep* 命令创建或更新密码中含有参数的连接。

此外也可以使用 *pmrep* 命令更新密码中带有或不带有参数的连接。

有关详细信息，请参阅《*Informatica 10.2 命令参考*》中的"pmrep 命令参考"一章。

PowerExchange 适配器

本节介绍版本 10.2 中新增的 PowerExchange 适配器功能。

Informatica PowerExchange 适配器

本节介绍版本 10.2 中新增的 Informatica 适配器功能。

PowerExchange for Amazon Redshift

从版本 10.2 开始，PowerExchange for Amazon Redshift 包括以下新功能：

- 在以下地区，您可以从 Amazon S3 段读取数据或者向 Amazon S3 段写入数据：
 - 亚太地区（孟买）
 - 亚太地区（首尔）
 - 加拿大（中部）
 - 中国（北京）

- 欧盟（伦敦）
- 美国东部（俄亥俄州）
- 您可以在 Spark 引擎上运行 Amazon Redshift 映射。运行映射时，数据集成服务会将映射推送到 Hadoop 群集并在 Spark 引擎上处理映射，这将大幅提升性能。
- 您可以使用 AWS 身份和访问管理 (IAM) 身份验证来安全地控制对 Amazon S3 资源的访问。
- 您可以通过 VPC 端点连接到虚拟私有云 (VPC) 中可用的 Amazon Redshift 群集。
- 您可以使用 AWS 身份和访问管理 (IAM) 身份验证在 EMR 群集上运行会话。

有关详细信息，请参阅《*Informatica PowerExchange for Amazon Redshift 10.2 用户指南*》。

PowerExchange for Amazon S3

从版本 10.2 开始，PowerExchange for Amazon S3 包括以下新功能：

- 在以下地区，您可以从 Amazon S3 段读取数据或者向 Amazon S3 段写入数据：
 - 亚太地区（孟买）
 - 亚太地区（首尔）
 - 加拿大（中部）
 - 中国（北京）
 - 欧盟（伦敦）
 - 美国东部（俄亥俄州）
- 在本机环境和 Spark 引擎中从 Amazon S3 段读取数据或者向 Amazon S3 段写入数据时，可以用以下格式压缩数据：

压缩格式	读取	写入
Bzip2	是	是
Deflate	否	是
Gzip	是	是
Lzo	是	是
无	是	是
Snappy	否	是

- 您可以在 Amazon S3 数据对象读取操作的高级属性的**源类型**选项下，选择要从中读取数据的源的类型。您可以选择**目录**或**文件**源类型。
- 您可以在 Amazon S3 数据对象属性的**资源格式**选项下，选择数据源的类型。您可以从以下源格式读取数据：
 - 二进制
 - 平面
 - Avro
 - Parquet
- 您可以通过 VPC 端点连接到虚拟私有云 (VPC) 中可用的 Amazon S3 段。

- 您可以在 Spark 引擎上运行 Amazon S3 映射。运行映射时，数据集成服务会将映射推送到 Hadoop 群集并在 Spark 引擎上处理映射。
- 您可以选择覆盖现有文件。您可以在 Amazon S3 数据对象写入操作属性中选择**覆盖文件（如果存在）**选项来覆盖现有数据。
- 您可以使用 AWS 身份和访问管理 (IAM) 身份验证来安全地控制对 Amazon S3 资源的访问。
- 您可以在**对象浏览器**视图下筛选元数据，从而优化搜索性能。
- 您可以使用 AWS 身份和访问管理 (IAM) 身份验证在 EMR 群集上运行会话。

有关详细信息，请参阅《*Informatica PowerExchange for Amazon S3 10.2 用户指南*》。

PowerExchange for HBase

从版本 10.2 开始，PowerExchange for HBase 包含以下新功能：

- 您可以使用 PowerExchange for HBase 从源读取数据，以及将其写入存储在 Azure HDInsight 的 WASB 文件系统中的目标。
- 您可以将群集配置与 HBase 连接相关联。群集配置是含有 Hadoop 群集配置信息的域中的一个对象。使用群集配置有助于数据集成服务将映射逻辑推送到 Hadoop 环境。

有关详细信息，请参阅《*Informatica PowerExchange for HBase 10.2 用户指南*》。

PowerExchange for HDFS

从版本 10.2 开始，可以将群集配置与 HDFS 连接关联起来。群集配置是含有 Hadoop 群集配置信息的域中的一个对象。使用群集配置有助于数据集成服务将映射逻辑推送到 Hadoop 环境。

有关详细信息，请参阅《*Informatica PowerExchange for HDFS 10.2 用户指南*》。

PowerExchange for Hive

从版本 10.2 开始，可以将群集配置与 Hive 连接关联起来。群集配置是含有 Hadoop 群集配置信息的域中的一个对象。使用群集配置有助于数据集成服务将映射逻辑推送到 Hadoop 环境。

有关详细信息，请参阅《*Informatica PowerExchange for Hive 10.2 用户指南*》。

PowerExchange for MapR-DB

从版本 10.2 开始，PowerExchange for MapR-DB 包含以下新功能：

- 您可以在 Spark 引擎上运行 MapR-DB 映射。运行映射时，数据集成服务会将映射推送到 Hadoop 群集并在 Spark 引擎上处理映射，这将大幅提升性能。
- 您可以为在 Spark 引擎上运行的 MapR-DB 映射配置动态分区。
- 您可以将群集配置与适用于 MapR-DB 的 HBase 连接关联起来。群集配置是含有 Hadoop 群集配置信息的域中的一个对象。使用群集配置有助于数据集成服务将映射逻辑推送到 Hadoop 环境。

有关详细信息，请参阅《*Informatica PowerExchange for MapR-DB 10.2 用户指南*》。

PowerExchange for Microsoft Azure Blob Storage

从版本 10.2 开始，可以从 Microsoft Azure Blob 存储的子目录中读取数据或将数据写入该子目录。您可以使用 **Blob 容器覆盖**和 **Blob 名称覆盖**字段从 Microsoft Azure Blob 存储子目录读取数据或将数据写入 Microsoft Azure Blob 子目录。

有关详细信息，请参阅《*Informatica PowerExchange for Microsoft Azure Blob Storage 10.2 用户指南*》。

PowerExchange for Microsoft Azure SQL Data Warehouse

从版本 10.2 开始，可以在启用了 Kerberos 的群集的 Hadoop 环境中运行 Microsoft Azure SQL 数据仓库映射。

有关详细信息，请参阅《*Informatica PowerExchange for Microsoft Azure SQL Data Warehouse 10.2 用户指南*》。

PowerExchange for Salesforce

从版本 10.2 开始，可以使用版本 39 Salesforce API 创建 Salesforce 连接并访问 Salesforce 对象。

有关详细信息，请参阅《*Informatica PowerExchange for Salesforce 10.2 用户指南*》。

PowerCenter PowerExchange 适配器

本节介绍版本 10.2 中新增的 PowerCenter 适配器功能。

PowerExchange for Amazon Redshift

从版本 10.2 开始，PowerExchange for Amazon Redshift 包括以下新功能：

- 您可以从中国（北京）地区读取数据或将数据写入中国（北京）地区。
- 在 PowerCenter Designer 中从 **AmazonRSCloudAdapter** 导入对象时，PowerCenter 集成服务会按字母顺序列出表名。
- 除了真空表中的现有恢复选项外，还可以选择**重新索引**选项来分析交错排序键列中值的分布。
- 您可以配置多部分上载选项，将单个对象作为一组独立部件进行上载。TransferManager API 会将单个对象的多个部分上载到 Amazon S3。上载后，Amazon S3 会组装这些部分，创建整个对象。数据内容较大且带宽较高时，TransferManager API 使用多部分上载选项来提高性能并增加吞吐量。
您可以在目标会话属性中配置**部分大小**和 **TransferManager 线程池大小**。
- PowerExchange for Amazon Redshift 使用 commons-beanutils.jar 文件解决访问属性时的潜在安全问题。以下是 commons-beanutils.jar 文件的位置：
<Informatica 安装目录>server/bin/javali/505100/commons-beanutils-1.9.3.jar

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica PowerExchange for Amazon Redshift 10.2 用户指南*》。

PowerExchange for Amazon S3

从版本 10.2 开始，PowerExchange for Amazon S3 包括以下新功能：

- 您可以从中国（北京）地区读取数据或将数据写入中国（北京）地区。
- 您可以从 Amazon S3 读取多个文件，并将数据写入目标。
- 您可以从一个源将多个文件写入到 Amazon S3 目标。您可以在目标会话属性中配置**分布列**选项。
- 创建映射任务以将数据写入 Amazon S3 目标时，可以配置分区来提高性能。您可以在目标会话属性中配置**合并分区文件**选项。
- 您可以在**暂存文件位置**属性中指定 PowerCenter 集成服务 可用的目录路径。
- 您可以配置多部分上载选项，将单个对象作为一组独立部件进行上载。TransferManager API 会将单个对象的多个部分上载到 Amazon S3。上载后，Amazon S3 会组装这些部分，创建整个对象。数据内容较大且带宽较高时，TransferManager API 使用多部分上载选项来提高性能并增加吞吐量。
您可以在目标会话属性中配置**部分大小**和 **TransferManager 线程池大小**。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica PowerExchange for Amazon S3 10.2 用户指南*》。

PowerExchange for Microsoft Dynamics CRM

从版本 10.2 开始，您可以在 PowerExchange for Microsoft Dynamics CRM 中使用以下目标会话属性：

- 添加行拒绝原因。选中该属性可将拒绝行的原因包含到拒绝文件中。
- 备用键名称。指示列是否为实体的备用键。指定备用键的名称。可以在更新和 upsert 操作中使用备用键。
- 您可以将 PowerExchange for Microsoft Dynamics CRM 配置为在 AIX 平台上运行。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Microsoft Dynamics CRM 10.2 用户指南》。

PowerExchange for SAP NetWeaver

从版本 10.2 开始，PowerExchange for SAP NetWeaver 包括以下新功能：

- 运行 ABAP 映射从 SAP 表中读取数据时，可以使用 STRING、SSTRING 和 RAWSTRING 数据类型。在 PowerCenter 中，SSTRING 数据类型以 SSTR 形式表示。
- 通过 IDocs 读取或写入数据时，可以使用 SSTRING 数据类型。
- 运行 ABAP 映射从 SAP 表中读取数据时，可以配置 HTTP 流。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for SAP NetWeaver 10.2 用户指南》。

规则规范

从版本 10.2 开始，可以从 Informatica Developer 的模型存储库中选择规则规范，并将规则规范添加到映射。还可以将规则规范部署为 Web 服务。

规则规范是 Developer tool 中的只读对象。按照向映射添加 Mapplet 的相同方式向映射添加规则规范。您可以继续选择从规则规范生成的 Mapplet，并将 Mapplet 添加到映射中。

如果希望映射应用当前规则规范所表示的逻辑，请向映射中添加规则规范。如果希望独立于规则规范使用或更新 Mapplet 逻辑，请将相应的 Mapplet 添加到映射中。

向映射中添加规则规范时，可以指定规则规范上的输出类型。默认情况下，规则规范具有单个输出端口，其中包含每个输入数据行的规则规范分析的最终结果。您可以配置规则规范，为规则规范中的每个规则集创建输出端口。

有关详细信息，请参阅《Informatica 10.2 Developer 映射指南》中的“Mapplet”一章。

安全

本节介绍版本 10.2 中新增的安全功能。

用户活动日志

从版本 10.2 开始，可以在用户活动日志中查看来自 Informatica 客户端应用程序的登录尝试。

用户活动数据包括来自 Informatica 客户端的每个登录尝试的下列属性：

- 应用程序名称
- 应用程序版本
- 应用程序主机的主机名或 IP 地址

如果客户端在登录请求中设置了自定义属性，则数据包括这些自定义属性。

有关详细信息，请参阅《Informatica 10.2 安全指南》中的“用户和组”一章。

转换语言

本节介绍 10.2 中新增的转换语言功能。

Informatica 转换语言

本节介绍版本 10.2 中新增的 Informatica 转换语言功能。

复杂函数

从版本 10.2 开始，转换语言引入了适用于复杂数据类型的复杂函数。使用复杂函数可处理 Spark 引擎上的层次结构数据。

转换语言包含以下复杂函数：

- ARRAY
- CAST
- COLLECT_LIST
- CONCAT_ARRAY
- RESPEC
- SIZE
- STRUCT
- STRUCT_AS

有关复杂函数的详细信息，请参阅《*Informatica 10.2 Developer 转换语言参考*》中的“函数”一章。

复杂运算符

从版本 10.2 开始，转换语言引入了适用于复杂数据类型的复杂运算符。在 Spark 引擎上运行的映射中，使用复杂运算符可访问层次结构数据中的元素。

转换语言包括以下复杂运算符：

- 下标运算符 []
- 点运算符 .

有关复杂运算符的详细信息，请参阅《*Informatica 10.2 Developer 转换语言参考*》中的“运算符”一章。

窗口函数

从版本 10.2 开始，转换语言引入了窗口功能。使用窗口函数可以在 Spark 引擎上处理大型数据集的较小子集。

转换语言包含以下窗口函数：

- LEAD。提供对位于当前行之后给定物理偏移量的行的访问权限。
- LAG。提供对位于当前行之前给定物理偏移量的行的访问权限。

有关详细信息，请参阅《*Informatica 10.2 转换语言参考*》中的“函数”一章。

转换

本节介绍版本 10.2 中新增的转换功能。

Informatica 转换

本节介绍版本 10.2 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：

奥地利

从版本 10.2 开始，可以对地址验证器转换进行配置，使其返回具有两个有效街道地址的邮箱的邮政地址代码标识符。例如，在两条街道交叉路口的一座建筑物可能在两条街道上都有一个地址。该建筑物可能更喜欢选择其中一个地址接收邮件。另一个地址仍然是有效地址，但邮递员不会使用它来投递邮件。

奥地利邮政公司会为这两个地址分配一个邮政地址代码。此外，奥地利邮政公司还会为不用来接收邮件的地址分配一个邮政地址代码标识符。该邮政地址代码标识符与首选地址的邮政地址代码相同。您可以将该邮政地址代码标识符与地址验证器转换结合使用来查找首选地址。

要查找某个奥地利地址的邮政地址代码标识符，请选择“邮政地址代码标识符奥地利”输出端口。在“奥地利补充”端口组中查找此端口。

要查找邮政地址标识符所代表的地址，请选择“邮政地址代码标识符奥地利”输入端口。在“离散值”端口组中查找此端口。

捷克共和国

从版本 10.2 开始，可以对地址验证器转换进行配置，使其向有效的捷克共和国地址添加 RUIAN ID 值。

您可以查找以下 RUIAN ID 值：

- RUIANAM_ID。唯一标识地址投递点。
要查找唯一标识地址投递点的 RUIAN ID 值，请选择“RUIAN 投递点标识符”输出端口。
- RUIANSO_ID。标识地址所在的建筑物。
要查找标识地址所在建筑物的 RUIAN ID 值，请选择“RUIAN 建筑物标识符”输出端口。
- RUIANTEA_ID。标识建筑物入口。
要查找标识建筑物入口的 RUIAN ID 值，请选择“RUIAN 建筑物入口标识符”输出端口。

在“捷克共和国补充”端口组中查找此端口。

中国香港特别行政区

地址验证器转换包括针对香港的以下功能：

香港地址支持多语言

从版本 10.2 开始，地址验证器转换可以读取和写入中文或英文香港地址。

使用“首选语言”属性可选择转换返回的地址的首选语言。默认语言为“中文”。要返回英文香港地址，请将该属性更新为“英文”。

使用“首选脚本”属性选择地址数据的首选字符集。默认字符集为“汉字”。要返回拉丁字符香港地址，请将该属性更新为“拉丁字符”或“ASCII”选项。选择拉丁字符脚本时，地址验证会将地址数据音译为拼音。

建议列表模式下的单行地址验证

从版本 10.2 开始，可以对地址验证器转换进行配置，使其为您在单行中输入的香港地址返回有效的建议。要返回建议，请将转换配置为在建议列表模式下运行。

使用原生中文和汉字脚本提交地址。地址验证器转换会读取汉字脚本中的地址，并返回汉字脚本中的地址建议。

请按以下格式提交香港地址：

[Province] [Locality] [Street] [House Number] [Building 1] [Building 2] [Sub-building]

提交不完整地址时，转换将为您输入的地址返回一个或多个地址建议。输入完整或几乎完整的地址时，转换将为您输入的地址返回一个建议。

要验证单行地址，请使用“完整地址”端口。

中国澳门特别行政区

地址验证器转换包括针对澳门的以下功能：

澳门地址支持多语言

从版本 10.2 开始，地址验证器转换可以读取和写入中文或葡萄牙语澳门地址。

使用“首选语言”属性可选择转换返回的地址的首选语言。默认语言为“中文”。要返回葡萄牙语澳门地址，请将该属性更新为 ALTERNATIVE_2。

使用“首选脚本”属性选择地址数据的首选字符集。默认字符集为“汉字”。要返回拉丁字符澳门地址，请将该属性更新为“拉丁字符”或“ASCII”选项。

注意：选择含有默认首选语言选项的拉丁字符脚本时，地址验证会将中文地址数据音译为粤语或普通话。选择含有 ALTERNATIVE_2 首选语言选项的拉丁字符脚本时，地址验证将返回葡萄牙语地址。

在建议列表模式下对原生澳门地址进行单行地址验证

从版本 10.2 开始，可以对地址验证器转换进行配置，使其为您在建议列表模式下在单行中输入的澳门地址返回有效的建议。在建议列表模式下输入不完整地址时，转换将为您输入的地址返回一个或多个地址建议。使用中文和汉字脚本提交地址。转换会返回汉字脚本中的中文地址建议。请按以下格式输入澳门地址：

[Locality] [Street] [House Number] [Building]

使用“首选语言”属性可选择地址的首选语言。默认首选语言为“中文”。使用“首选脚本”属性选择地址数据的首选字符集。默认首选脚本为汉字。要验证单行地址，请在“完整地址”端口中输入地址。

中国台湾

从版本 10.2 开始，可以对地址验证器转换进行配置，使其返回中文或英文台湾地址。

使用“首选语言”属性可选择转换返回的地址的首选语言。默认语言为繁体中文。要返回英文台湾地址，请将该属性更新为“英文”。

使用“首选脚本”属性选择地址数据的首选字符集。默认字符集为“汉字”。要返回拉丁字符台湾地址，请将该属性更新为“拉丁字符”或“ASCII”选项。

注意：原生脚本中的台湾地址结构会在单行列出所有地址元素。您可以在“格式化的地址行”端口中将地址作为单个字符串提交。

格式化输入的地址时，请按以下顺序输入地址中的元素：

Postal Code, Locality, Dependent Locality, Street, Dependent Street, House or Building Number, Building Name, Sub-Building Name

美国

地址验证器转换包括针对美国的以下功能：

支持与安全哈希算法兼容的 CASS 数据文件版本

从版本 10.2 开始，地址验证器转换可读取符合 SHA-256 标准的 CASS 认证数据文件。

最新 CASS 认证文件的编号为 USA5C101.MD 至 USA5C126.MD。要在认证模式下验证美国地址，必须使用最新文件。

注意：符合 SHA-256 标准的文件与旧版本 Informatica 不兼容。

在认证模式下支持“无人开门”地址

从版本 10.2 开始，可以对地址验证器转换进行配置，使其标识没有为邮递员提供门或入口点的美国地址。邮递员可能无法将大件物品投递到该地址。

美国邮政署维护着一个可以投递到邮箱但无法投递到物理入口的地址列表。例如，住所可能将邮箱设在上锁的门之外或设在乡村路线上。地址引用数据包括美国邮政署认可的无法投递的地址列表。在认证模式下验证地址时，地址验证可以返回地址是否可投递状态。

要标识 DNA 地址，请选择“投递点验证无人开门”端口。在“特定于美国”端口组中查找此端口。

在认证模式下支持“无安全位置”地址

从版本 10.2 开始，可以对地址验证器转换进行配置，使其标识没有提供安全邮箱或邮件签收点的美国地址。邮递员可能无法将大件物品投递到该地址。

美国邮政署维护着一个邮箱不安全的地址列表。例如，对零售店来说，如果邮递员可以进入店铺，但找不到邮箱或找不到员工签收邮件，那么零售店就不是安全的地点。地址引用数据包括美国邮政署认可的不安全地址列表。在认证模式下验证地址时，地址验证可以返回地址的不安全状态。

要标识 DNA 地址，请选择“投递点验证不安全地点”端口。在“特定于美国”端口组中查找此端口。

支持“只有邮政信箱的投递区”

从版本 10.2 开始，可以对地址验证器转换进行配置，使其标识只有邮政信箱地址而没有其他地址的邮政编码。某个邮政编码下的所有地址都是邮政信箱地址时，该邮政编码就代表只有邮政信箱的投递区。

地址验证器转换会将值 Y 添加到地址中，以指示它含有的邮政编码位于只有邮政信箱的投递区。该值有助于邮递员更轻松的分拣邮件。例如，只有邮政信箱的投递区内的邮箱可能都位于一个邮局大楼内。这时，邮递员只需跑一趟就可以将所有邮件投递到只有邮政信箱的投递区。

要标识只有邮政信箱的投递区，请选择“邮政信箱投递区指示器”端口。在“特定于美国”端口组中查找此端口。

有关详细信息，请参阅《*Informatica 10.2 Developer 转换指南*》和《*Informatica 10.2 地址验证器端口参考*》。

数据处理器转换

本节介绍新增的数据处理器转换功能。

JsonStreamer

使用数据处理器转换中的 JsonStreamer 对象可处理大型 JSON 文件。转换会将非常大的 JSON 文件拆分为多个完整的 JSON 消息。然后，转换可以调用其他数据处理器转换组件或分层到关系转换来完成处理。

有关详细信息，请参阅《*Informatica Data Transformation 10.2 用户指南*》中的“流转化器”一章。

RunPCWebService

使用 RunPCWebService 操作可以调用数据处理器转换中的 PowerCenter Mapplet。

有关详细信息，请参阅《*Informatica Data Transformation 10.2 用户指南*》中的“操作”一章。

PowerCenter 转换

计算表达式

从版本 10.2 开始，可以计算在表达式转换的表达式编辑器中配置的表达式。测试某个表达式时，可以输入示例数据，然后对该表达式求值。

有关计算表达式的详细信息，请参阅《*Informatica PowerCenter 10.2 转换指南*》中的“使用转换”一章以及“表达式转换”一章。

工作流

本节介绍版本 10.2 中新增的工作流功能。

Informatica 工作流

本节介绍版本 10.2 中新增的 Informatica 工作流功能。

人工任务分发属性

从版本 10.2 开始，可以将能够处理人工任务实例的用户或组列表存储在外部数据库表中。配置人工任务以根据源数据列中的值定义任务实例时，可以选择该表。

该表标识可以处理任务实例的用户或组，并指定要与每个用户或组关联的列值。例如，有用户加入或离开项目时，更新该表即可，而不必更改工作流配置。当工作流运行时，数据集成服务使用该表中的当前信息将任务实例分配给用户或组。

将用户或组与源数据列中的值关联时，还可以指定数值或日期值的范围。当一个或多个记录包含指定范围内的值时，数据集成服务会将任务实例分配给指定的用户或组。

有关详细信息，请参阅《*Informatica 10.2 Developer 工作流指南*》中的“人工任务”一章。

人工任务通知属性

从版本 10.2 开始，可以编辑在人工任务中配置的电子邮件通知的主题行。此外还可以将工作流变量添加到通知的主题行。

人工任务在工作流中完成时以及人工任务定义的任务实例更改状态时，人工任务可以发送电子邮件通知。要为人工任务配置通知，请更新工作流中人工任务的通知属性。要为任务实例配置通知，请更新定义任务实例的人工任务中的步骤上的通知属性。

为人工任务实例配置通知时，除了通知指定的任何收件人外，还可以选择一个选项来通知任务实例所有者。任务实例为单个用户所拥有时，将应用该选项。选择通知任务实例所有者的选项时，可以选择将“收件人”字段留空。

有关详细信息，请参阅《*Informatica 10.2 Developer 工作流指南*》中的“人工任务”一章。

从 PowerCenter 导入

从版本 10.2 开始，可以将具有多个管道、会话、工作流和工作集的映射从 PowerCenter 导入到模型存储库中。工作流内的会话将作为模型存储库中的映射任务导入。工作流作为模型存储库内的工作流导入。工作流中的工作集会展开，然后将对象导入模型存储库。

映射中的多个管道会根据目标负载顺序作为单独的映射导入到模型存储库中。如果工作流包含运行具有多个管道的映射的会话，则导入进程将为 PowerCenter 映射中的每个管道创建单独的模型存储库映射和映射任务，以保持目标负载顺序不变。

有关从 PowerCenter 导入的详细信息，请参阅《*Informatica 10.2 Developer 映射指南*》中的“从 PowerCenter 导入”一章以及《*Informatica 10.2 Developer 工作流指南*》中的“工作流”一章。

第 26 章

更改 (10.2)

本章包括以下主题：

- [支持更改, 282](#)
- [应用程序服务, 285](#)
- [大数据, 286](#)
- [命令行程序, 291](#)
- [Enterprise Information Catalog, 291](#)
- [Informatica Analyst, 292](#)
- [Intelligent Streaming, 292](#)
- [PowerExchange 适配器, 292](#)
- [安全, 294](#)
- [转换, 295](#)
- [工作流, 296](#)

支持更改

本节介绍版本 10.2 中的支持更改。

Big Data Hadoop 发行版支持

Informatica 大数据产品支持各种 Hadoop 发行版。在每个版本中，Informatica 会增加、延续和终止对 Hadoop 发行版版本的支持。Informatica 可能会在将来的版本中恢复对延续版本的支持。

下表列出了 Informatica 10.2 大数据产品支持的 Hadoop 发行版版本：

产品	Amazon EMR	Azure HDInsight	Cloudera CDH	Hortonworks HDP	IBM BigInsights	MapR
Big Data Management	5.4, 5.8	3.5, 3.6	5.9, 5.10, 5.11, 5.12, 5.13	2.4, 2.5, 2.6	4.2	5.2 MEP 2.0 5.2 MEP 3.0
Informatica Intelligent Streaming	5.8	NA	5.11, 5.12, 5.13	2.6	NA	5.2 MEP 2.0
Enterprise Information Catalog	NA	3.6	5.8、5.9、5.10 和 5.11	2.5, 2.6	4.2.x	3.1
Intelligent Data Lake	5.4	3.6	5.11, 5.12	2.6	4.2	5.2 MEP 2.0

要查看支持的版本的最新列表，请参阅 Informatica 客户门户上的"产品可用性列表"，网址为：
<https://network.informatica.com/community/informatica-network/product-availability-matrices>。

Big Data Management Hadoop 发行版

下表列出了 Big Data Management 10.2 中支持的 Hadoop 发行版版本和更改：

Hadoop 发行版	支持的发行版版本	10.2 中的更改
Amazon EMR	5.8	增加了对版本 5.8 的支持。终止了对版本 5.0 和 5.4 的支持。 注意: 要将 Amazon EMR 5.8 与 Big Data Management 10.2 一起使用，必须应用 Emergency Bug Fix 10571。请查阅知识库文章 KB 525399 。
Azure HDInsight	3.5.x 3.6.x	增加了对版本 3.6 的支持。
Cloudera CDH	5.10.x 5.11.x 5.12.x 5.13.x	增加了对版本 5.12 和 5.13 的支持。 终止了对版本 5.8 和 5.9 的支持。
Hortonworks HDP	2.5x 2.6x	终止了对版本 2.3 和 2.4 的支持。 注意: 要将 Hortonworks 2.5 与 Big Data Management 10.2 一起使用，必须应用紧急缺陷修复程序。请查阅以下知识库文章： - Hortonworks 2.5 支持： KB 521847 。
MapR	5.2 MEP 3.0.x	增加了对版本 5.2 MEP 3.0 的支持。 终止了对版本 5.2 MEP 1.x 和 5.2 MEP 2.x 的支持。

Informatica 大数据产品支持各种 Hadoop 发行版。在每个版本中，Informatica 会增加、延期和终止对 Hadoop 发行版版本的支持。Informatica 可能会在将来的版本中恢复对延续版本的支持。

要查看支持的版本的最新列表，请参阅 Informatica Network 上的"产品可用性列表"，网址为：<https://network.informatica.com/community/informatica-network/product-availability-matrices>。

Enterprise Information Catalog Hadoop 发行版

下表列出了 Enterprise Information Catalog 10.2 中支持的 Hadoop 发行版版本和更改：

Hadoop 发行版	支持的发行版版本	自 10.1.1 HotFix1 以来的更改
Azure HDInsight	3.6	增加了对 Azure HDInsight 的支持。
Cloudera CDH	5.8、5.9、5.10 和 5.11	无更改。
Hortonworks HDP	2.5.x（Kerberos 版本）和 2.6.x（非 Kerberos 版本）	增加了对 2.6 非 Kerberos 版本的支持。
IBM BigInsights	4.2	无更改。

Intelligent Data Lake Hadoop 发行版

下表列出了 Intelligent Data Lake 10.2 中支持的 Hadoop 发行版版本和更改：

Hadoop 发行版	支持的发行版版本	自 10.1.1 HotFix1 以来的更改
Amazon EMR	5.4	增加了对版本 5.4 的支持。 终止了对版本 5.0 的支持。
Azure HDInsight	3.6	增加了对版本 3.6 的支持。 终止了对版本 3.5 的支持。
Cloudera CDH	5.10 5.11 5.12	增加了对版本 5.10 和 5.12 的支持。 终止了对版本 5.8 的支持。 延续了对版本 5.9 的支持。
Hortonworks HDP	2.6	终止了对版本 2.3 的支持。 延期了对版本 2.4 和 2.5 的支持。
IBM BigInsights	4.2	无更改。
MapR	5.2 MEP 2.0	增加了对 MapR 的支持。

Intelligent StreamingHadoop 分布

下表列出了 Intelligent Streaming 10.2 中支持的 Hadoop 发行版版本和更改：

发行版	支持的版本	自 10.1.1 HotFix1 以来的更改
Amazon EMR	5.4 5.8	增加了对 5.8 的支持。
Cloudera CDH	5.10.x 5.11.x 5.12.x 5.13.x	增加了对 5.13 的支持。 终止了对版本 5.8 的支持。 延续了对版本 5.9 的支持。
Hortonworks HDP	2.5.x 2.6.x	终止了对版本 2.3 的支持。 延续了对版本 2.4 的支持。
MapR	5.2 MEP 2.0	增加了对版本 5.2 MEP 2.0 的支持。

要查看支持的版本的最新列表，请参阅 Informatica Network 上的"产品可用性列表"，网址为：<https://network.informatica.com/community/informatica-network/product-availability-matrices>。

Metadata Manager

Custom Metadata Configurator（已弃用）

从版本 10.2 开始，Informatica 弃用了 Metadata Manager 中的 Custom Metadata Configurator。

您可以使用加载模板从元数据源文件将元数据加载到自定义资源中。为使用 Custom Metadata Configurator 模板的模型创建一个加载模板。

有关使用加载模板的详细信息，请参阅《*Informatica Metadata Manager 10.2 自定义元数据集成指南*》中的“使用加载模板创建自定义 XConnect”。

应用程序服务

本节介绍版本 10.2 中对应用程序服务的更改。

内容管理服务

从版本 10.2 开始，在运行 `infacmd cms purge` 命令之前不需要更新模型存储库中的搜索索引。`infacmd cms purge` 命令会在清除引用数据仓库中未使用的表之前更新搜索索引。

以前，您需要在运行该命令之前更新搜索索引，以便模型存储库拥有最新的引用表列表。内容管理服务会使用索引中的对象列表来选择要删除的表。

有关详细信息，请参阅《*Informatica 10.2 应用程序服务指南*》中的“内容管理服务”一章。

数据集成服务

本节介绍版本 10.2 中对数据集成服务的更改。

执行选项

从版本 10.2 开始，可在"属性"视图上为数据集成服务配置以下执行选项：

- 按需执行池大小上限。控制可并发运行的按需作业数。作业包括数据预览、剖析作业、REST 和 SQL 查询、Web 服务请求以及从 Developer tool 运行的映射。
- 本地批处理执行池大小上限。控制每个数据集成服务进程可并发运行的已部署本地作业数。
- Hadoop 批处理执行池大小上限。控制可并发运行的已部署 Hadoop 作业数。

以前，需要配置"执行池大小上限"属性来控制数据集成服务进程可以并发运行的最大作业数。

升级到 10.2 后，执行池大小上限的值将升级到以下属性：

- 按需批处理执行池大小上限。继承"执行池大小上限"属性的值。
- 本地批处理执行池大小上限。继承"执行池大小上限"属性的值。
- Hadoop 批处理执行池大小上限。如果原始值已从 10 更改为其他值，则继承"执行池大小上限"属性的值。如果值为 10，则 Hadoop 批处理池会保持默认大小，即 100。

有关详细信息，请参阅《Informatica 10.2 应用程序服务指南》中的"数据集成服务"一章。

大数据

本节介绍版本 10.2 中对大数据的更改。

Hadoop 连接

从版本 10.2 开始，以下更改会影响 Hadoop 连接属性。

您可以使用以下属性来配置 Hadoop 连接：

属性	说明
群集配置	与 Hadoop 环境关联的群集配置的名称。 出现在常规属性中。
将拒绝文件写入 Hadoop	选择该属性可在运行映射时将拒绝文件移动到属性拒绝文件目录中列出的 HDFS 位置。 出现在拒绝目录属性中。
拒绝文件目录	运行映射时 HDFS 上 Hadoop 映射文件所在的目录。 出现在拒绝目录属性中
Blaze 作业监视器地址	Blaze 作业监视器的主机名和端口号。 出现在 Blaze 配置中。
YARN 队列名称	Spark 引擎用来指定群集上的可用资源的 YARN 调度程序队列名称。 出现在 Blaze 配置中。

从版本 10.2 开始，以下属性被重命名：

当前名称	以前名称	说明
ImpersonationUserName	HiveUserName	Hadoop 模拟用户：数据集成服务为在 Hadoop 环境中运行映射而模拟的用户名。
Hive 暂存数据库名称	数据库名称	Hive 暂存表的命名空间。 出现在公共属性中。 以前出现在 Hive 属性中。
HiveWarehouseDirectory	HiveWarehouseDirectoryOnHDFS	仓库（此仓库对群集来说属于本地仓库）的默认数据库的绝对 HDFS 文件路径。
Blaze 暂存目录	HDFS 上的临时工作目录 CadiWorkingDirectory	Blaze 引擎用于存储临时文件的目录的 HDFS 文件路径。 出现在 Blaze 配置中。
Blaze 用户名	Blaze 服务用户名 CadiUserName	Blaze 服务和 Blaze 服务日志的所有者。 出现在 Blaze 配置中。
YARN 队列名称	Yarn 队列名称 CadiAppYarnQueueName	由在群集上指定可用资源的 Blaze 引擎使用的 YARN 计划程序队列名称。 出现在 Blaze 配置中。
BlazeMaxPort	CadiMaxPort	Blaze 引擎的端口号范围的最大值。
BlazeMinPort	CadiMinPort	Blaze 引擎的端口号范围的最小值。
BlazeExecutionParameterList	CadiExecutionParameterList	应用到 Blaze 引擎的配置参数的可选列表。
SparkYarnQueueName	YarnQueueName	Spark 引擎用来指定群集上的可用资源的 YARN 调度程序队列名称。
Spark 暂存目录	Spark HDFS 暂存目录	Spark 引擎用于存储运行作业的临时文件的目录的 HDFS 文件路径。

从版本 10.2 开始，以下属性将从连接中删除并导入到群集配置中：

属性	说明
资源管理器地址	Hadoop 中提交资源的请求或生成 YARN 应用程序的服务。 作为属性 yarn.resourcemanager.address 导入到群集配置中。 以前出现在 Hadoop 群集属性中。
默认文件系统 URI	用于访问默认的 Hadoop 分布式文件系统的 URI。 作为属性 fs.defaultFS 或 fs.default.name 导入到群集配置中。 以前出现在 Hadoop 群集属性中。

从版本 10.2 开始，以下属性被弃用，并从连接中删除：

属性	说明
类型	连接类型。 以前出现在常规属性。
元存储执行模式*	连接到远程元存储还是本地元存储的控制机制。 以前出现在 Hive 配置中。
元存储数据库 URI*	用于在本地元存储设置下访问数据存储的 JDBC 连接 URI。 以前出现在 Hive 配置中。
元存储数据库驱动程序*	JDBC 数据存储的驱动程序类名称。 以前出现在 Hive 配置中。
元存储数据库用户名*	元存储数据库用户名。 以前出现在 Hive 配置中。
元存储数据库密码*	元存储用户名的密码。 以前出现在 Hive 配置中。
远程元存储 URI*	用于在远程元存储设置下访问元数据的元存储 URI。 此属性作为属性 hive.metastore.uris 导入到群集配置中。 以前出现在 Hive 配置中。
作业监视 URL	MapReduce JobHistory 服务器的 URL。 以前出现在 Hive 配置中。
* 这些属性在版本 10.2 中已弃用。升级到 10.2 后，在以前的版本中设置的属性值将保存在存储库中，但它们不会出现在连接属性中。	

HBase 连接属性

从版本 10.2 开始，以下属性将从连接中删除并导入到群集配置中：

属性	说明
ZooKeeper 主机	托管 ZooKeeper 服务器的计算机的名称。
ZooKeeper 端口	托管 ZooKeeper 服务器的计算机的端口号。
启用 Kerberos 连接	启用 Informatica 域可与使用 Kerberos 身份验证的 HBase 主服务器或区域服务器进行通信。
HBase 主服务器主体	HBase 主服务器的服务主体名称 (SPN)。
HBase 区域服务器主体	HBase 区域服务器的服务主体名称 (SPN)。

Hive 连接属性

从版本 10.2 开始，PowerExchange for Hive 具有以下更改：

- 如果希望 Hive 驱动程序在 Hadoop 群集中运行映射，则不能使用 PowerExchange for Hive 连接。要使用 Hive 驱动程序在 Hadoop 群集中运行映射，请使用 Hadoop 连接。
- 以下属性从连接中删除并导入到群集配置中：

属性	说明
默认 FS URI	用于访问默认的 Hadoop 分布式文件系统的 URI。
JobTracker/Yarn 资源管理器 URI	Hadoop 中将 MapReduce 任务提交到群集中特定节点的服务。
HDFS 上的 Hive 仓库目录	仓库（此仓库对群集来说属于本地仓库）的默认数据库的绝对 HDFS 文件路径。
元存储执行模式	连接到远程元存储还是本地元存储的控制机制。
元存储数据库 URI	用于在本地元存储设置下访问数据存储的 JDBC 连接 URI。
元存储数据库驱动程序	JDBC 数据存储的驱动程序类名称。
元存储数据库用户名	元存储数据库用户名。
元存储数据库密码	元存储用户名的密码。
远程元存储 URI	用于在远程元存储设置下访问元数据的元存储 URI。 此属性作为属性 <code>hive.metastore.uris</code> 导入到群集配置中。

适用于 MapR-DB 的 HBase 连接属性

从版本 10.2 开始，启用 **Kerberos 连接** 属性已从适用于 MapR-DB 的 HBase 连接中删除，并导入到群集配置中。

映射运行时属性

本节列出对映射运行时属性的更改。

执行环境

从版本 10.2 开始，可以将拒绝文件目录配置为 Hadoop 执行环境中的新属性。

名称	值
拒绝文件目录	<p>在 Hadoop 环境中运行映射时，HDFS 上的 Hadoop 映射文件的目录。</p> <p>Blaze 引擎可以将拒绝文件写入到 Hadoop 环境中，用于平面文件、HDFS 和 Hive 目标。Spark 和 Hive 引擎可以将拒绝文件写入到 Hadoop 环境中，用于平面文件和 HDFS 目标。</p> <p>请选择以下选项之一：</p> <ul style="list-style-type: none">- 在数据集成服务计算机中。数据集成服务会根据 RejectDir 系统参数存储拒绝文件。- 在 Hadoop 群集中。拒绝文件会移动到在 Hadoop 连接中配置的拒绝目录。如果未配置该目录，映射将失败。- 推迟到 Hadoop 连接。将根据是否在 Hadoop 连接属性中启用了拒绝目录来移动拒绝文件。如果已启用拒绝目录，则将拒绝文件移动到在 Hadoop 连接中配置的拒绝目录。否则，数据集成服务会根据 RejectDir 系统参数存储拒绝文件。

监视

从版本 10.2 开始，管理员工具中“摘要统计”视图中的 AllHiveSourceTables 行包括从以下来源读取的记录：

- 映射中的原始 Hive 源。
- Hive 引擎定义的暂存 Hive 表。
- 在每个查询中的两个链接的 MapReduce 作业之间暂存数据。

如果 LDTM 会话包含一个 MapReduce 作业，则 AllHiveSourceTables 统计信息只包括映射中的原始 Hive 源。

有关详细信息，请参阅《*Big Data Management 10.2 用户指南*》中的“监控 Hadoop 环境中的映射”一章。

S3 访问和密钥属性

从版本 10.2 开始，群集配置的敏感属性列表中包括以下属性：

- fs.s3a.access.key
- fs.s3a.secret.key
- fs.s3n.awsAccessKeyId
- fs.s3n.awsSecretAccessKey
- fs.s3.awsAccessKeyId
- fs.s3.awsSecretAccessKey

生成群集配置存档文件以便在运行 Developer tool 的计算机上部署时，敏感属性会包括在内，但会进行脱敏处理。

以前，这些属性在运行数据集成服务和 Developer tool 的计算机上的 .xml 配置文件中配置。

有关敏感属性的详细信息，请参阅《*Informatica Big Data Management 10.2 管理员指南*》。

Sqoop

从版本 10.2 开始，如果创建密码文件来访问数据库，Sqoop 将忽略密码文件。Sqoop 使用在 JDBC 连接的密码字段中配置的值。

以前，您可以创建一个密码文件来访问数据库。

有关详细信息，请参阅《*Informatica Big Data Management 10.2 用户指南*》中的“Hadoop 环境中的映射对象”一章。

命令行程序

本节介绍版本 10.2 中对命令进行的更改。

infacmd ihs 命令

过时的命令

下表介绍了已过时的 infacmd ihs 命令：

命令	说明
BackupData	将内部 Hadoop 群集中的 HDFS 数据备份为 zip 文件。备份数据时，Informatica 群集服务会保存 Enterprise Information Catalog 创建的所有数据，例如 HBase 数据、扫描程序数据和引入数据。
removesnapshot	删除现有 HDFS 快照以成功运行 infacmd ihs BackupData 命令从而备份 HDFS 数据。

infacmd ldm 命令

已更改命令

下表介绍了已更改的 infacmd ldm 命令：

命令	更改说明
BackupData	从版本 10.2 开始，此命令的名称已更改为 BackupContents。
LocalDestination	从版本 10.2 开始，BackupContents 命令添加了 -of 选项。
restoreData	从版本 10.2 开始，此命令的名称已更改为 restoreContents。

有关详细信息，请参阅《*Informatica 10.2 命令引用*》中的“infacmd ldm 命令引用”一章。

Enterprise Information Catalog

本节介绍版本 10.2 中对 Informatica Enterprise Information Catalog 进行的更改。

产品名称更改

从版本 10.2 开始，Enterprise Information Catalog 包括以下名称更改：

- Informatica Live Data Map 产品更名为 Informatica Enterprise Information Catalog。
- Informatica Live Data Map Administrator 工具更名为 Informatica Catalog Administrator。
- 安装程序从 Live Data Map 更名为 Enterprise Information Catalog。

Informatica Analyst

本节介绍版本 10.2 中对 Analyst 工具的更改。

参数

本节介绍对 Analyst 工具参数的更改。

系统参数

从版本 10.2 开始，Analyst 工具按以下格式显示系统参数的文件路径：\${[参数名]}/[路径]。

以前，Analyst 工具显示数据对象的本地文件路径，且不解析系统参数。

有关查看数据对象的详细信息，请参阅《*Informatica 10.2 Analyst 工具指南*》。

Intelligent Streaming

本节介绍版本 10.2 中对 Informatica Intelligent Streaming 的更改。

Kafka 数据对象更改

从版本 10.2 开始，当您配置数据操作读取属性时，可以指定 Kafka 源开始从 Kafka 主题读取 Kafka 消息的时间。您可以从配置为执行 Kerberos 身份验证的 Kafka 群集进行读取或写入至该群集。

有关详细信息，请参阅《*Informatica Intelligent Streaming 10.2 用户指南*》中的“流式映射中的源和目标”一章。

PowerExchange 适配器

本节介绍版本 10.2 中对 PowerExchange 适配器的更改。

Informatica PowerExchange 适配器

本节介绍版本 10.2 中对 Informatica 适配器的更改。

PowerExchange for Amazon S3

从版本 10.2 开始，PowerExchange for Amazon S3 发生了以下更改：

- 执行读取和写入操作时，可以提供文件夹路径，但不在高级属性中指定段名称，格式为 `/<folder_name>`。数据集成服务会将此文件夹路径附加到在连接属性中指定的文件夹路径。
以前，您在执行读取和写入操作时需要在高级属性中同时指定段名称及文件夹路径，格式为 `<bucket_name>/<folder_name>`。
- 您可以在元数据导入浏览器的左面板中查看段名称目录及子目录列表，并在右面板中查看所选的文件列表。
此前，PowerExchange for Amazon S3 会在元数据导入浏览器的左面板中显示段名称列表，并在右面板中显示文件夹路径和文件名。
- PowerExchange for Amazon S3 会自动为 Amazon S3 数据对象创建数据对象读取操作和数据对象写入操作。
以前，您必须手动为 Amazon S3 数据对象创建数据对象读取操作和数据对象写入操作。

有关详细信息，请参阅《*Informatica PowerExchange for Amazon S3 10.2 用户指南*》。

PowerCenter PowerExchange 适配器

本节介绍版本 10.2 中对 PowerCenter 适配器的更改。

PowerExchange for Amazon Redshift

从版本 10.2 开始，必须提供 Amazon Redshift 表的架构名称才能成功运行映射。

以前，即使选择了公共架构，映射也会运行。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica PowerExchange for Amazon Redshift 10.2 用户指南*》。

PowerExchange for Email Server

从版本 10.2 开始，PowerExchange for Email Server 会与 Informatica 服务一起安装。

在以前的版本中，PowerExchange for Email Server 有单独的安装程序。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica PowerExchange for Email Server 10.2 用户指南*》。

PowerExchange for JD Edwards EnterpriseOne

从版本 10.2 开始，PowerExchange for JD Edwards EnterpriseOne 会与 Informatica 服务一起安装。

在以前的版本中，PowerExchange for JD Edwards EnterpriseOne 有单独的安装程序。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica PowerExchange for JD Edwards EnterpriseOne 10.2 用户指南*》。

PowerExchange for JD Edwards World

从版本 10.2 开始，PowerExchange for JD Edwards World 会与 Informatica 服务一起安装。

在以前的版本中，PowerExchange for JD Edwards World 有单独的安装程序。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica PowerExchange for JD Edwards World 10.2 用户指南*》。

PowerExchange for LDAP

从版本 10.2 开始，PowerExchange for LDAP 会与 Informatica 服务一起安装。

在以前的版本中，PowerExchange for LDAP 有单独的安装程序。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for LDAP 10.2 用户指南》。

PowerExchange for Lotus Notes

从版本 10.2 开始，PowerExchange for Lotus Notes 会与 Informatica 服务一起安装。

在以前的版本中，PowerExchange for Lotus Notes 有单独的安装程序。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Lotus Notes 10.2 用户指南》。

PowerExchange for Oracle E-Business Suite

从版本 10.2 开始，PowerExchange for Oracle E-Business Suite 会与 Informatica 服务一起安装。

在以前的版本中，PowerExchange for Oracle E-Business Suite 有单独的安装程序。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Oracle E-Business Suite 10.2 用户指南》。

PowerExchange for SAP NetWeaver

从版本 10.2 开始，Informatica 不会将安全传输打包在 Informatica 安装程序 .zip 文件中名为 Secure 的单独文件夹中。Informatica 会同时在以下文件夹中打包标准和安全传输：

- Unicode cofiles: Informatica 安装程序 zip 文件/saptrans/mySAP/UC/cofiles
- Unicode 数据文件: Informatica 安装程序 zip 文件/saptrans/mySAP/UC/data
- 非 Unicode cofiles: Informatica 安装程序 zip 文件/saptrans/mySAP/NUC/cofiles
- 非 Unicode 数据文件: Informatica 安装程序 zip 文件/saptrans/mySAP/NUC/data

在以前的版本中，Informatica 将安全传输打包到下列文件夹中：

- Unicode cofiles: Informatica 安装程序 zip 文件/saptrans/mySAP/UC/Secure/cofiles
- Unicode 数据文件: Informatica 安装程序 zip 文件/saptrans/mySAP/UC/Secure/data
- 非 Unicode cofiles: Informatica 安装程序 zip 文件/saptrans/mySAP/NUC/Secure/cofiles
- 非 Unicode 数据文件: Informatica 安装程序 zip 文件/saptrans/mySAP/NUC/Secure/data

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for SAP NetWeaver 10.2 用户指南》。

PowerExchange for Siebel

从版本 10.2 开始，PowerExchange for Siebel 会与 Informatica 服务一起安装。

在以前的版本中，PowerExchange for Siebel 有单独的安装程序。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Siebel 10.2 用户指南》。

安全

本节介绍版本 10.2 中对安全功能的更改。

SAML 身份验证

从版本 10.2 开始，必须在域级别以及域中的所有网关节点上配置安全断言标记语言 (SAML) 身份验证。

以前，只能在域级别配置 SAML 身份验证。

有关详细信息，请参阅《Informatica 10.2 安全指南》中的“Informatica Web 应用程序的 SAML 身份验证”一章。

转换

本节介绍版本 10.2 中已更改的转换行为。

Informatica 转换

本节介绍版本 10.2 中对 Informatica 转换的更改。

地址验证器转换

本节介绍地址验证器转换的更改。

地址验证器转换对地址功能做了以下更新：

所有国家/地区

从版本 10.2 开始，地址验证器转换使用 5.11.0 版本的 Informatica Address Verification 软件引擎。该引擎支持 Informatica 在版本 10.2 中向地址验证器转换添加的功能。

以前，该转换使用 5.9.0 版本的 Informatica Address Verification 软件引擎。

日本

从版本 10.2 开始，只需配置一个映射就能返回最新日本地址的 Choumei Aza 代码。要返回该代码，请选择“最新 Choumei Aza 代码日本”端口。使用该代码可以查找日本邮政公司认可的任何旧地址的最新版本。

以前，只能使用“新版 Choumei Aza 代码日本”端口来返回某个地址对应的 Choumei Aza 代码的增量更改。由于该转换不包括“最新 Choumei Aza 代码日本”端口，您需要配置两个或多个映射才能验证最新的 Choumei Aza 代码及其对应的地址。

英国

从版本 10.2 开始，可以对地址验证器转换进行配置，使其从英国皇家邮政公司邮政编码地址文件中返回邮政区划、行政区划和传统郡县信息。该转换会返回省端口上的信息。

以前，该转换会在信息与邮政区划有关时返回邮政郡信息。

下表显示了可为每种类型信息选择的端口：

郡信息类型	地址元素
邮政区划	省 1
行政区划	省 2
传统	省 3

更新了多个国家/地区的认证标准

从版本 10.2 开始, Informatica 支持地址验证软件使用的以下认证标准:

- 澳大利亚邮政公司的地址匹配审批系统 (AMAS)。更新到 2017 周期。
- 新西兰邮政公司的 SendRight 认证。更新到 2017 周期。
- 加拿大邮政公司的软件评估和认可计划 (SERP)。更新到 2017 周期。

Informatica 继续支持美国邮政署的最新版编码准确性支持系统 (CASS) 标准以及法国邮政公司的最新版 Service National de L'Adresse (SNA) 标准。

有关详细信息, 请参阅《*Informatica 10.2 Developer 转换指南*》和《*Informatica 10.2 地址验证器端口参考*》。

有关从版本 5.9.0 到版本 5.11.0 中的 Informatica Address Verification 软件引擎的更新的全面信息, 请参阅《*Informatica Address Verification 5.11.0 发行指南*》。

表达式转换

从版本 10.2 开始, 可以使用带有窗口属性的窗口函数或聚合函数, 将表达式转换配置为 Spark 引擎上的主动转换。

以前, 表达式转换只能是被动转换。

有关详细信息, 请参阅《*Big Data Management 10.2 管理员指南*》。

规范器转换

从版本 10.2 开始, 在规范器转换的高级属性中不再提供禁用“生成第一级输出组”的选项。

以前, 可以选择此选项抑制生成第一级输出组。

有关详细信息, 请参阅《*Informatica Big Data Management 10.2 开发人员转换指南*》。

工作流

本节介绍版本 10.2 中已更改的工作流行为。

Informatica 工作流

本节介绍版本 10.2 中对 Informatica 工作流行为的更改。

任务实例通知中的工作流变量

从版本 10.2 开始, 工作流变量 `$taskEvent.startOwner` 更名为 `$taskEvent.owner`。该变量的用法在版本 10.2 中没有变化。

有关详细信息, 请参阅《*Informatica 10.2 Developer 工作流指南*》中的“人工任务”一章。

发布任务 (10.2)

本章包括以下主题：

- [PowerExchange 适配器, 297](#)

PowerExchange 适配器

本节介绍版本 10.2 中 PowerExchange 适配器的发布任务。

PowerCenter PowerExchange 适配器

本节介绍版本 10.2 中 PowerCenter 适配器的发布任务。

PowerExchange for Amazon Redshift

从版本 10.2 开始，对于选择公共架构的现有映射，请确保架构名称正确并且适用于 Redshift 表。公共架构可能不对所有表有效。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica 10.2 PowerExchange for Amazon Redshift 用户指南》。

PowerExchange for Amazon S3

从版本 10.2 开始，当您从 9.5.1 或 9.6.1 升级时，升级过程不会保留连接中的所有属性值。升级后，必须重新配置以下属性：

属性	说明
访问密钥	用于访问 Amazon 帐户资源的访问密钥 ID。如果不使用 AWS Identity and Access Management (IAM) 身份验证，则为必需。 注意: 创建连接之前，请确保您有有效的 AWS 凭据。
密钥	用于访问 Amazon 帐户资源的密码访问密钥。此值与访问密钥关联，并唯一标识帐户。如果指定了访问密钥 ID，则必须指定此值。如果不使用 AWS Identity and Access Management (IAM) 身份验证，则为必需。
主对称密钥	可选。在启用客户端加密时提供 Base64 格式的 256 位 AES 加密密钥。可以使用第三方工具生成密钥。 如果指定一个值，请确保在目标会话属性中指定客户端加密作为加密类型。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica 10.2 PowerExchange for Amazon S3 用户指南》。

PowerExchange for Microsoft Dynamics CRM

从早期版本升级时，必须将 .jar 文件复制到版本 10.2 的安装位置。

- 对于客户端，如果从 9.x 升级到 10.2，请从以下 9.x 安装文件夹 <Informatica 安装目录>\clients\java\jre\lib\security 将 local_policy.jar、US_export_policy.jar 和 cacerts 文件复制到以下 10.2 安装文件夹 <Informatica 安装目录>\clients\java\32bit\jre\lib\security。
- 如果从 10.x 升级到 10.2，请从以下 10.x 安装文件夹 <Informatica 安装目录>\clients\java\32bit\jre\lib\security 将 local_policy.jar、US_export_policy.jar 和 cacerts 文件复制到对应的 10.2 文件夹。
- 对于服务器，请从以前版本的 <Informatica 安装目录>\java\jre\lib\security 文件夹将 local_policy.jar、US_export_policy.jar 和 cacerts 文件复制到对应的 10.2 文件夹。

从早期版本升级时，必须将 msdcrm 文件夹复制到版本 10.2 的安装位置。

- 对于客户端，请从以前版本的 <Informatica 安装目录>\clients\PowerCenterClient\client\bin\javalib 文件夹将 msdcrm 文件夹复制到对应的 10.2 文件夹。
- 对于服务器，请从以前版本的 <Informatica 安装目录>\server\bin\javalib 文件夹将 msdcrm 文件夹复制到对应的 10.2 文件夹。

PowerExchange for SAP NetWeaver

从版本 10.2 开始，Informatica 对适用于 PowerCenter 的 PowerExchange for SAP NetWeaver 支持实施了以下更改：

终止了对 CPI-C 协议的支持

从版本 10.2 开始，Informatica 终止了对 CPI-C 协议的支持。

在从 SAP 表读取数据时使用 RFC 或 HTTP 协议来生成并安装 ABAP 程序。

如果升级使用 CPI-C 协议生成的 ABAP 映射，则必须完成以下任务：

1. 使用流模式 (RFC/HTTP) 重新生成并重新安装 ABAP 程序。
2. 创建具有适当的授权配置文件的系统用户或通信用户，以便在 SAP 和 Informatica 之间启用免对话通信。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for SAP NetWeaver 10.2 用户指南》。

终止了对 ABAP 表读取器标准传输的支持

从版本 10.2 开始，Informatica 终止了对 ABAP 表读取器标准传输的支持。Informatica 将不为 ABAP 表读取器提供标准传输。Informatica 仅为 ABAP 表读取器提供安全传输。

如果从早期版本进行升级，必须删除标准传输并安装安全传输。

有关详细信息，请参阅《Informatica PowerExchange for SAP NetWeaver 10.2 传输版本安装说明》。

为 ABAP 表读取器映射增加了对 HTTP 流式处理的支持

从版本 10.2 开始，当您运行 ABAP 映射以从 SAP 表读取数据时，可以配置 HTTP 流式处理。

要对升级后的 ABAP 映射使用 HTTP 流模式，请执行以下任务：

1. 在流模式下重新生成并重新安装 ABAP 程序。
2. 创建 SAP ABAP HTTP 流式处理连接。
3. 将会话配置为使用 SAP 流读取器、SAP ABAP HTTP 流连接和 SAP R/3 应用程序连接。

注意：如果您配置了 HTTP 流，但未在流模式下重新生成并重新安装 ABAP 程序，会话将失败。

第 VI 部分： 版本 10.1.1

本部分包含以下章节：

- [新功能、更改和发布任务 \(10.1.1 HotFix 1\), 300](#)
- [新功能、更改和发布任务 \(10.1.1 Update 2\), 305](#)
- [新功能、更改和发布任务 \(10.1.1 Update 1\), 312](#)
- [新产品 \(10.1.1\), 314](#)
- [新功能 \(10.1.1\), 316](#)
- [更改 \(10.1.1\), 335](#)
- [发布任务 \(10.1.1\), 345](#)

第 28 章

新功能、更改和发布任务 (10.1.1 HotFix 1)

本章包括以下主题：

- [新产品 \(10.1.1 HotFix 1\), 300](#)
- [新功能 \(10.1.1 HotFix 1\), 300](#)
- [更改 \(10.1.1 HotFix 1\), 303](#)

新产品 (10.1.1 HotFix 1)

本节介绍版本 10.1.1 HotFix 1 中的新产品。

PowerExchange for Cloud Applications

从版本 10.1.1 HotFix 1 开始，您可以使用 PowerExchange for Cloud Applications 从 PowerCenter 连接到 Informatica Cloud。如果连接在 Informatica Cloud 中可用，您可以从数据源中读取数据或向其写入数据。PowerCenter 中不需要有适用于各个云应用程序的 PowerExchange。

有关详细信息，请参阅《*Informatica PowerExchange for Cloud Applications 10.1.1 HotFix 1 用户指南*》。

新功能 (10.1.1 HotFix 1)

本节介绍版本 10.1.1 HotFix 1 中的新功能。

命令行程序

本节介绍版本 10.1.1 HotFix 1 中新增的命令。

infacmd dis 命令 (10.1.1 HF1)

下表介绍了新的 infacmd dis 命令：

命令	说明
disableMappingValidationEnvironment	为部署到数据集成服务的映射禁用映射验证环境。
enableMappingValidationEnvironment	为部署到数据集成服务的映射启用映射验证环境。
setMappingExecutionEnvironment	为部署到数据集成服务的映射指定映射执行环境。

有关详细信息，请参阅《Informatica 10.1.1 HotFix1 命令引用》中的“infacmd dis 命令引用”一章。

infacmd mrs 命令 (10.1.1 HF1)

下表介绍了新的 infacmd mrs 命令：

命令	说明
disableMappingValidationEnvironment	为从 Developer tool 运行的映射禁用映射验证环境。
enableMappingValidationEnvironment	为从 Developer tool 运行的映射启用映射验证环境。
setMappingExecutionEnvironment	为从 Developer tool 运行的映射指定映射执行环境。

有关详细信息，请参阅《Informatica 10.1.1 HotFix1 命令引用》中的“infacmd mrs 命令引用”一章。

infacmd ps 命令

下表介绍了新的 infacmd ps 命令：

命令	说明
restoreProfilesAndScorecards	将配置文件和结果卡从以前的版本还原到版本 10.1.1 HotFix 1。

有关详细信息，请参阅《Informatica 10.1.1 HotFix 1 命令引用》中的“infacmd ps 命令引用”一章。

Informatica Analyst

本节介绍版本 10.1.1 HotFix 1 中新增的 Analyst 工具功能。

配置文件和结果卡

本节介绍针对配置文件和结果卡新增的 Analyst 工具功能。

无效行工作表

从版本 10.1.1 HotFix1 开始，在将数据导出到文件对话框中选择数据 > 全部选项后，结果卡导出结果会包含无效的源行。

有关结果卡的详细信息，请参阅《Informatica 10.1.1 HotFix1 Data Discovery 指南》中的“Informatica Analyst 中的结果卡”一章。

PowerCenter

本节介绍版本 10.1.1 HotFix 1 中新增的 PowerCenter 功能。

针对 Greenplum 的下推优化

从版本 10.1.1 HotFix 1 开始，当连接类型为 ODBC 时，PowerCenter 集成服务可以使用源端下推优化和完全下推优化将 TRUNC(DATE)、CONCAT() 和 TO_CHAR(DATE) 函数推送到 Greenplum。

有关详细信息，请参阅《*Informatica PowerCenter 10.1.1 HotFix 1 高级工作流指南*》。

下推优化 Microsoft Azure SQL 数据仓库

从版本 10.1.1 HotFix 1 开始，如连接类型为 ODBC，则可配置源侧或完整下推优化，将转换逻辑推送到 Microsoft Azure SQL 数据仓库。

有关详细信息，请参阅《*Informatica PowerCenter 10.1.1 HotFix 1 高级工作流指南*》。

PowerExchange 适配器

本节介绍版本 10.1.1 HotFix 1 中新增的 PowerExchange 适配器功能。

PowerCenter(R) PowerExchange 适配器

本节介绍版本 10.1.1 HotFix 1 中新增的 PowerCenter 适配器功能。

PowerExchange for Amazon Redshift

本节介绍版本 10.1.1 HotFix 1 中新增的 PowerExchange for Amazon Redshift 功能：

- 可以在以下区域读取或写入数据：
 - 亚太地区（孟买）
 - 加拿大（中部）
 - 美国东部（俄亥俄州）
- PowerExchange for Amazon Redshift 支持星号下推运算符 (*)，使用源端、目标端或完整下推优化可将该运算符推送到 Amazon Redshift 数据库。
- 对于客户端和服务端加密，可以在连接中配置由 AWS 密钥管理服务 (AWS KMS) 生成的客户主密钥 ID。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica 10.1.1 HotFix 1 PowerExchange for Amazon Redshift 用户指南*》。

PowerExchange for Amazon S3

本节介绍版本 10.1.1 HotFix 1 中新增的 PowerExchange for Amazon S3 功能：

- 可以在以下区域读取或写入数据：
 - 亚太地区（孟买）
 - 加拿大（中部）
 - 美国东部（俄亥俄州）
- 对于客户端和服务端加密，可以在连接中配置由 AWS 密钥管理服务 (AWS KMS) 生成的客户主密钥 ID。
- 向 Amazon S3 存储桶写入数据时，可以采用 GZIP 格式压缩数据。
- 您可以在运行映射时替代 Amazon S3 文件夹路径。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica PowerExchange for Amazon S3 10.1.1 HotFix 1 用户指南*》。

PowerExchange for Microsoft Azure Blob Storage

从版本 10.1.1 HotFix 1 开始，您可以使用附加 blob 类型目标会话属性向 Microsoft Azure Blob Storage 写入数据。

有关详细信息，请参阅《*Informatica PowerExchange for Microsoft Azure Blob Storage 10.1.1 HotFix 1 用户指南*》。

PowerExchange for Microsoft Azure SQL Data Warehouse

从版本 10.1.1 HotFix 1 开始，您可以在 PowerExchange for Microsoft Azure SQL Data Warehouse 中使用以下目标会话属性：

- 更新为更新内容。PowerCenter 集成服务将所有行更新为更新内容。
- 更新否则插入。PowerCenter 集成服务将更新现有行，并插入其他行（就像这些行已标记为要插入一样）。
- 删除。PowerCenter 集成服务从 Microsoft Azure SQL Data Warehouse 中删除指定的记录。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica PowerExchange for Microsoft Azure SQL Data Warehouse 10.1.1 HotFix 1 用户指南*》。

PowerExchange for Microsoft Dynamics CRM

从版本 10.1.1 HotFix 1 开始，您可以在 PowerExchange for Microsoft Dynamics CRM 中使用以下目标会话属性：

- 添加行拒绝原因。选中该属性可将拒绝行的原因包含到拒绝文件中。
- 备用键名称。指示列是否为实体的备用键。指定备用键的名称。可以在更新和 upsert 操作中使用备用键。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica PowerExchange for Microsoft Dynamics CRM 10.1.1 HotFix 1 用户指南*》。

PowerExchange for SAP NetWeaver

从版本 10.1.1 HotFix 1 开始，通过 ABAP 读取 SAP 表中的数据时，PowerExchange for SAP NetWeaver 支持 SSTRING 数据类型。在 PowerCenter 中，SSTRING 数据类型以 SSTR 形式表示。

有关详细信息，请参阅《*Informatica PowerExchange for SAP NetWeaver 10.1.1 HotFix 1 用户指南*》。

更改 (10.1.1 HotFix 1)

本节介绍版本 10.1.1 HotFix 1 中的更改。

支持更改

从版本 10.1.1 HF1 开始，以下更改适用于第三方平台和系统的 Informatica 支持：

Big Data Management Hadoop 发行版

下表列出了 10.1.1 HotFix 1 中支持的 Hadoop 发行版版本和更改：

发行版	支持的版本	10.1.1 HotFix 1 中的更改
Amazon EMR	5.4	为了能够支持 Amazon EMR 5.4，请将 EBF-9585 应用于 Big Data Management 10.1.1 HotFix 1。 Big Data Management 版本 10.1.1 Update 2 支持 Amazon EMR 5.0。
Azure HDInsight	3.5	增加了对版本 3.5 的支持。
Cloudera CDH	5.8、5.9、5.10 和 5.11	增加了对版本 5.10 和 5.11 的支持。
Hortonworks HDP	2.3、2.4、2.5 和 2.6	增加了对版本 2.6 的支持。
IBM BigInsights	4.2	无更改。
MapR	5.2.0 MEP 二进制文件 版本 1.0	无更改。

要查看支持的版本的最新列表，请参阅 Informatica 客户门户上的“产品可用性列表”，网址为：
<https://network.informatica.com/community/informatica-network/product-availability-matrices>。

第 29 章

新功能、更改和发布任务 (10.1.1 Update 2)

本章包括以下主题：

- [新产品 \(10.1.1 Update 2\), 305](#)
- [新功能 \(10.1.1 Update 2\), 305](#)
- [更改 \(10.1.1 Update 2\), 308](#)

新产品 (10.1.1 Update 2)

本节介绍版本 10.1.1 Update 2 中的新产品。

PowerExchange for MapR-DB

从版本 10.1.1 Update 2 开始，您可以使用 PowerExchange for MapR-DB 从 MapR-DB 二进制表读取数据以及向其写入数据。

PowerExchange for MapR-DB 使用 HBase API 连接到 MapR-DB。要连接到 MapR-DB 表，必须创建一个 HBase 连接，其中您必须将数据库类型指定为 MapR-DB。您必须创建 HBase 数据对象读取或写入操作并将其添加到映射中，以便读取或写入数据。

可以在本地环境中或在 Hadoop 环境中的 Blaze 引擎上验证和运行映射。

有关详细信息，请参阅《*Informatica PowerExchange for MapR-DB 10.1.1 Update 2 用户指南*》。

新功能 (10.1.1 Update 2)

本节介绍版本 10.1.1 Update 2 中的新功能。

Big Data Management

本节介绍版本 10.1.1 Update 2 中新增的大数据功能。

在使用 Blaze 运行时引擎的映射上截断 Hive 表分区

从版本 10.1.1 Update 2 开始，您可以在使用 Blaze 运行时引擎的映射上截断 Hive 表分区。

有关截断 Hive 目标中的分区的详细信息，请参阅《*Informatica 10.1.1 Update 2 Big Data Management 用户指南*》。

Blaze 引擎上用于已分区列的筛选器

从版本 10.1.1 Update 2 开始，Blaze 引擎可以将已分区列上的筛选器下推到 Hive 源以提高性能。

当映射在 Hive 源的已分区列上包含筛选器转换时，Blaze 引擎只读取包含满足筛选条件的数据的数据分区。要使 Blaze 引擎能够读取特定的分区，筛选器转换在映射中必须是源之后的下一个转换。

有关详细信息，请参阅《*Informatica 10.1.1 Update 2 Big Data Management 用户指南*》。

Spark 引擎上的 OraOop 支持

从版本 10.1.1 Update 2 开始，您可以对 OraOop 进行配置以在 Spark 引擎上运行 Sqoop 映射。从 Oracle 读取数据或向其写入数据时，您可以配置直接参数以便 Sqoop 能够使用 OraOop。

OraOop 是 Oracle 专用的 Sqoop 插件，它使用本地协议连接到 Oracle 数据库。配置 OraOop 时，性能会提高。

有关详细信息，请参阅《*Informatica 10.1.1 Update 2 Big Data Management 用户指南*》。

Cloudera 群集中本地 Teradata 映射的 Sqoop 支持

从版本 10.1.1 Update 2 开始，如果您使用 Teradata PT 连接在 Cloudera 群集和 Blaze 引擎上运行映射，数据集成服务会在运行时调用由 Teradata 提供技术支持的 Cloudera Connector。然后，数据集成服务会通过 Sqoop 运行映射。

有关详细信息，请参阅《*Informatica 10.1.1 Update 2 PowerExchange for Teradata Parallel Transporter API 用户指南*》。

Blaze 引擎和 Spark 引擎上的计划程序支持

从版本 10.1.1 Update 2 开始，以下计划程序可用于 Blaze 引擎和 Spark 引擎上的 Hadoop 发行版：

- 公平计划程序。将资源分配给作业，以便所有作业在一段时间内收到相等的平均资源份额。
- 容量计划程序。用于将 Hadoop 应用程序作为多租户共享群集运行。您可以配置容量计划程序，不论其是否带有节点标签。节点标签可用于将具有相似特征的节点分组到一起。

有关详细信息，请参阅《*Informatica 10.1.1 Update 2 Big Data Management 用户指南*》的“Hadoop 环境中的映射”一章。

Blaze 引擎和 Spark 引擎上对 YARN 队列的支持

从版本 10.1.1 Update 2 开始，您可以将 Blaze 作业和 Spark 作业定向到特定的 YARN 计划程序队列。队列允许多个租户共享群集。当您将应用程序提交到 YARN 时，计划程序会将它们分配到一个队列。您可以在 Hadoop 连接属性中配置 YARN 队列。

有关详细信息，请参阅《*Informatica 10.1.1 Update 2 Big Data Management 用户指南*》的“Hadoop 环境中的映射”一章。

IBM BigInsights 4.2 上的 Hadoop 安全功能

从版本 10.1.1 Update 2 开始，您可以在 IBM BigInsights 4.2 Hadoop 发行版上使用以下 Hadoop 安全功能：

- Apache Knox
- Apache Ranger
- HDFS 透明加密

有关详细信息，请参阅《*Informatica 10.1.1 Update 2 Big Data Management 安全指南*》

SSL/TLS 安全模式

从版本 10.1.1 Update 2 开始，您可以在 Cloudera 和 HortonWorks Hadoop 发行版上使用 SSL 和 TLS 安全模式，包括以下安全方法和插件：

- Kerberos 身份验证
- Apache Ranger
- Apache Sentry
- 名称节点的高可用性
- 资源管理器的高可用性

有关详细信息，请参阅《*Informatica 10.1.1 Update 2 Big Data Management 安装和配置指南*》。

Amazon S3 上的 Hive 源和目标

从版本 10.1.1 Update 2 开始，对于配置了以下 Hadoop 发行版的群集，Big Data Management 支持从 Amazon S3 存储桶上的 Hive 读取以及向其写入：

- Amazon EMR
- Cloudera
- HortonWorks
- MapR
- BigInsights

有关详细信息，请参阅《*Informatica 10.1.1 Update 2 Big Data Management 用户指南*》。

Enterprise Information Catalog

本节介绍 Enterprise Information Catalog 版本 10.1.1 Update 2 中的新功能。

文件系统资源

从版本 10.1.1 Update 2 开始，您可以创建**文件系统**资源，以便从 Windows 和 Linux 文件系统文件中的文件导入元数据。

有关详细信息，请参阅《*Informatica 10.1.1 Update 2 Live Data Map 管理员指南*》。

已启用 Apache Ranger 的群集

从版本 10.1.1 Update 2 开始，您可以在已启用 Apache Ranger 的群集上部署 Enterprise Information Catalog。Apache Ranger 可提供一个安全框架来管理群集的安全。

用于部署 Informatica 群集服务的增强 SSH 支持

从版本 10.1.1 Update 2 开始，您可以在启用了 Centrify 的主机上部署 Informatica 群集服务。Centrify 与现有 Active Directory 基础架构集成，以管理远程 Linux 主机上的用户身份验证。

Intelligent Data Lake

本节介绍版本 10.1.1 Update 2 中新增的 Intelligent Data Lake 功能。

Hadoop 生态系统

从版本 10.1.1 Update 2 开始，您可以使用以下 Hadoop 发行版作为 Hadoop 数据湖：

- Cloudera CDH 5.9
- Hortonworks HDP 2.3、2.4 和 2.5

- Azure HDInsight 3.5
- Amazon EMR 5.0
- IBM BigInsights 4.2

对数据准备服务使用 MariaDB

从版本 10.1.1 Update 2 开始，您可以对数据准备服务存储库使用 MariaDB 10.0.28。

查看列级沿袭

从版本 10.1.1 Update 2 开始，数据分析人员可以查看活动（例如数据资产复制、导入、导出、发布和上传）对应的表中各个列的沿袭。

SSL/TLS 支持

从版本 10.1.1 Update 2 开始，您可以将 Intelligent Data Lake 与已启用 SSL/TLS 的 Cloudera 5.9 群集集成。

Informatica PowerExchange 适配器

本节介绍版本 10.1.1 Update 2 中新增的 Informatica 适配器功能。

PowerExchange for Amazon Redshift

从版本 10.1.1 Update 2 开始，您可以为 Amazon Redshift 对象选择多个架构。

有关详细信息，请参阅《Informatica 10.1.1 Update 2 PowerExchange for Amazon Redshift 用户指南》。

更改 (10.1.1 Update 2)

本节介绍版本 10.1.1 Update 2 中的更改。

支持更改

本节介绍版本 10.1.1 Update 2 中的支持更改。

Big Data Management 的发行版支持更改

下表列出了 10.1.1 Update 2 中支持的 Hadoop 发行版版本和更改：

发行版	支持的版本	10.1.1 Update 2 中的更改
Amazon EMR	5.0.0	无更改。
Azure HDInsight	3.5 *	增加了对版本 3.5 的支持 终止了对版本 3.4 的支持。
ClouderaCDH	5.8、5.9 和 5.10 *	增加了对版本 5.10 的支持。
Hortonworks HDP	2.3、2.4 和 2.5	增加了对版本 2.3 和 2.4 的支持。
IBM BigInsights	4.2	无更改。

发行版	支持的版本	10.1.1 Update 2 中的更改
MapR	5.2	恢复了支持。 增加了对版本 5.2 的支持。 终止了对版本 5.1 的支持。
* Azure HDInsight 3.5 和 Cloudera CDH 5.10 可用于技术预览版。支持技术预览版功能，但尚未做好生产准备。Informatica 建议您只在非生产环境中使用。		

有关 Hadoop 支持的完整列表，请参阅 Informatica Network 上的“Product Availability Matrix”（产品可用性列表）：

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

终止为 Blaze 引擎上的 Teradata Connector for Hadoop (TDCH) 和 Teradata PT 对象提供支持

从版本 10.1.1 Update 2 开始，Informatica 已终止为 Blaze 引擎上的 Teradata Connector for Hadoop (TDCH) 提供支持。10.1.1 Update 2 中的 Sqoop 连接配置取决于 Hadoop 发行版：IBM Biginsights 和 MapR

您可以通过 JDBC 连接配置 Sqoop 连接。有关通过 JDBC 连接配置 Sqoop 连接的信息，请参阅《Informatica 10.1.1 Update 2 Big Data Management 用户指南》。

ClouderaCDH

您可以通过 Teradata PT 连接和由 Teradata 提供技术支持的 Cloudera Connector 来配置 Sqoop 连接。

1. 下载由 Teradata 提供技术支持的 Cloudera Connector 的 .jar 文件，并将它们复制到运行数据集成服务的节点。有关详细信息，请参阅《Informatica 10.1.1 Update 2 PowerExchange for Teradata Parallel Transporter API 用户指南》。
2. 将在 InfaTDCHConfig.txt 文件中定义的配置参数移动到 Teradata PT 连接中的**其他 Sqoop 参数**字段中。有关您可以指定的参数列表，请参阅由 Teradata 提供技术支持的 Cloudera Connector 文档。

Hortonworks HDP

您可以通过 Teradata PT 连接和 Hortonworks Connector for Teradata 来配置 Sqoop 连接。

1. 下载 Hortonworks Connector for Teradata 的 .jar 文件，并将它们复制到运行数据集成服务的节点。有关详细信息，请参阅《Informatica 10.1.1 Update 2 PowerExchange for Teradata Parallel Transporter API 用户指南》。
2. 将在 InfaTDCHConfig.txt 文件中定义的配置参数移动到 Teradata PT 连接中的**其他 Sqoop 参数**字段中。有关您可以指定的参数列表，请参见 Hortonworks Connector for Teradata 文档。

注意：您可以继续通过 Teradata PT 连接在 Hive 引擎上使用 TDCH。

已停止对通过 Teradata PT 数据对象和 Teradata PT 连接的 Sqoop 连接的支持

从版本 10.1.1 Update 2 开始，Informatica 已停止为 Cloudera CDH 和 Hortonworks 提供通过 Teradata PT 数据对象和 Teradata PT 连接的 Sqoop 连接。将在未来版本中停止提供支持。

要使用 TDCH 和 Sqoop 从 Teradata 读取数据或向其写入数据，Informatica 建议您通过 JDBC 连接和关系数据对象来配置 Sqoop 连接。

Big Data Management

本节介绍版本 10.1.1 Update 2 中对大数据进行的更改。

Sqoop

从版本 10.1.1 Update 2 开始，无法再使用 `--username` 和 `--password` 参数替代 Sqoop 映射中的用户名和密码。Sqoop 会使用您在 JDBC 连接的**用户名**和**密码**字段中配置的值。

有关详细信息，请参阅《*Informatica 10.1.1 Update 2 Big Data Management 用户指南*》。

Enterprise Information Catalog

本节介绍版本 10.1.1 Update 2 中对 Enterprise Information Catalog 进行的更改。

资产路径

从版本 10.1.1 Update 2 开始，您可以在“资产详细信息”视图中查看资产的路径，以及有关该资产的其他常规信息。

有关详细信息，请参阅《*Informatica 10.1.1 Update 2 Enterprise Information Catalog 用户指南*》。

配置文件结果部分中的业务术语

从版本 10.1.1 Update 2 开始，Tabular 资产的配置文件结果部分还包含业务术语。在以前的版本中，配置文件结果部分包含列名称、数据类型和数据域。

有关详细信息，请参阅《*Informatica 10.1.1 Update 2 Enterprise Information Catalog 用户指南*》。

使用 URL 作为属性值

从版本 10.1.1 Update 2 开始，如果您已将自定义属性配置为允许输入 URL 作为属性值，则可以将多个 URL 作为属性值分配给技术资产。

有关详细信息，请参阅《*Informatica 10.1.1 Update 2 Enterprise Information Catalog 用户指南*》。

检测 CSV 文件标头

从版本 10.1.1 Update 2 开始，您可以配置以下资源，以便自动检测从中提取元数据的 CSV 文件的标头：

- Amazon S3
- HDFS
- 文件系统

有关详细信息，请参阅《*Informatica 10.1.1 Update 2 Live Data Map 管理员指南*》。

Amazon Redshift 资源

从版本 10.1.1 Update 2 开始，您可以为 Amazon Redshift 资源导入多个架构。

有关详细信息，请参阅《*Informatica 10.1.1 Update 2 Live Data Map 管理员指南*》。

数据集成服务上的 Hive 资源剖析

从版本 10.1.1 Update 2 开始，您可以在数据集成服务上运行 Hive 资源以进行剖析。

有关详细信息，请参阅《*Informatica 10.1.1 Update 2 Live Data Map 管理员指南*》。

Informatica PowerExchange 适配器

本节介绍版本 10.1.1 Update 2 中对 Informatica 适配器进行的更改。

PowerExchange for Amazon Redshift

从版本 10.1.1 Update 2 开始，您可以为 Amazon Redshift 对象选择多个架构。要选择多个架构，请将连接属性中的**架构**字段留空。在早期版本中，您必须选择架构，并且只能选择一个架构。

如果升级到版本 10.1.1 Update 2，则在早期版本中创建的 PowerExchange for Redshift 映射在连接属性中必须具有相关的架构名称。否则，当您在版本 10.1.1 Update 2 上运行这些映射时，映射会失败。

有关详细信息，请参阅《*Informatica 10.1.1 Update 2 PowerExchange for Amazon Redshift 用户指南*》。

第 30 章

新功能、更改和发布任务 (10.1.1 Update 1)

本章包括以下主题：

- [新功能 \(10.1.1 Update 1\), 312](#)
- [更改 \(10.1.1 Update 1\), 312](#)
- [发布任务 \(10.1.1 Update 1\), 313](#)

新功能 (10.1.1 Update 1)

本节介绍版本 10.1.1 Update 1 中的新功能。

Big Data Management

本节介绍版本 10.1.1 Update 1 中新增的大数据功能。

本地 Teradata 映射的 Sqoop 支持

从版本 10.1.1 Update 1 开始，如果您使用 Teradata PT 连接在 Hortonworks 群集和 Blaze 引擎上运行映射，数据集成服务会在运行时调用 Hortonworks Connector for Teradata。然后，数据集成服务会通过 Sqoop 运行映射。

有关详细信息，请参阅《*Informatica 10.1.1 Update 1 PowerExchange for Teradata Parallel Transporter API 用户指南*》。

本地 Teradata 映射的 SQL 替代支持

从版本 10.1.1 Update 1 开始，如果您使用 Teradata PT 连接在 Hortonworks 群集和 Blaze 引擎上运行映射，那么您可以配置 SQL 替代查询。您还可以参数化 SQL 替代查询。

有关详细信息，请参阅《*Informatica 10.1.1 Update 1 PowerExchange for Teradata Parallel Transporter API 用户指南*》。

更改 (10.1.1 Update 1)

本节介绍版本 10.1.1 Update 1 中的更改。

Informatica PowerExchange 适配器

本节介绍版本 10.1.1 Update 1 中的 PowerExchange 适配器更改。

PowerExchange for Amazon S3

从版本 10.1.1 Update 1 开始，PowerExchange for Amazon S3 具有以下用于 Amazon S3 数据对象读取和写入操作的高级属性：

- 文件夹路径
- 分多个部分下载 S3 文件
- 暂存目录

在以前的版本中，用于 Amazon S3 数据对象读取和写入操作的高级属性为：

- S3 文件夹路径
- 启用分多个部分下载 S3 文件
- 本地 Temp 文件夹路径

有关详细信息，请参阅《*Informatica 10.1.1 Update 1 PowerExchange for Amazon S3 用户指南*》。

发布任务 (10.1.1 Update 1)

本节介绍版本 10.1.1 Update 1 的发布任务。

Informatica PowerExchange 适配器

本节介绍版本 10.1.1 Update 1 的 PowerExchange 适配器发布任务。

PowerExchange for Teradata Parallel Transporter API

从版本 10.1.1 Update 1 开始，如果您使用 Teradata PT 连接在 Hortonworks 群集和 Blaze 引擎上运行映射，数据集成服务会在运行时调用 Hortonworks Connector for Teradata。然后，数据集成服务会通过 Sqoop 运行映射。

如果您已将 Teradata Connector for Hadoop (TDCH) 配置为在 Blaze 引擎上运行 Teradata 映射并安装了 10.1.1 Update 1，则数据集成服务将忽略 TDCH 配置。必须执行以下升级任务才能在 Blaze 引擎上运行 Teradata 映射：

1. 安装 10.1.1 Update 1。
2. 下载 Hortonworks Connector for Teradata 的 JAR 文件。
3. 将在 InfaTDCHConfig.txt 文件中定义的配置参数移动到 Teradata PT 连接中的**其他 Sqoop 参数**字段中。有关您可以指定的参数列表，请参阅 Hortonworks Connector for Teradata 文档。

注意：如果您已将 TDCH 配置为在 Blaze 引擎和 Hortonworks 以外的其他发行版上运行 Teradata 映射，请不要安装 10.1.1 Update 1。您可以继续使用版本 10.1.1 在 Blaze 引擎和 Hortonworks 以外的其他发行版上通过 TDCH 运行映射。

有关详细信息，请参阅《*Informatica 10.1.1 Update 1 PowerExchange for Teradata Parallel Transporter API 用户指南*》。

第 31 章

新产品 (10.1.1)

本章包括以下主题：

- [智能流, 314](#)

智能流

随着大数据技术的出现，组织希望从数据速度获得最大利益，还希望在数据变为可用时立即捕获并处理数据，以及实时对事件做出响应。通过添加实时流式传输功能，组织可以从降低的延迟获得以下好处：创建完整的最新客户视图，向客户提供实时的运营情报，改进欺诈检测，降低安全风险，改进有形资产管理，改善整体客户体验，以及使决策制定流程整体上提高几个量级。

在 10.1.1 中，Informatica 引入了新产品 Intelligent Streaming，此产品可帮助 IT 组织通过以下方法从实时队列获得最大价值：流式传输并处理数据，然后以近实时方式提取有意义的业务价值。客户能够以较高的准确度实时处理各种数据类型，以及来自非传统源的数据类型（例如，网站日志文件数据、传感器数据、消息总线数据和计算机数据）。

Intelligent Streaming 构建为 Informatica Intelligent Data Platform 的功能扩展，并为 IT 组织提供以下好处：

- 创建和运行流式传输（持续处理）映射。
- 从实时队列（例如 Apache Kafka 和 JMS）收集事件。
- 转换数据，为转换后的数据创建业务规则，检测实时模式，以及驱动自动响应或警告。
- 提供在运行时管理和监视流的功能。
- 提供至少一次的交付保证。
- 基于处理的行数或执行时间将生命周期控制粒度化。
- 重用和维护事件处理逻辑，包括批处理映射（在进行了一些修改之后）。

Intelligent Streaming 具有以下功能：

捕获和传输流数据

可以从诸如 Kafka 或 JMS 等源以 JSON、XML 或 Avro 格式流式传输以下数据类型：

- 应用程序日志数据和基础结构日志数据
- 来自关系数据库的更改数据捕获 (Change Data Capture, CDC)
- 来自 Web 服务器的点击流
- 社交媒体事件流

- 来自物联网 (Internet Of Things, IoT) 设备的时间序列数据
- 消息总线数据
- 可编程逻辑控制器 (Programmable Logic Controller, PLC) 数据
- 来自设备的销售点数据

此外, Informatica 客户还可以利用 Informatica Vibe Data Stream (单独许可) 实时收集数据 (例如, 传感器数据和计算机日志数据) 以及将数据插入到 Kafka 队列。然后, Intelligent Streaming 便可以处理该数据。

细化、扩充、分析和处理流数据

可以使用基础处理平台实时运行以下复杂数据转换, 而无需编写代码或脚本:

- 适用于流式传输的窗口转换 (带有滑动和翻转窗口的选项) 用例。
- 现在, 筛选器转换、表达式转换、联合转换、路由器转换、汇总转换、连接器转换、查找转换、Java 转换、排序器转换可以与流式传输映射结合使用, 并在 Spark 流式传输上执行。
- 查找转换可以与平面文件、HDFS、Sqoop 和 Hive 结合使用。

发布数据

可以将数据流式传输到各种类型的目标, 例如 Kafka、HDFS 和 NoSQL 数据库以及企业消息传送系统。

Intelligent Streaming 构建于 Informatica Big Data Platform 平台之上, 并为该平台扩展了流式传输功能。Intelligent Streaming 使用 Spark 流式传输来处理流数据。它使用 YARN 更高效地管理 Spark 群集上的资源, 并使用第三方分发来连接到 Hadoop 环境以及向其推送作业处理。

可以使用 Informatica Developer (Developer tool) 创建流式传输映射。然后可以使用 Hadoop 运行时环境和 Spark 引擎运行该映射。可以配置高可用性以在 Hadoop 群集上运行流式传输映射。

有关 Intelligent Streaming 的详细信息, 请参阅《*Informatica Intelligent Streaming 用户指南*》。

第 32 章

新功能 (10.1.1)

本章包括以下主题：

- [应用程序服务, 316](#)
- [大数据, 317](#)
- [Business Glossary , 320](#)
- [命令行程序, 321](#)
- [Enterprise Information Catalog, 322](#)
- [Informatica Analyst, 325](#)
- [Informatica 安装, 325](#)
- [Intelligent Data Lake, 326](#)
- [映射 , 327](#)
- [Metadata Manager, 327](#)
- [PowerExchange 适配器, 328](#)
- [安全, 329](#)
- [转换, 330](#)
- [Web 服务 , 333](#)
- [工作流, 333](#)

应用程序服务

本节介绍版本 10.1.1 中新增的系统服务功能。

分析服务

从版本 10.1.1 开始，可以对分析服务进行配置，使其将异常管理任务的所有审计数据存储存储在单个数据库中。该数据库会存储 Analyst 工具用户对分析服务指定的人工任务实例执行的工作的记录。

可以在 Administrator 工具中在分析服务的人工任务属性上设置审计表的数据库连接和架构。指定连接和架构之后，可以使用 Administrator 工具中的**操作**菜单选项来创建审计数据库内容。或者，也可以使用 infacmd as 命令来设置数据库和架构以及创建审计数据库内容。要设置数据库和架构，请运行 infacmd as updateServiceOptions。要创建数据库内容，请运行 infacmd as createExceptionAuditTables。

如果不指定连接和架构，则分析服务会在存储任务实例数据的数据库中创建每个任务实例的审计数据。

有关详细信息，请参阅《*Informatica 10.1.1 应用程序服务指南*》和《*Informatica 10.1.1 命令参考*》。

大数据

本节介绍版本 10.1.1 中新增的大数据功能。

Blaze 引擎

从版本 10.1.1 开始，Blaze 引擎具有以下新功能：

Blaze 引擎上的 Hive 源和目标

从版本 10.1.1 开始，Blaze 引擎上的 Hive 源和目标具有以下附加支持：

- 精度为 38 的 Hive 十进制数据类型值
- Hive 表名称、列名称和架构名称中带引号的标识符
- 使用已分区的 Hive 表作为目标
- 使用位于存储桶中的 Hive 表作为源和目标
- Hive 源的 SQL 替代
- Hive 源和目标的表锁定
- 创建或替换 Hive 目标的目标表
- 截断 Hive 目标和 Hive 分区表的目标表

有关详细信息，请参阅《*Informatica Big Data Management 10.1.1 用户指南*》中的“Hadoop 环境中的映射对象”一章。

Blaze 引擎上的转换支持

从版本 10.1.1 开始，Blaze 引擎上的转换具有以下附加支持：

- 查找转换。可以将 SQL 替代和筛选器查询与 Hive 查找源结合使用。
- 排序器转换。当排序器转换连接到平面文件目标时，支持全局排序。要维持全局排序顺序，必须在平面文件目标中启用“维持行顺序”属性。如果排序器转换位于映射的中游，则会在本地对行进行排序。
- 更新策略转换。支持更新策略转换，但存在一些限制。

有关详细信息，请参阅《*Informatica Big Data Management 10.1.1 用户指南*》中的“Hadoop 环境中的映射对象”一章。

Blaze 引擎监视

从版本 10.1.1 开始，Blaze 摘要报告中提供了有关映射作业的更详细统计信息。在 Blaze 作业监视器中，成功的网格任务的名称旁边会显示一个绿色的摘要报告按钮，单击此按钮可打开 Blaze 摘要报告。

Blaze 摘要报告包含有关映射作业的以下信息：

- 各个段所用的时间。关于网格任务中的段的饼图。
- 映射属性。包含有关映射作业的基本信息的表。
- 小任务的执行时间。关于选定段中的所有小任务的时间序列图。
- 选定的小任务的信息。各个小任务的源行计数和目标行计数以及缓存信息。

注意：Blaze 摘要报告仍处于测试阶段。它包含大部分主要功能，但尚不完整。

Blaze 引擎日志

从版本 10.1.1 开始，Blaze 引擎上提供了以下错误日志记录增强：

- 当日志跟踪级别设置为详细初始化或详细数据时，LDTM 日志中会提供执行统计信息。该日志包含以下映射执行详细信息：
 - 每个任务的开始时间、结束时间和状态
 - Blaze 作业监视器 URL
 - 小任务总数以及成功和失败/已取消的小任务的数量
 - 源和目标的已处理行数以及已拒绝行数
 - 每个已执行段中的转换的数据错误（如果有）
- LDTM 日志包含以下转换统计信息：
 - 源和目标的输出行数
 - 源和目标的错误行数
- 会话日志还会显示网格任务中所有段的列表以及相应的 Blaze 作业监视器链接。单击链接可查看相应段的执行详细信息。

有关详细信息，请参阅《*Informatica Big Data Management 用户指南*》中的“监视 Hadoop 环境中的映射”一章。

安装和配置

本节介绍与大数据安装和配置相关的新功能。

地址引用数据安装

从版本 10.1.1 开始，Informatica Big Data Management 随 shell 脚本一起安装，您可以使用该脚本来安装地址引用数据文件。该脚本将在您指定的计算节点上安装引用数据文件。

在 Hadoop 环境中运行地址验证映射时，引用数据文件必须驻留在要对其运行映射的每个计算节点上。使用该脚本可通过单个操作在多个节点上安装引用数据文件。

该 shell 脚本的名称是 `copyRefDataToComputeNodes.sh`。

可以在 Informatica Big Data Management 安装中的以下目录找到该脚本：

[Informatica 安装目录]/tools/dq/av

运行该脚本时，可以输入以下信息：

- 引用数据文件的当前位置。
- 该脚本将文件安装到的目录。
- 包含计算节点名称的文件的位置。
- 运行该脚本的用户的用户名。

如果您未输入这些信息，该脚本将使用一系列默认值来标识文件位置和用户名。

有关详细信息，请参阅《*Informatica Big Data Management 10.1.1 安装和配置指南*》。

静默模式下的 Hadoop 配置管理器

从版本 10.1.1 开始，可以在静默模式下使用 Hadoop 配置管理器来配置 Big Data Management。

有关在静默模式下配置 Big Data Management 的详细信息，请参阅《*Informatica Big Data Management 10.1.1 安装和配置指南*》。

在 Ambari 堆栈中安装

从版本 10.1.1 开始，可以使用 Ambari 配置管理器在 Ambari 堆栈中安装 Big Data Management 作为服务。

有关在 Ambari 堆栈中安装 Big Data Management 的详细信息，请参阅《*Informatica 10.1.1 Big Data Management 安装和配置指南*》。

用于在 HDInsight 群集中填充 HDFS 的脚本

从版本 10.1.1 开始，在针对 Big Data Management 配置群集时，可以使用脚本在 Azure HDInsight 群集上填充 HDFS 文件系统。

有关使用脚本填充 HDFS 文件的详细信息，请参阅《*Informatica Big Data Management 10.1.1 安装和配置指南*》。

Spark 引擎

从版本 10.1.1 开始，Spark 引擎具有以下新功能：

二进制数据类型

从版本 10.1.1 开始，Spark 引擎对以下函数支持二进制数据类型：

- DEC_BASE64
- ENC_BASE64
- MD5
- UUID4
- UUID_UNPARSE
- CRC32
- COMPRESS
- DECOMPRESS（忽略精度）
- AES 加密
- AES 解密

注意：Spark 引擎对联接条件和查找条件不支持二进制数据类型。

有关详细信息，请参阅《*Informatica Big Data Management 10.1.1 用户指南*》中的“函数引用”一章。

Spark 引擎上的转换支持

从版本 10.1.1 开始，Spark 引擎上的转换具有以下附加支持：

- 支持 Java 转换，但存在一些限制。
- 查找转换可以访问 Hive 查找源。

有关详细信息，请参阅《*Informatica Big Data Management 10.1.1 用户指南*》中的“Hadoop 环境中的映射对象”一章。

Spark 引擎作业运行的运行时统计信息

从版本 10.1.1 开始，可以查看在 Spark 引擎上运行的映射作业的摘要统计信息和详细统计信息。

可以在**摘要统计信息**视图中查看以下 Spark 摘要统计信息：

- 源。映射源文件的名称。
- 目标。目标文件的名称。

- 行数。为源和目标读取的行数。

详细统计信息视图显示 Spark 引擎作业运行的行计数的图形。

有关详细信息，请参阅《*Informatica Big Data Management 10.1.1 用户指南*》中的“Hadoop 环境中的映射对象”一章。

安全

本节介绍版本 10.1.1 中新增的大数据安全功能。

Hive 源的细粒度 SQL 授权支持

从版本 10.1.1 开始，可以对 Hive 连接进行配置，使其在 Hive 源表使用细粒度 SQL 授权时遵守此授权级别。在 Hive 连接中启用**遵守细粒度 SQL 授权**选项可遵守为 Hive 表和视图配置的行级别与列级别限制。

有关详细信息，请参阅《*Informatica 10.1.1 Big Data Management 安全指南*》的“Big Data Management 安全简介”一章中的“授权”一节。

Spark 引擎安全支持

从版本 10.1.1 开始，Spark 引擎支持以下附加的安全系统：

- Cloudera CDH 群集上的 Apache Sentry
- Hortonworks HDP 群集上的 Apache Ranger
- Spark 引擎支持的 Hadoop 分发上的 HDFS 透明加密
- Spark 引擎支持的 Hadoop 分发上的操作系统配置文件

有关详细信息，请参阅《*Informatica Big Data Management 10.1.1 安全指南*》中的“Big Data Management 安全简介”一章。

Sqoop

从版本 10.1.1 开始，可以在配置 Sqoop 时使用以下新功能：

- 可以在 Blaze 引擎上运行 Sqoop 映射。
- 可以在 Spark 引擎上运行 Sqoop 映射，以从 Oracle 数据库读取数据或向其写入数据。
- 在 Blaze 和 Spark 引擎上运行 Sqoop 映射时，可以配置分区。还可以在使用 Kerberos 身份验证的 Hadoop 群集上运行映射。
- 在 Blaze 引擎上运行 Sqoop 映射以从 Teradata 读取数据或向其写入数据时，可以使用以下专用连接器：
 - 由 Teradata 提供技术支持的 Cloudera 连接器
 - Hortonworks Connector for Teradata

这些专用连接器使用本地协议来连接到 Teradata 数据库。

有关详细信息，请参阅《*Informatica 10.1.1 Big Data Management 用户指南*》。

Business Glossary

本节介绍版本 10.1.1 中新增的 Business Glossary 功能。

将富文本导出为纯文本

从版本 10.1.1 开始，可以将富文本词汇表内容导出为纯文本。此导出选项在词汇表导出向导和命令行程序中都可
用。

有关详细信息，请参阅《*Informatica 10.1.1 Business Glossary 指南*》中的“词汇表管理”一章。

包含冲突资产的富文本内容

从版本 10.1.1 开始，当 Analyst 工具检测到冲突资产时，可以选择从导入文件导入格式设置为富文本或数据类型
为长字符串的属性。

此导入选项在词汇表导入向导和命令行程序中都可
用。

有关详细信息，请参阅《*Informatica 10.1.1 Business Glossary 指南*》中的“词汇表管理”一章。

命令行程序

本节介绍版本 10.1.1 中的新命令。

infacmd as 命令

下表介绍了新的 infacmd as 命令：

命令	说明
CreateExceptionAuditTables	为分析服务指定的人工任务实例创建审计表。
DeleteExceptionAuditTables	删除分析服务指定的人工任务实例的审计表。

下表介绍了 infacmd as 命令的新选项：

命令	说明
UpdateServiceOptions	<ul style="list-style-type: none">- HumanTaskDataIntegrationService.exceptionDbName 标识用于存储异常管理任务的审计跟踪表的数据库。- HumanTaskDataIntegrationService.exceptionSchemaName 标识用于存储异常管理任务的审计跟踪表的架构。

有关详细信息，请参阅《*Informatica 10.1.1 命令参考*》中的“Infacmd as 命令参考”一章。

infacmd dis 命令

下表介绍了新的 infacmd dis 命令：

命令	说明
replaceMappingHadoopRuntimeConnections	将已部署的应用程序中所有映射的 Hadoop 连接替换为其他 Hadoop 连接。数据集成服务使用 Hadoop 连接来连接到 Hadoop 群集，以便在 Hadoop 环境中运行映射。

有关详细信息，请参阅《Informatica 10.1.1 命令参考》中的“infacmd dis 命令参考”一章。

infacmd mrs 命令

下表介绍了新的 infacmd mrs 命令：

命令	说明
replaceMappingHadoopRuntimeConnections	将存储库中所有映射的 Hadoop 连接替换为其他 Hadoop 连接。数据集成服务使用 Hadoop 连接来连接到 Hadoop 群集，以便在 Hadoop 环境中运行映射。

有关详细信息，请参阅《Informatica 10.1.1 命令参考》中的“infacmd mrs 命令参考”一章。

pmrep 命令

下表介绍了更新的 pmrep 命令选项：

命令	说明
验证	包含以下更新后的选项： -n (对象名称)。必需。要验证的对象的名称。如果使用 -i 参数，请勿使用此选项。 验证不可重用的会话时，请包含工作流名称。按以下格式输入工作流名称和会话名称： <工作流名称>.<会话实例名称> 在不可重用工作集中验证不可重用会话时，请使用以下格式输入工作流名称、工作集名称和会话名称。 <工作流名称>.<工作集名称>.<会话实例名称>

有关详细信息，请参阅《Informatica 10.1.1 命令参考》中的“pmrep 命令参考”一章。

Enterprise Information Catalog

本节介绍 Enterprise Information Catalog 版本 10.1.1 中的新功能。

Business Glossary 集成

从版本 10.1.1 开始，Analyst 工具业务词汇表与 Enterprise Information Catalog 完全集成。

可以使用业务词汇表资产执行以下任务：

查看目录中的业务词汇表资产。

可以在 Enterprise Information Catalog 中搜索和查看业务术语、类别或策略的完整详细信息。查看业务术语的详细信息时，Enterprise Information Catalog 还会显示与该术语相关的词汇表资产、技术资产和其他资产（例如 Metadata Manager 对象）。

查看目录中的业务词汇表资产时，可以在 Analyst 工具业务词汇表中打开该资产进行进一步分析。

将资产与业务术语相关联。

可以将业务术语与技术资产相关联，以使资产变得更易于理解并且在目录中更易于识别。例如，可以将业务术语“Movie Details”与名为“mv_dt”的关系表相关联。在“资产详细信息”视图以及沿袭和影响图（可选）中，Enterprise Information Catalog 会在搜索结构中的资产名称旁边显示术语“Movie Details”。

将术语与资产相关联时，Enterprise Information Catalog 会根据数据域发现提供智能的关联建议。

有关业务词汇表资产的详细信息，请参阅《*Informatica 10.1.1 Enterprise Information Catalog 用户指南*》中的“查看资产”一章。

列相似性剖析

从版本 10.1.1 开始，可以配置和执行列相似性剖析。列相似性剖析会暗示对从数据源提取的元数据进行准备以在企业数据中发现相似列。然后，可以将数据域附加到相似列，以便更快更高效地在 Enterprise Information Catalog 中搜索相似数据。

Enterprise Information Catalog 对以下资源扫描器支持列相似性剖析：

- Amazon Redshift
- Amazon S3
- Salesforce
- HDFS
- Hive
- IBM DB2
- IBM DB2 for z/OS
- IBM Netezza
- JDBC
- Microsoft SQL Server
- Oracle
- Sybase
- Teradata
- SAP

数据域和数据域组

从版本 10.1.1 开始，可以在 Enterprise Information Catalog 中创建数据域和数据域组。可以在数据域组中对逻辑数据域进行分组。

数据域是基于列数据或列名称的语义的预定义或用户定义模型存储库对象。示例包括社会保障号、电话号码和信用卡号。

可以基于 Informatica Analyst 工具或 Informatica Developer tool 中定义的数据规则或列名称规则创建数据域。或者，也可以基于目录中的现有列创建数据域。可以定义邻近规则，以配置从目录中已配置的现有数据域推理新的数据域的过程。

沿袭和影响分析

从版本 10.1.1 开始，沿袭和影响图的功能已得到扩展。“沿袭和影响”视图还包含一个表格式影响摘要，其中列出了对所研究的资产造成影响的资产以及受所研究的资产影响的资产。

“沿袭和影响”视图具有以下增强功能：

图表增强功能

沿袭和影响图具有以下增强功能：

- 默认情况下，沿袭和影响图会显示源、所研究的资产和数据的目标。使用滑块控件可显示中间资产（按照与种子资产的距离，每拖动滑块一下显示一个中间资产）或完全展开图表。还可以展开特定数据流路径中的所有资产。
- 可以显示所研究的资产的子资产，可以一直向下展开至列级别或字段级别。对资产进行向下钻取时，图表会显示所选子资产以及这些子资产链接到的资产。
- 可以在图表中显示与技术资产关联的业务术语。
- 可以打印图表，也可以将图表导出到可缩放向量图形 (.svg) 文件。

影响分析

打开资产的“沿袭和影响”视图时，可以从图表视图切换到表格式资产摘要。表格式资产摘要列出了对所研究的资产造成影响的资产以及受所研究的资产影响的资产。可以将资产摘要导出到 Microsoft Excel 文件，以创建报告或进一步分析数据。

有关沿袭和影响分析的详细信息，请参阅《Informatica 10.1.1 Enterprise Information Catalog 用户指南》中的“查看沿袭和影响”一章。

用户和用户组的权限

从版本 10.1.1 开始，可以配置用户和用户组对 Enterprise Information Catalog 中已配置的资源权限。可以指定在 Enterprise Information Catalog 中查看资源元数据的权限，也可以指定在 Enterprise Information Catalog 中查看和扩充资源元数据的权限。还可以拒绝特定用户和用户组在 Enterprise Information Catalog 中查看或扩充资源元数据的权限。

新的资源类型

从版本 10.1.1 开始，您可以为以下数据源类型创建资源：

Oracle Business Intelligence

从 Oracle 提供的包含分析和报告功能的业务情报工具提取元数据。

Informatica Master Data Management

从 Informatica Master Data Management 提取有关组织中的重要信息的元数据。

Microsoft SQL Server Integration Service

从 Microsoft SQL Server Integration Service 提取有关数据集成和工作流应用程序的元数据。

SAP

从集成了多个业务应用程序和解决方案的 SAP 应用程序平台提取元数据。

Amazon Elastic MapReduce 上的 Hive

使用 Hive 资源从 Amazon Elastic MapReduce 中的文件提取元数据。

Azure HDInsight 上的 Hive

使用 Hive 资源从 Azure HDInsight 中的文件提取元数据。

同义词定义文件

从版本 10.1.1 开始，可以将同义词定义文件上传到 Enterprise Information Catalog。同义词定义文件包含为目录中的表名称、列名称、数据域和其他资产定义的同义词。可以使用已定义的同义词在 Enterprise Information Catalog 中搜索资产。

通用连接框架

从版本 10.1.1 开始，Enterprise Information Catalog 引入了通用连接框架。使用此框架，您可以构建自定义资源以从 MITI 支持的各种数据源提取元数据。

Informatica Analyst

本节介绍版本 10.1.1 中新增的 Analyst 工具功能。

配置文件

本节介绍针对配置文件和结果卡新增的 Analyst 工具功能。

对结果卡进行向下钻取

从版本 10.1.1 开始，当您在结果卡仪表板中单击一个数据序列或数据点时，映射到该数据序列或数据点的结果卡会显示在资产列表窗格中。

有关结果卡的详细信息，请参阅《*Informatica 10.1.1 Data Discovery 指南*》中的“Informatica Analyst 中的结果卡”一章。

Informatica 安装

本节介绍版本 10.1.1 中新增的安装功能。

Informatica Upgrade Advisor

从版本 10.1.1 开始，可以在执行升级之前运行 Informatica Upgrade Advisor 以检查域中的冲突和已弃用的服务。

有关升级顾问的详细信息，请参阅 *Informatica 升级指南*。

Intelligent Data Lake

本节介绍版本 10.1.1 中新增的 Intelligent Data Lake 功能。

外部源中的表的数据预览

从版本 10.1.1 开始，如果已编录这些源，您可以预览外部（Hadoop 数据湖外部）表的采样数据。管理员需要配置与 Sqoop 的 JDBC 连接，并为分析人员提供必要的权限。分析人员可以使用这些连接来连接到数据源，以查看来自不在数据湖中的资产的数据。

有关详细信息，请参阅《*10.1.1 Intelligent Data Lake 用户指南*》中的“发现数据”一章。

从外部源的表中导入数据

从版本 10.1.1 开始，如果已编录诸如 Oracle 和 Teradata 等外部源（Hadoop 数据湖外部），则可以从这些源的表中将数据导入到数据湖。管理员需要将与 Sqoop 的 JDBC 连接配置到外部源，并向分析人员提供访问权限。分析人员可以使用这些连接来预览数据资产并根据需要将它们导入到数据湖。

有关详细信息，请参阅《*10.1.1 Intelligent Data Lake 用户指南*》中的“发现数据”一章。

将数据导出到外部目标

从版本 10.1.1 开始，您可以将数据资产或出版物导出到诸如 Oracle 和 Teradata 等外部目标（Hadoop 数据湖外部）。管理员需要将与 Sqoop 的 JDBC 连接配置到外部源，并向分析人员提供访问权限。分析人员可以使用这些连接将数据资产导出到外部数据库。

有关详细信息，请参阅《*10.1.1 Intelligent Data Lake 用户指南*》中的“发现数据”一章。

为数据准备配置采样标准

从版本 10.1.1 开始，可以指定最符合您对给定数据资产的数据准备需求的采样标准。您可以选择在准备过程中只包含几列并筛选数据，还可以选择要采样的行数以及选择“随机”或“前 N 行”作为样本。

有关详细信息，请参阅《*10.1.1 Intelligent Data Lake 用户指南*》中的“准备数据”一章。

对工作表执行查找

从版本 10.1.1 开始，您可以执行查找。使用查找函数可在另一个工作表中查找键列，并在查找的该工作表中相应的其他列中提取值。

有关详细信息，请参阅《*10.1.1 Intelligent Data Lake 用户指南*》中的“准备数据”一章。

下载为 TDE 文件

从版本 10.1.1 开始，您可以将数据湖资产中的数据下载为 TDE 文件。您可以直接在 Tableau 中打开已下载的文件。您可以搜索任何数据资产并将其下载为 CSV 文件或 TDE 文件。

有关详细信息，请参阅《*10.1.1 Intelligent Data Lake 用户指南*》中的“发现数据”一章。

Sentry 和 Ranger 支持

从版本 10.1.1 开始，Intelligent Data Lake 在 Cloudera 和 Hortonworks 上支持 Sentry 和 Ranger。Ranger 和 Sentry 可提供一个集中式安全框架来管理 Cloudera 和 Hortonworks 上的粒度级别访问控制。您可以创建授权规则或策略来控制数据访问。Sentry 和 Ranger 对数据湖资产支持基于 SQL 的授权。

映射

本节介绍版本 10.1.1 中新增的映射功能。

Informatica 映射

本节介绍版本 10.1.1 中新增的 Informatica 映射功能。

将参数导出到参数文件

可以从 Developer tool 导出映射参数文件或工作流参数文件。可以导出包含您在 Developer tool 中定义的映射参数或工作流参数的参数文件。Developer tool 会以 .xml 格式创建参数文件。可以从映射的 **参数** 选项卡或工作流的 **参数** 选项卡导出参数。可以在运行已部署的映射或工作流时使用参数文件。

有关详细信息，请参阅《*Informatica Developer 10.1.1 映射指南*》中的“映射参数”一章或《*Informatica Developer 10.1.1 工作流指南*》中的“工作流参数”一章。

Metadata Manager

本节介绍版本 10.1.1 中新增的 Metadata Manager 功能。

Cloudera Navigator 资源的数据集提取

从版本 10.1.1 开始，Metadata Manager 可以从 Cloudera Navigator 提取 HDFS 数据集。Metadata Manager 会在 HDFS 数据集逻辑组内的元数据目录中显示数据集。

有关 Cloudera Navigator 资源的详细信息，请参阅《*Informatica 10.1.1 Metadata Manager 管理员指南*》中的“数据库管理资源”一章。

Informatica Platform 资源的映射提取

从版本 10.1.1 开始，Informatica Platform 资源可以提取已部署的工作流中的映射的元数据。

除了提取直接部署到应用程序的映射之外，基于版本 10.1.1 应用程序的 Informatica Platform 资源还可以提取已部署的工作流中的映射的元数据。

在提取已部署的工作流中的映射时，Metadata Manager 会将工作流名称和映射任务名称作为前缀添加到映射名称。Metadata Manager 会在映射逻辑组内的元数据目录中显示映射。

有关 Informatica Platform 资源的详细信息，请参阅《*Informatica 10.1.1 Metadata Manager 管理员指南*》中的“数据集成资源”一章。

PowerExchange 适配器

本节介绍版本 10.1.1 中新增的 PowerExchange 适配器功能。

Informatica PowerExchange® 适配器

本节介绍版本 10.1.1 中新增的 Informatica 适配器功能。

PowerExchange for Amazon Redshift

从版本 10.1.1 开始，可以启用 PowerExchange for Amazon Redshift 以在 Blaze 引擎上运行映射。运行映射时，数据集成服务会将映射推送到 Hadoop 群集并在 Blaze 引擎上处理映射，这将大幅提升性能。

有关详细信息，请参阅《*Informatica PowerExchange for Amazon Redshift 10.1.1 用户指南*》。

PowerExchange for Cassandra

从版本 10.1.1 开始，PowerExchange for Cassandra 支持以下功能：

- 可以将以下高级 ODBC 驱动程序配置与 PowerExchange for Cassandra 结合使用：
 - 负载平衡策略。决定如何根据指定的 DC 感知策略或循环策略在 Cassandra 群集中的节点之间分配查询。
 - 筛选。限制驱动程序到预定义的主机集的连接。
- 可以在 ODBC 驱动程序中启用以下参数以优化性能：
 - 标志感知缩短查询延迟并降低 Cassandra 节点的负载。
 - 延迟感知。在发送查询时忽略性能不佳的 Cassandra 节点。
 - 空值插入。允许在 INSERT 语句中指定空值。
 - 区分大小写。允许以区分大小写的方式指定架构名称、表名称和列名称。
- 可以处理包含 date、smallint 以及 tinyint 数据类型的 Cassandra 源和目标。

有关详细信息，请参阅《*Informatica PowerExchange for Cassandra 10.1.1 用户指南*》。

PowerExchange for HBase

从版本 10.1.1 开始，可以启用 PowerExchange for HBase 以在 Blaze 或 Spark 引擎上运行映射。运行映射时，数据集成服务会将映射推送到 Hadoop 群集并在选定的引擎上处理映射，这将大幅提升性能。

有关详细信息，请参阅《*Informatica PowerExchange for HBase 10.1.1 用户指南*》。

PowerExchange for Hive

从版本 10.1.1 开始，可以在本地环境中在映射中的 Hive 数据对象上配置查找转换。

有关详细信息，请参阅《*Informatica PowerExchange for Hive 10.1.1 用户指南*》。

PowerCenter® PowerExchange 适配器

本节介绍版本 10.1.1 中新增的 PowerCenter 适配器功能。

PowerExchange for Amazon Redshift

从版本 10.1.1 开始，使用 PowerExchange for Amazon Redshift 可以执行以下任务：

- 可为 Amazon Redshift 源和目标配置分区。可以配置分区信息，使 PowerCenter 集成服务能够在运行时确定要创建的分区数量。
- 可以在映射中包括管道查找转换。

- 当连接类型为 ODBC 并且选择 Redshift 作为 ODBC 子类型时，PowerCenter 集成服务可以将表达式函数、汇总器函数、运算符函数、联合函数、排序器函数和筛选器函数推送到 Amazon Redshift 源和目标。
- 可以在映射中配置高级筛选器属性。
- 可为映射中的源对象和目标对象配置 pre-SQL 和 post-SQL 查询。
- 可以对源转换进行配置，使其选择 Amazon Redshift 表中的不同行并对数据进行排序。
- 可以将源表名称和目标表名称参数化，以替代映射中的表名称。
- 可为映射中的源对象和目标对象定义 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。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica 10.1.1 PowerExchange for Amazon Redshift 用户指南》。

PowerExchange for Cassandra

从版本 10.1.1 开始，PowerExchange for Cassandra 支持以下功能：

- 可以将以下高级 ODBC 驱动程序配置与 PowerExchange for Cassandra 结合使用：
 - 负载均衡策略。决定如何根据指定的 DC 感知策略或循环策略在 Cassandra 群集中的节点之间分配查询。
 - 筛选。限制驱动程序到预定义的主机集的连接。
- 可以在 ODBC 驱动程序中启用以下参数以优化性能：
 - 标志感知缩短查询延迟并降低 Cassandra 节点的负载。
 - 延迟感知。在发送查询时忽略性能不佳的 Cassandra 节点。
 - 空值插入。允许在 INSERT 语句中指定空值。
 - 区分大小写。允许以区分大小写的方式指定架构名称、表名称和列名称。
- 可以处理包含 date、smallint 以及 tinyint 数据类型的 Cassandra 源和目标。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Cassandra 10.1.1 用户指南》。

PowerExchange for Vertica

从版本 10.1.1 开始，PowerExchange for Vertica 支持将数据压缩为 GZIP 格式。使用批量模式向 Vertica 目标写入大量数据时，可以将会话配置为创建暂存文件。在 UNIX 操作系统上，如果启用文件暂存，还可以将数据压缩为 GZIP 格式。压缩数据后，可以减少在网络中传输的数据量并提高会话性能。

要压缩数据，必须向 PowerCenter 存储库重新注册 PowerExchange for Vertica 插件。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica PowerExchange for Vertica 10.1.1 用户指南》。

安全

本节介绍版本 10.1.1 中的新安全功能。

自定义 Kerberos 库

从版本 10.1.1 开始，可以对 Informatica 域中的自定义或本地数据库客户端进程和 Informatica 进程进行配置，使这些进程使用自定义 Kerberos 库而非 Informatica 所使用的默认 Kerberos 库。

有关详细信息，请参阅《Informatica 10.1.1 安全指南》中的“Kerberos 身份验证设置”一章。

已启用 Kerberos 的域中的计划程序服务支持

从版本 10.1.1 开始，可以使用计划程序服务在使用 Kerberos 身份验证的域中运行映射、工作流、配置文件和结果卡。

Informatica Web 应用程序的单点登录

从版本 10.1.1 开始，可以配置基于安全断言标记语言 (Security Assertion Markup Language, SAML) 的单点登录 (Single Sign-on, SSO)，以登录到 Administrator 工具、Analyst 工具和 Monitoring 工具。

安全断言标记语言是一种基于 XML 的数据格式，用于在服务提供程序和标识提供程序之间交换身份验证和授权信息。在 Informatica 域中，Informatica Web 应用程序是服务提供程序。Microsoft Active Directory 联合身份验证服务 (AD FS) 2.0 是标识提供程序，它将使用组织的 LDAP 或 Active Directory 标识存储对 Web 应用程序用户进行身份验证。

有关详细信息，请参阅《Informatica 10.1.1 安全指南》中的“Informatica Web 应用程序的单点登录”一章。

转换

本节介绍版本 10.1.1 中新增的转换功能。

Informatica 转换

本节介绍版本 10.1.1 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 10.1.1 开始，可以向输出地址添加“计数”端口。“计数”端口值指示转换在交互模式或建议列表模式下返回的一组建议中每个地址的位置。

例如，“计数”端口为该集的第一个地址返回数字 1，为该集的第二个地址返回数字 2。对于地址验证返回的每个地址，该数字以 1 递增。

可在“状态信息”端口组中查找“计数”端口。

中国

多语言地址解析和验证

从版本 10.1.1 开始，可以对地址验证器转换进行配置，使其采用转译的拉丁文字（拼音）或英语返回有效的中国地址中的街道描述符和街道方向信息。转换将采用汉字文字返回地址中的其他元素。

要指定输出语言，请在转换上设置“首选语言”高级属性。

建议列表模式下对中国地址的单行验证

从版本 10.1.1 开始，可以对地址验证器转换进行配置，使其为您在快速填写模式下在单行中输入的中国地址返回有效的建议。要在单行中输入地址，请从“多行”端口组中选择“完整地址”端口。请采用汉字文字输入地址。

输入不完整地址时，转换将为您输入的地址返回一个或多个地址建议。输入完整的有效地址时，转换将从引用数据库返回该地址的有效版本。

爱尔兰

多语言地址解析和验证

从版本 10.1.1 开始，可以对地址验证器转换进行配置，使其采用爱尔兰语读取和写入爱尔兰地址的街道、区域和郡信息。

除了维护英语地址之外，爱尔兰邮政服务机构 An Post 还会维护地址的爱尔兰语信息。可以在输入地址中包含爱尔兰语街道、区域和郡信息，还可以检索地址的有效英语版本。可以输入英语地址，也可以检索包含爱尔兰语街道、区域和郡信息的地址。地址验证将采用英语返回所有其他信息。

要指定输出语言，请在转换上设置“首选语言”高级属性。

爱尔兰地址中的屋顶级地理坐标

从版本 10.1.1 开始，可以对地址验证器转换进行配置，使其返回爱尔兰地址的屋顶级地理坐标。

要返回地理坐标，请向输出地址添加“地理编码填写”端口。可在“地理编码”端口组中查找“地理编码填写”端口。要指定屋顶级地理坐标，请在转换上设置“地理编码数据类型”高级属性。

爱尔兰地址中的首选描述符支持

从版本 10.1.1 开始，可以对地址验证器转换进行配置，使其采用英语返回以下元素的短格式或长格式：

- 街道描述符
- 方向值

要指定元素的首选项，请在转换上设置“全局首选描述符”高级属性。

注意：地址验证器转换会将所有街道信息写入到爱尔兰语地址中的街道名称字段。

意大利

从版本 10.1.1 开始，可以对地址验证器转换进行配置，使其向有效的意大利地址添加 ISTAT 代码。ISTAT 代码包含一些字符，用于标识地址所属的省、自治市和地区。意大利国家统计局 (Italian National Institute of Statistics, ISTAT) 负责维护 ISTAT 代码。

要向地址添加 ISTAT 代码，请选择“ISTAT 代码”端口。可在“意大利补充”端口组中查找“ISTAT 代码”端口。

日本

日本地址的地理编码扩充

从版本 10.1.1 开始，可以对地址验证器转换进行配置，使其返回日本地址的标准地理坐标。

转换可以返回多个准确度级别的地理坐标。当有效地址包含适用于 Ban 级别的信息时，转换将返回门牌号级别地理坐标。当有效地址包含适用于 Chome 级别的信息时，转换将返回街道级别地理坐标。如果地址不包含 Ban 或 Chome 信息，地址验证将返回区域级别地理坐标。

要返回地理坐标，请向输出地址添加“地理编码填写”端口。可在“地理编码”端口组中查找“地理编码填写”端口。

建议列表模式下对日本地址的单行验证

从版本 10.1.1 开始，可以对地址验证器转换进行配置，使其为您在建议列表模式下在单行中输入的日本地址返回有效的建议。可以检索采用日文汉字文字或日语假名文字输入的地址的建议。要在单行中输入地址，请从“多行”端口组中选择“完整地址”端口。

输入不完整地址时，转换将为您输入的地址返回一个或多个地址建议。输入完整的有效地址时，转换将从引用数据库返回该地址的有效版本。

韩国

韩国地址中的 Revised Romanization 转译支持

从版本 10.1.1 开始，地址验证器转换可以使用 Revised Romanization 系统在朝鲜文字符集和拉丁文字符集之间转译地址。要指定韩国输出地址的字符集，请使用“首选脚本”高级属性。

对韩国地址中的邮政编码验证进行的更新

从版本 10.1.1 开始，地址验证器转换将向不包含邮政编码的完全有效的输入地址添加一个五位数的邮政编码。该五位数的邮政编码代表韩国正在使用的当前邮政编码格式。转换可以向完全有效的基于场地的地址和完全有效的基于街道的地址添加五位数的邮政编码。

要验证采用较旧的基于场地格式的地址，请使用“匹配扩展存档”高级属性。

西班牙

从版本 10.1.1 开始，可以对地址验证器转换进行配置，使其向有效的西班牙地址添加 INE 代码。INE 代码包含一些字符，用于在地址中标识省、市镇和街道。西班牙国家统计局 (National Institute of Statistics, INE) 负责维护 INE 代码。

要向地址添加 INE 代码，请选择以下一个或多个端口：

- INE 市代码
- INE 省/市/自治区代码
- INE 街道代码

可在“西班牙补充”端口组中查找“INE 代码”端口。

美国

对 CASS Cycle O 要求的支持

从版本 10.1.1 开始，地址验证器转换增加了支持编码准确性支持系统 (Coding Accuracy Support System, CASS) Cycle O 标准的建议要求的功能。

针对 Cycle O 标准做好准备，转换包含以下功能：

- 私人邮箱和商业邮件接收机构标识

美国邮政署在 Cycle O 中更新了对私人邮箱 (Private Mailbox, PMB) 地址和商业邮件接收机构 (Commercial Mail Receiving Agency, CMRA) 地址的 CASS 要求。为了遵守 Cycle O 标准，地址验证器转换将在 CMRA 地址中的私人邮箱号前面添加 PMB 作为前缀。如果地址中的私人邮箱号前面带有英镑符号 (#)，则转换会将英镑符号转换为 PMB。为了遵守 Cycle O 标准，转换不会使用 PMB 号码来验证地址的收件人地点验证 (Delivery Point Validation, DPV) 数据。

- 用于邮政信箱街道地址 (Post Office Box Street Address, PBSA) 标识的“DPV PBSA 指示器”端口

美国邮政署能够识别采用街道地址格式的邮政信箱街道地址。要在地址集中标识 PBSA 地址，请使用“DPV PBSA 指示器”端口。可在“特定于美国”端口组中查找“DPV PBSA 指示器”端口。

例如，以下地址标识 South Center 街道邮局的邮政信箱号 3094：

131 S Center St Unit 3094
Collierville TN 38027-0419

- 用于 Form 3553 填写的“DPV 邮政编码验证”端口

“DPV 邮政编码验证”端口指示地址是否可用于包括到 CASS Form 3553 上的地址总数。如果地址通过收件人地点验证但不包含可投递的 ZIP+4 代码，则不能将该地址包括到地址总数。可在“特定于美国”端口组中查找“DPV 邮政编码验证”端口。

改进了对美国地址中的非标准第一行数据的解析

从版本 10.1.1 开始，地址验证器转换可将非标准邮箱数据解析为子建筑物元素。非标准数据可能标识大学校园邮箱或法院法庭。

美国地址中的全局首选描述符支持

从版本 10.1.1 开始，可以返回美国地址中以下元素的短格式或长格式：

- 街道描述符
- 方向值
- 子建筑物描述符

要指定转换返回的元素格式，请在转换上设置“全局首选描述符”高级属性。

有关详细信息，请参阅《*Informatica 10.1.1 Developer 转换指南*》和《*Informatica 10.1.1 地址验证器端口参考*》。

写入转换

从版本 10.1.1 开始，从映射中的现有转换创建写入转换时，可以指定该写入转换的输入端口的链接类型。

可以按名称将端口链接起来。此外，在动态映射中，可以按名称将端口链接起来，还可以基于映射流创建动态端口，或者也可以在运行时基于链接策略将端口链接起来。

有关详细信息，请参阅《*Informatica 10.1.1 Developer 转换指南*》中的“写入转换”一章。

Web 服务

本节介绍版本 10.1.1 中新增的 Web 服务功能。

Informatica Web 服务

本节介绍版本 10.1.1 中新增的 Informatica Web 服务功能。

REST Web 服务

可以创建使用 JSON 或 XML 格式将数据返回到 Web 服务客户端的 Informatica REST Web 服务。

Informatica REST Web 服务是一种 Web 服务，可接收 HTTP 请求以执行 GET 操作。GET 操作用于检索数据。REST 请求是来自 Internet 浏览器的简单 URI 字符串。客户端可以通过向 URI 添加筛选器参数来限制 Web 服务输出数据。

可以在 Developer tool 中定义 REST Web 服务资源。REST Web 服务资源包含 REST Web 服务响应消息的定义和返回响应的映射。创建 Informatica REST Web 服务时，可以从数据对象定义资源，或者也可以手动定义资源。

工作流

本节介绍版本 10.1.1 中新增的工作流功能。

Informatica 工作流

本节介绍版本 10.1.1 中新增的 Informatica 工作流功能。

终止事件

从版本 10.1.1 开始，可以向工作流添加终止事件。终止事件定义了一个在结束事件之前的点，此时可以结束事件。一个工作流可以包含一个或多个终止事件。

将任务或网关连接到终止事件时，如果任务输出满足序列流上的条件，工作流会终止。在工作流中的任何其他任务可以运行之前，终止事件会中止工作流。

如果工作流数据能够到达某个点（此时不需要运行其他任务），则可以向工作流添加终止事件。例如，可以添加终止事件以结束包含一个映射任务和一个人工任务的工作流。请将映射任务连接到独占网关，然后将网关连接到一个人工任务和一个终止事件。如果映射任务为人工任务生成异常记录数据，工作流将遵循到人工任务的序列流。如果映射任务未生成异常记录数据，工作流将遵循到终止事件的序列流。

有关详细信息，请参阅《*Informatica 10.1.1 Developer 工作流指南*》。

对人工任务的用户权限

从版本 10.1.1 开始，可以设置对人工任务数据的用户权限。权限指定了用户可以查看的数据，以及用户可以在 Analyst 工具中对人工任务实例执行的操作类型。在设计工作流时，可以在人工任务的步骤中设置权限。权限会应用到可以查看或编辑该步骤所定义的任务实例的所有用户。

默认情况下，Analyst 工具用户可以查看所有数据以及对他们所处理的任务实例执行任何操作。

可以设置查看权限和编辑权限。查看权限定义了 Analyst 工具为步骤所定义的任务实例显示哪些数据。编辑权限定义了用户可以执行哪些操作以更新任务实例数据。查看权限优先于编辑权限。如果已授予对某个列的编辑权限但未授予对该列的查看权限，则 Analyst 工具用户无法编辑列数据。

有关详细信息，请参阅《*Informatica 10.1.1 Developer 工作流指南*》。

人工任务实例通知中的工作流变量

从版本 10.1.1 开始，可以使用工作流变量将有关人工任务实例的信息写入到电子邮件通知。当用户完成、上报或重新分配任务实例时，这些变量会记录有关任务实例的信息。

要显示变量的列表，请打开该人工任务，然后选择定义了人工任务实例的步骤。在**通知**视图中，选择电子邮件通知的邮件正文，然后按 `$+Ctrl+空格键`。

通知可以显示以下变量：

`$taskEvent.eventTime`

工作流引擎指示用户上报、重新分配或完成任务实例时的时间。

`$taskEvent.startOwner`

在工作流引擎上报或完成任务时的任务实例所有者。或者在引擎重新分配任务实例之后的任务实例所有者。

`$taskEvent.status`

在引擎指示用户上报、重新分配或完成任务实例之后的任务实例状态。状态名称为 `READY` 和 `IN_PROGRESS`。

`$taskEvent.taskEventType`

引擎发出的指示的类型。变量值为 `escalate`、`reassign` 和 `complete`。

`$taskEvent.taskId`

Analyst 工具显示的任务实例标识符。

有关详细信息，请参阅《*Informatica 10.1.1 Developer 工作流指南*》。

第 33 章

更改 (10.1.1)

本章包括以下主题：

- [支持更改, 335](#)
- [大数据, 337](#)
- [Business Glossary , 338](#)
- [数据集成服务, 339](#)
- [数据类型, 339](#)
- [Informatica Analyst, 340](#)
- [Informatica Developer, 340](#)
- [映射, 340](#)
- [Enterprise Information Catalog, 341](#)
- [Metadata Manager, 341](#)
- [PowerExchange 适配器, 342](#)
- [转换, 343](#)
- [工作流, 344](#)
- [文档, 344](#)

支持更改

本节介绍版本 10.1.1 HotFix 2 中的支持更改。

Big Data Management Hive 引擎

从版本 10.1.1 开始，Informatica 已终止为 Hive 引擎用于运行映射的 HiveServer2 提供支持。

在以前的版本中，Hive 引擎支持使用 Hive 驱动程序和 HiveServer2 在 Hadoop 环境中运行映射。HiveServer2 和 Hive 驱动程序可将 HiveQL 查询转换为在 Hadoop 群集上处理的 MapReduce 或 Tez 作业。

如果您安装了 Big Data Management 10.1.1 或升级到版本 10.1.1，则在运行映射时，Hive 引擎会使用 Hive 驱动程序。Hive 引擎不再支持使用 HiveServer2 在 Hadoop 环境中运行映射。Hadoop 群集上仍然支持使用 HiveServer2 服务的 Hive 源和目标。

要在 Hadoop 环境中运行映射，Informatica 建议您选择所有运行时引擎。数据集成服务会使用专有的基于规则的方法来确定最适用于运行映射的引擎。

有关为 Hadoop 发行版配置运行时引擎的信息，请参阅《*Informatica Big Data Management 10.1.1 安装和配置指南*》。有关运行时引擎支持的映射对象的信息，请参阅《*Informatica Big Data Management 10.1.1 用户指南*》。

支持更改 - Big Data Management Hadoop 发行版

下表列出了 Big Data Management 10.1.1 中支持的 Hadoop 发行版本和更改：

在发行时，版本 10.1.1 支持以下 Hadoop 发行版：

- Azure HDInsight 版本 3.4
- Cloudera CDH 版本 5.8
- IBM BigInsights 版本 4.2
- Hortonworks HDP 版本 2.5
- Amazon EMR 版本 5.0

要查看支持的版本的最新列表，请参阅 Informatica 客户门户上的“产品可用性列表”，网址为：<https://network.informatica.com/community/informatica-network/product-availability-matrices>。

MapR 支持

从版本 10.1.1 开始，Informatica 推迟了对 MapR 群集上的 Big Data Management 的支持。要在 MapR 群集上运行映射，请使用 Big Data Management 10.1。Informatica 计划在将来的版本中恢复此支持。

本文档以示例的形式保留了某些对 MapR 的引用。请将这些示例的结构应用到 Hadoop 分发。

Amazon EMR 支持

从版本 10.1.1 开始，可以在 Amazon EMR 环境中安装 Big Data Management。可以从以下安装方法中选择一种方法：

- 下载 RPM 包并从中进行安装。在 Amazon EMR 环境中安装 Big Data Management 时，应在用于运行模型存储库服务、数据集成服务和其他服务的本地计算机上安装 Big Data Management 元素。
- 在 Amazon 云环境中安装 Informatica 实例。在 Amazon 云中创建 Big Data Management 的实施时，应使要在其中安装并运行 Big Data Management 的虚拟机联机。

有关在 Amazon EMR 上安装和配置 Big Data Management 的详细信息，请参阅《*Informatica Big Data Management 10.1.1 安装和配置指南*》。

Big Data Management Spark 支持

从版本 10.1.1 开始，可以在所有受支持的 Hadoop 分发上配置 Spark 引擎。可以将 Big Data Management 配置为使用以下 Spark 版本之一，具体取决于所使用的 Hadoop 分发：

- 适用于 Cloudera cdh5u8 分发的 Cloudera Spark 1.6 和 Apache Spark 2.0.1。
- 适用于所有 Hadoop 分发的 Apache Spark 2.0.1。

有关详细信息，请参阅《*Informatica Big Data Management 10.1.1 安装和配置指南*》。

Data Analyzer

从版本 10.1.1 开始，Informatica 停止了对 Data Analyzer 的支持。Informatica 建议使用第三方报告工具来运行 PowerCenter 报告和 Metadata Manager 报告。您可以使用建议的 SQL 查询构建 PowerCenter 早期版本随附的所有报告。

操作系统

从版本 10.1.1 开始，Informatica 增加了对以下操作系统的支持：

Solaris 11

适用于 Informatica 客户端的 Windows 10

PowerExchange for SAP NetWeaver

从版本 10.1.1 开始，Informatica 对 PowerExchange for SAP NetWeaver 支持实施了以下更改：

支持更改	支持级别	注释
分析业务组件	删除的支持	从版本 10.1.1 开始，Informatica 停止了对分析服务组件 (Analytic Business Components, ABC) 功能的支持。无法使用 ABC 存储库中的对象来读取和转换 SAP 数据。Informatica 不会提供 ABC 传输文件。
SAP R/3 版本 4.7	删除的支持	从版本 10.1.1 开始，Informatica 停止了对 SAP R/3 4.7 系统的支持。请升级到 SAP ECC 版本 5.0 或更高版本。

报告和仪表板服务

从版本 10.1.1 开始，Informatica 停止了对报告和仪表板服务的支持。Informatica 建议使用第三方报告工具来运行 PowerCenter 报告和 Metadata Manager 报告。您可以使用建议的 SQL 查询构建 PowerCenter 早期版本随附的所有报告。

报告服务

从版本 10.1.1 开始，Informatica 停止了对报告服务的支持。Informatica 建议使用第三方报告工具来运行 PowerCenter 报告和 Metadata Manager 报告。您可以使用建议的 SQL 查询构建 PowerCenter 早期版本随附的所有报告。

大数据

本节介绍版本 10.1.1 中对 Big Data 的更改。

Hadoop 环境中支持的函数

从版本 10.1.1 开始，以下支持更改会影响 Hadoop 环境中的函数：

功能	说明	更改
AES_DECRYPT	将加密数据返回字符串格式。	在 Spark 引擎上受支持。 以前只在 Blaze 和 Hive 引擎上受支持。
AES_ENCRYPT	返回加密格式的数据。	在 Spark 引擎上受支持。 以前只在 Blaze 和 Hive 引擎上受支持。

功能	说明	更改
COMPRESS	使用 zlib 1.2.1 压缩算法压缩数据。	在 Spark 引擎上受支持。 以前只在 Blaze 和 Hive 引擎上受支持。
CRC32	返回 32 位循环冗余检查 (CRC32) 值。	在 Spark 引擎上受支持。 以前只在 Blaze 和 Hive 引擎上受支持。
DECOMPRESS	使用 zlib 1.2.1 压缩算法解压缩数据。	在 Spark 引擎上受支持，但存在一定限制。 以前只在 Blaze 和 Hive 引擎上受支持。
DEC_BASE64	解码 base 64 编码值，并返回使用数据的二进制数据表现形式的字符串。	在 Spark 引擎上受支持。 以前只在 Blaze 和 Hive 引擎上受支持。
ENC_BASE64	通过使用多用途 Internet 邮件扩展(MIME) 编码方式将二进制数据转换为字符串数据来编码数据。	在 Spark 引擎上受支持。 以前只在 Blaze 和 Hive 引擎上受支持。
MD5	计算输入值的校验和。函数使用消息摘要算法 5(MD5)。	在 Spark 引擎上受支持。 以前只在 Blaze 和 Hive 引擎上受支持。
UUID4	返回符合 RFC 4122 中所述 UUID 规范的变量 4 的随机生成 16 字节二进制值。	在 Spark 引擎上受支持，且不受任何限制。 以前在 Blaze 引擎上受支持且不受任何限制，而且在 Spark 和 Hive 引擎上也受支持，但存在一定限制。
UUID_UNPARSE	将 16 字节二进制值转换为 RFC 4122 中指定的 36 字符字符串表示形式。	在 Spark 引擎上受支持，且不受任何限制。 以前在 Blaze 引擎上受支持且不受任何限制，而且在 Spark 和 Hive 引擎上也受支持，但存在一定限制。

Hadoop 配置管理器

从版本 10.1.1 开始，Big Data Management 配置实用程序进行了以下更改：

- 该实用程序已重命名为 Hadoop 配置管理器。
- Hadoop 配置管理器支持在 Azure HDInsight 群集和其他 Hadoop 群集上配置 Big Data Management。

有关 Hadoop 配置管理器的详细信息，请参阅《*Informatica Big Data Management 10.1.1 安装和配置指南*》。

Business Glossary

本节介绍版本 10.1.1 中对 Business Glossary 的更改。

导出文件限制

从版本 10.1.1 开始，Analyst 工具和命令行中的 Business Glossary 导出具有以下行为更改：

Microsoft Excel 导出文件单元格中的字符被截断

在一个 Microsoft Excel 单元格中导出包含超过 32,767 个字符的词汇表资产时，Analyst 工具会自动将单元格中的字符截断至低于 32,763 的值。

Microsoft Excel 仅支持每个单元格最多 32,767 个字符。在以前的版本中，当导出词汇表时，Microsoft Excel 会截断一个单元格中包含超过 32,767 个字符的长文本属性，从而导致数据丢失而不显示任何警告。

有关导入和导出的详细信息，请参阅《*Informatica 10.1.1 Business Glossary 指南*》中的“词汇表管理”一章。

数据集成服务

本节介绍版本 10.1.1 中对数据集成服务的更改。

数据集成属性中的执行选项

从版本 10.1.1 开始，在编辑以下数据集成服务属性时不再需要重新启动数据集成服务：

- 缓存目录
- 主目录
- 最大并行数
- 拒绝文件目录
- 源目录
- 状态存储
- 目标目录
- 临时目录

在以前的版本中，在编辑了以上属性时必须重新启动数据集成服务。

数据类型

本节介绍版本 10.1.1 中对数据类型的更改。

Informatica 数据类型

本节介绍对 Developer tool 中转换数据类型的更改。

双精度数据类型

从版本 10.1.1 开始，可以编辑双精度数据类型的精度和小数位数。小数位数必须小于或等于精度。

以前，精度设置为 15 位，小数位数设置为 0 位。

有关详细信息，请参阅《*Informatica 10.1.1 Developer tool 指南*》中的“数据类型参考”附录。

Informatica Analyst

本节介绍版本 10.1.1 中对 Analyst 工具的更改。

配置文件

本节介绍针对配置文件新增的 Analyst 工具功能。

运行时环境

从版本 10.1.1 开始，在选择 Hive 选项作为运行时环境之后，应选择 Hadoop 连接用于运行配置文件。

在以前的版本中，在选择 Hive 选项作为运行时环境之后，应选择 Hive 连接用于运行配置文件。

有关运行时环境的详细信息，请参阅《*Informatica 10.1.1 Data Discovery 指南*》中的“Informatica Analyst 中的列配置文件”一章。

Informatica Developer

本节介绍版本 10.1.1 中对 Developer tool 的更改。

配置文件

本节介绍针对配置文件新增的 Developer tool 功能。

运行时环境

从版本 10.1.1 开始，在选择 Hive 选项作为运行时环境之后，应选择 Hadoop 连接用于运行配置文件。

在以前的版本中，在选择 Hive 选项作为运行时环境之后，应选择 Hive 连接用于运行配置文件。

有关运行时环境的详细信息，请参阅《*Informatica 10.1.1 Data Discovery 指南*》中的“数据对象配置文件”一章。

映射

本节介绍版本 10.1.1 中对映射的更改。

Informatica 映射

本节介绍版本 10.1.1 中对 Informatica 映射的更改。

对动态端口中生成的端口进行重新排序

从版本 10.1.1 开始，可以根据以下选项更改生成的端口的顺序：

- 上游转换的组或动态端口中的端口顺序。
- 动态端口的输入规则的顺序。
- 具有静态端口的最近转换中的端口顺序。

默认设置是根据上游转换中的端口进行重新排序。

在以前的版本中，可以根据动态端口的输入规则的顺序对生成的端口进行重新排序。

有关详细信息，请参阅《*Informatica 10.1.1 Developer 映射指南*》中的“动态映射”一章。

Enterprise Information Catalog

本节介绍版本 10.1.1 中对 Enterprise Information Catalog 的更改。

HDFS 扫描器增强

从版本 10.1.1 开始，可以使用 HDFS 资源扫描器从平面文件类型提取元数据。

关系视图

从版本 10.1.1 开始，可以查看选定的资产的业务术语、相关词汇表资产、相关技术资产和相似列。

在以前的版本中，可以查看资产关系，例如列、数据域、表和视图。

有关关系视图的详细信息，请参阅《*Informatica 10.1.1 Enterprise Information Catalog 用户指南*》中的“查看关系”一章。

Metadata Manager

本节介绍版本 10.1.1 中对 Metadata Manager 的更改。

Cloudera 导航器资源

从版本 10.1.1 开始，Cloudera Navigator 资源具有以下行为更改：

增量加载更改

Cloudera Navigator 资源的增量加载在默认情况下处于禁用状态。在以前的版本中，增量加载在默认情况下处于启用状态。

如果启用了增量加载，则当 Cloudera 管理员在上次成功的元数据加载之后在 Cloudera Navigator 中调用清除操作时，Metadata Manager 将执行完整的元数据加载。

此外，还有一些新的准则，这些准则说明了在哪些情况下需要禁用增量加载。

搜索查询更改

可以使用搜索查询从元数据加载中排除实体类型和 HDFS 实体。例如，可以使用搜索查询排除 YARN 或 Oozie 作业执行。

数据沿袭更改

为了降低数据沿袭图的复杂性，Metadata Manager 进行了以下更改：

- Metadata Manager 不再显示 Hive 查询模板部件的数据沿袭。可以改为对 Hive 查询模板运行数据沿袭分析。

- 对于已分区 Hive 表，Metadata Manager 将显示表中每个列和包含相关 HDFS 实体的父目录之间的数据沿袭链接。在以前的版本中，Metadata Manager 显示每个列和每个相关 HDFS 实体之间的数据沿袭链接。

有关 Cloudera Navigator 资源的详细信息，请参阅《*Informatica 10.1.1 Metadata Manager 管理员指南*》中的“数据库管理资源”一章。

Netezza 资源

从版本 10.1.1 开始，Metadata Manager 支持对 Netezza 资源使用多个架构。

Netezza 资源具有以下行为更改：

- 创建或编辑 Netezza 资源时，可以选择要从中提取元数据的架构。可以选择一个或多个架构。
- Metadata Manager 会按架构在元数据目录中组织 Netezza 对象。数据库不会显示在元数据目录中。
- 配置与 Netezza 的连接分配时，可以选择要将连接分配给的架构。

由于这些更改，Netezza 资源如同其他类型的关系资源一样工作。

在以前的版本中，在创建或编辑 Netezza 资源时，无法选择要从中提取元数据的架构。如果从包含多个架构的 Netezza 数据库创建了资源，Metadata Manager 会忽略架构信息。Metadata Manager 会按数据库在元数据目录中组织 Netezza 对象。配置与 Netezza 的连接分配时，可以选择要将连接分配给的数据库。

有关 Netezza 资源的详细信息，请参阅《*Informatica 10.1.1 Metadata Manager 管理员指南*》中的“数据库管理资源”一章。

PowerExchange 适配器

本节介绍版本 10.1.1 中对 PowerExchange 适配器的更改。

Informatica PowerExchange 适配器

本节介绍版本 10.1.1 中对 Informatica 适配器的更改。

PowerExchange for Hive

从版本 10.1.1 开始，PowerExchange for Hive 具有以下适用于 Hive 连接的连接模式：

- 访问作为源或目标的 Hive
- 使用 Hive 在 Hadoop 群集中运行映射

在以前的版本中，连接模式如下：

- 访问 HiveServer2 以运行映射
- 访问 Hive CLI 以运行映射

有关详细信息，请参阅《*Informatica 10.1.1 PowerExchange for Hive 用户指南*》。

PowerExchange for Tableau

从版本 10.1.1 开始，PowerExchange for Tableau 具有以下更改：

- PowerExchange for Tableau 随 Informatica 10.1.1 一起安装。
在以前的版本中，PowerExchange for Tableau 有单独的安装程序。

- 配置目标操作以发布 Tableau 数据提取 (Tableau Data Extract, TDE) 文件时，可以在高级属性中使用附加操作向 Tableau Server 和 Tableau Online 中的现有 TDE 文件添加数据。
在以前的版本中，可以将附加操作配置为仅将 TDE 文件发布到 Tableau Desktop。

有关详细信息，请参阅《*Informatica 10.1.1 PowerExchange for Tableau 用户指南*》。

PowerCenter PowerExchange 适配器

本节介绍版本 10.1.1 中对 PowerCenter 适配器的更改。

PowerExchange for Essbase

从 10.1.1 版本开始，PowerExchange for Essbase 随 PowerCenter 一起安装。

在以前的版本中，PowerExchange for Essbase 有单独的安装程序。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica 10.1.1 PowerExchange for Essbase 用户指南*》。

PowerExchange for Greenplum

从 10.1.1 版本开始，PowerExchange for Greenplum 随 PowerCenter 一起安装。

在以前的版本中，PowerExchange for Greenplum 有单独的安装程序。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica 10.1.1 PowerExchange for Greenplum 用户指南*》。

PowerExchange for Microsoft Dynamics CRM

从 10.1.1 版本开始，PowerExchange for Microsoft Dynamics CRM 随 PowerCenter 一起安装。

在以前的版本中，PowerExchange for Microsoft Dynamics CRM 有单独的安装程序。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica 10.1.1 PowerExchange for Microsoft Dynamics CRM 用户指南*》。

PowerExchange for Tableau

从版本 10.1.1 开始，PowerExchange for Tableau 具有以下更改：

- PowerExchange for Tableau 随 PowerCenter 一起安装。
在以前的版本中，PowerExchange for Tableau 有单独的安装程序。
- 配置目标操作以发布 Tableau 数据提取 (TDE) 文件时，可以在会话属性中配置附加操作以向 Tableau Server 和 Tableau Online 中的现有 TDE 文件添加数据。
在以前的版本中，可以将附加操作配置为仅将 TDE 文件发布到 Tableau Desktop。

有关详细信息，请参阅《*适用于 PowerCenter 的 Informatica 10.1.1 PowerExchange for Tableau 用户指南*》。

转换

本节介绍版本 10.1.1 中已更改的转换行为。

Informatica 转换

本节介绍版本 10.1.1 中对 Informatica 转换的更改。

地址验证器转换

从版本 10.1.1 开始，地址验证器转换使用 5.9.0 版本的 Informatica Address Verification 软件引擎。该引擎支持 Informatica 在版本 10.1.1 中向地址验证器转换添加的功能。

在以前的版本中，转换使用版本 5.8.1 的引擎。

有关详细信息，请参阅《*Informatica 10.1.1 Developer 转换指南*》和《*Informatica 10.1.1 地址验证器端口参考*》。

工作流

本节介绍版本 10.1.1 中已更改的工作流行为。

Informatica 工作流

本节介绍版本 10.1.1 中对 Informatica 工作流行为的更改。

嵌套的非独占网关

从版本 10.1.1 开始，可以向两个非独占网关或两个独占网关之间的序列流添加一对或多对网关。

在以前的版本中，如果向两个非独占网关之间的序列流添加一对网关，这会使工作流失效。

有关详细信息，请参阅《*Informatica 10.1.1 Developer 工作流指南*》。

文档

本节介绍版本 10.1.1 中的文档更改。

Metadata Manager 文档

从版本 10.1.1 开始，《*Informatica Metadata Manager 存储库报告参考*》已过时，因为 Informatica 停止了对报告和仪表板服务以及 JasperReports Server 的支持。

PowerExchange for SAP NetWeaver 文档

从版本 10.1.1 开始，以下指南已过时，因为 Informatica 停止了对分析服务组件功能的支持：

- *Informatica PowerExchange for SAP NetWeaver 分析服务组件指南*
- *Informatica PowerExchange for SAP NetWeaver 分析服务组件传输版本安装须知*

第 34 章

发布任务 (10.1.1)

本章包括以下主题：

- [Metadata Manager, 345](#)

Metadata Manager

本节介绍版本 10.1.1 中 Metadata Manager 的发布任务。

业务情报资源

从版本 10.1.1 开始，部分业务情报资源的**工作线程**配置属性已替换为**多个线程**配置属性。如果在上一版本的 Metadata Manager 中设置了“工作线程”属性，请在升级之后将“多个线程”属性设置为相同值。

为以下资源更新“多个线程”属性的值：

- 业务对象
- Cognos
- Oracle Business Intelligence Enterprise Edition
- Tableau

“多个线程”配置属性控制 Metadata Manager 代理用于异步提取元数据的工作线程的数量。如果在升级之后未更新“多个线程”属性，Metadata Manager 代理会计算工作线程的数量。根据运行 Metadata Manager 代理的计算机上的 JVM 体系结构和可用 CPU 核心数，Metadata Manager 代理将分配一个到六个线程。

有关“多个线程”配置属性的详细信息，请参阅《*Informatica 10.1.1 Metadata Manager 管理员指南*》中的“业务情报资源”一章。

Cloudera 导航器资源

从版本 10.1 开始，必须配置 Cloudera Navigator 服务器的 Java 堆大小和 Metadata Manager 服务的最大堆大小。如果未正确配置堆大小，元数据加载可能会失败。

将 Cloudera Navigator 服务器的 Java 堆大小设置为 2 GB 或以上。如果堆大小不足，资源加载将失败并显示“连接被拒绝”错误。

将 Metadata Manager 服务的最大堆大小设置为 4 GB 或以上。如果要同时执行多个资源加载，请将每个资源加载的最大堆大小增加至少 1 GB。例如，要同时加载两个 Cloudera Navigator 资源，请将最大堆大小增加 2 GB。因此，应将 Metadata Manager 服务的**堆大小上限**属性设置为 6144 MB (6 GB) 或以上。如果最大堆大小不足，加载将失败并显示内存不足错误。

有关 Cloudera Navigator 资源的详细信息，请参阅《*Informatica 10.1.1 Metadata Manager 管理员指南*》中的“数据库管理资源”一章。

Tableau 资源

从版本 10.1.1 开始，Tableau 模型具有次要更改。因此，必须在升级之后清除并重新加载 Tableau 资源。

有关 Tableau 资源的详细信息，请参阅《*Informatica 10.1.1 Metadata Manager 管理员指南*》中的“业务情报资源”一章。

第 VII 部分： 版本 10.1

本部分包含以下章节：

- [新产品 \(10.1\), 348](#)
- [新增功能 \(10.1\), 352](#)
- [更改 \(10.1\), 373](#)
- [发布任务 \(10.1\), 382](#)

第 35 章

新产品 (10.1)

本章包括以下主题：

- [Intelligent Data Lake, 348](#)
- [PowerExchange 适配器, 350](#)

Intelligent Data Lake

随着大数据技术的出现，许多组织采用名为数据湖的新型信息存储模型来解决数据管理挑战。数据湖模型正应用于多种多样的用例，例如，业务情报、分析、法规遵从性和欺诈检测。

数据湖是一个共享存储库，来自各个源的原始数据和企业数据存储在湖中。它通常是基于分布式 Hadoop 群集构建的，该群集可提供经济实用且可扩展的持久层和计算层。Hadoop 使得可以存储来自组织内外部各个企业系统的大量结构化数据和非结构化数据。湖泊中的数据可以包括原始数据和已细化数据、主数据和事务数据、日志文件以及计算机数据。

组织还希望提供一些方法，以便各种类型的用户可以访问并使用 Hadoop 数据湖内部的所有企业数据和数据湖外部的数据。组织希望数据分析人员和数据科学家能够使用数据湖进行即席自助分析以推动业务创新，而不受复杂的基础技术所困扰，也不必掌握高超的编程技能。IT 人员和数据管理人员希望监视企业中与数据相关的用户活动。但如果没有在数据智能的驱动下，奠定坚实的数据管理和治理基础，数据湖就有可能变成数据沼泽。

在版本 10.1 中，Informatica 引入了 Intelligent Data Lake，此新产品可帮助客户从基于 Hadoop 的数据湖获取更高价值并使数据对组织中的所有用户可用。

Intelligent Data Lake 是一款面向数据分析人员和数据科学家的协作式大数据发现和准备自助解决方案。此解决方案可使分析人员能够快速发现原始数据并将其转换为洞察力，并使 IT 人员能够确保质量、可见性和治理。使用 Intelligent Data Lake，分析人员可以将更多人员用在分析上，并减少发现和准备数据所用的时间。

Intelligent Data Lake 可提供以下好处：

- 数据分析人员可以使用语义搜索和智能建议快速而轻松地查找并研究数据湖内外部的可信数据资产。
- 数据分析人员可以使用类似于 Excel 的电子表格界面以自助方式对数据湖中的数据进行转换、清理和扩充，而不需要编码技能。
- 数据分析人员可以发布数据并与社区的其余人员分享知识，还可以使用所选 BI 工具或分析工具来分析数据。
- IT 人员和管理人员可以监视与湖泊中的数据使用相关的用户活动。
- IT 人员可以跟踪数据沿袭，以验证数据是否从正确的源传出以及是否传入到正确的目标。
- IT 人员可以对数据湖强制执行适当的安全和治理。
- IT 人员可以在可重复且可计划的数据交付流程中将数据分析人员所做的工作操作化。

Intelligent Data Lake 具有以下功能：

搜索

- 使用智能搜索和基于推理的结果查找湖泊和其他企业系统中的数据。
- 使用系统属性和自定义的分类基于动态方面筛选资产。

研究

- 获取资产的概览，包括自定义属性、数据质量的剖析统计信息、业务内容的数据域和使用信息。
- 通过进行众包元数据扩充和标记来添加业务上下文信息。
- 预览示例数据，以基于用户凭据感知数据资产。
- 获取资产沿袭，以了解数据的来源和目标并建立数据信任。
- 基于数据资产与其他表或视图、用户、报告和数据域的关联来了解数据资产与企业中其他资产的关系。
- 使用沿袭视图和关系视图以渐进方式发现其他资产。

获取

- 使用基于湖泊的界面将个人的分隔文件上载到湖泊。
系统会自动创建最佳格式的 Hive 表以供上载。
- 创建、附加到或覆盖已上载的数据的资产。

协作

- 通过将数据资产添加到项目来组织工作。
- 将具有不同角色和不同特权的协作者（例如，共有者、编辑者或查看者）添加到项目中。

建议

- 通过使用基于其他用户的行为和共享知识的建议来提高工作效率。
- 获取可在项目中使用的备用资产的建议。
- 获取可在项目中使用的其他资产的建议。
- 建议基于项目中的内容而异。

准备

- 通过类似于 Excel 的环境以交互方式使用示例数据指定转换。
- 查看工作表级别概览和列级别概览，包括值分布以及数字分布和日期分布。
- 以脚本步骤的形式添加转换，并立即在工作表上查看结果。
- 使用字符串运算、数学运算、日期运算和逻辑运算执行列级别数据清理和数据转换。
- 执行工作表级别操作，以组合、合并、汇总或筛选数据。
- 在基础表中的数据更改时刷新工作表中的示例。
- 从现有工作表派生工作表，并在父工作表更改时获取警告。
- 所有转换步骤均存储在可以交互回放的脚本中。

发布

- 使用基础 Hadoop 系统的功能运行大规模数据转换，而不必进行编码或脚本编写。
- 对湖泊中实在很大的数据集运行数据准备步骤，以创建新的数据资产。
- 将湖泊中的数据作为 Hive 表发布到所需的数据库。
- 创建、附加或覆盖已发布的数据的资产。

数据资产操作

- 从湖泊中将数据导出到 CSV 文件。
- 将数据复制到其他数据库或表。
- 删除数据资产（如果用户凭据允许的话）。

我的活动

- 跟踪上载活动及其状态。
- 跟踪发布及其状态。
- 在发生错误时查看日志文件，并根据需要与 IT 管理员共享信息。

IT 监视

- 通过在审计数据库上构建报告来跟踪用户活动、数据资产活动和项目活动。
- 查找诸如最活跃用户、最大数据集、优先的更新、重用频率最高的资产和最活跃项目等信息。

IT 操作化

- 将分析人员所做的即席工作操作化。
- 使用 Informatica Developer 自定义和优化从分析人员所创建的脚本转换的 Informatica Big Data Management 映射。
- 部署、计划和监视 Informatica Big Data Management 映射，以确保在正确的时间向正确的目标交付数据资产。
- 确保根据安全策略授予对数据湖中各个数据库和表的访问权限。

PowerExchange 适配器

Informatica PowerExchange 适配器

本节介绍版本 10.1 中新的 Informatica 适配器。

PowerExchange for Amazon Redshift

从版本 10.1 开始，可以使用 PowerExchange for Amazon Redshift 在 Amazon Redshift 中读取和写入数据。可以将 Amazon Redshift 业务实体作为读取和写入数据对象导入，以创建并运行映射，从而可从 Amazon Redshift 实体中提取数据或将数据加载到 Amazon Redshift 实体中。

有关详细信息，请参阅《*Informatica PowerExchange for Amazon Redshift 10.1 用户指南*》。

PowerExchange for Amazon S3

从版本 10.1 开始，可以创建 Amazon S3 连接，以指定要包括在数据对象中的 Amazon S3 源和目标的位置。可以在数据对象读取和写入操作中使用该 Amazon S3 连接。可以在本地环境中或在 Hadoop 环境中的 Blaze 引擎上验证并运行映射。

有关详细信息，请参阅《*Informatica PowerExchange for Amazon S3 10.1 用户指南*》。

PowerExchange for Microsoft Azure Blob 存储

从版本 10.1 开始，您可以使用 PowerExchange for Microsoft Azure Blob 存储在 Microsoft Azure Blob 存储中读取和写入数据。您可以创建 Microsoft Azure Blob 存储连接，以在 Microsoft Azure Blob 存储数据对象中读取或写入 Microsoft Azure Blob 存储数据。您可以在本地和 Hadoop 环境中验证和运行映射。

有关详细信息，请参阅《*Informatica PowerExchange for Microsoft Azure Blob 存储 10.1 用户指南*》。

[PowerExchange for Microsoft Azure SQL 数据仓库](#)

从版本 10.1 开始，您可以使用 PowerExchange for Microsoft Azure SQL 数据仓库在 Microsoft Azure SQL 数据仓库中读取和写入数据。您可以在本地和 Hadoop 环境中验证和运行映射。

有关详细信息，请参阅《*Informatica PowerExchange for Microsoft Azure SQL 数据仓库 10.1 用户指南*》。

第 36 章

新增功能 (10.1)

本章包括以下主题：

- [应用程序服务, 352](#)
- [大数据, 353](#)
- [Business Glossary, 355](#)
- [连接, 355](#)
- [命令行程序, 356](#)
- [文档, 360](#)
- [异常管理, 360](#)
- [Informatica Administrator, 361](#)
- [Informatica Analyst, 362](#)
- [Informatica Developer, 363](#)
- [Informatica Development Platform, 364](#)
- [Live Data Map, 365](#)
- [映射, 366](#)
- [Metadata Manager, 367](#)
- [PowerCenter, 369](#)
- [PowerExchange 适配器, 369](#)
- [安全性, 370](#)
- [转换, 370](#)
- [工作流, 372](#)

应用程序服务

本节介绍版本 10.1 中的新应用程序服务功能。

系统服务

本节介绍版本 10.1 中的新系统服务功能。

用于配置文件和结果卡的计划程序服务

从版本 10.1 开始，您可以使用计划程序服务计划配置文件运行和结果卡运行，以在特定时间或按特定时间间隔运行。

有关计划的详细信息，请参阅《*Informatica 10.1 管理员指南*》中的“计划”一章。

设置计划的时区

从版本 10.1 开始，选择运行计划的日期和时间时，也要选择时区。设置时区时，请确保作业在您预期的时间运行，无论数据集成服务在哪个位置运行。

有关计划的详细信息，请参阅《*Informatica 10.1 管理员指南*》中的“计划”一章。

大数据

本节介绍版本 10.1 中新增的大数据功能。

Hadoop 生态系统

Big Data Management 10.1 中的支持

从版本 10.1 开始，Informatica 支持以下 Hadoop 发行版的更新版本。

- Azure HDInsight 3.3
- Cloudera CDH 5.5
- MapR 5.1

有关 Big Data Management 10.1 支持的 Hadoop 发行版的完整列表，请参阅《*Informatica Big Data Management 10.1 安装和配置指南*》。

Hadoop 安全系统

从版本 10.1 开始，Informatica 支持 Hadoop 生态系统上的以下安全系统：

- Apache Knox
- Apache Ranger
- Apache Sentry
- HDFS 透明加密

限制适用于安全系统和 Hadoop 分布平台的某些组合。有关对这些技术的 Informatica 支持的详细信息，请参阅《*Informatica Big Data Management 10.1 安全指南*》。

Spark 运行时引擎

从版本 10.1 开始，您可以将映射推入 Hadoop 环境中的 Apache Spark 引擎。

Spark 是 Apache 项目，包含可在 Hadoop 群集中运行映射的运行时引擎。配置特定于 Spark 引擎的 Hadoop 连接属性。创建映射后，您可以使用与 Blaze 和 Hive 引擎相同的方式对其进行验证和查看执行计划。

将映射逻辑推入 Spark 引擎时，数据集成服务会生成 Scala 程序并将它打包到应用程序中。它将应用程序发送到 Spark 执行器，该执行器会将其提交到 Hadoop 群集中的资源管理器。资源管理器会标识资源以运行应用程序。您可以在 Administrator 工具中监视作业。

有关使用 Spark 运行映射的详细信息，请参阅《*Informatica Big Data Management 10.1 用户指南*》。

针对关系源和目标的 Sqoop 连接

从版本 10.1 开始，您可以使用 Sqoop 通过 MapReduce 程序处理关系数据库和 HDFS 之间的数据。您可以使用 Sqoop 导入和导出数据。使用 Sqoop 时，您无需在 Hadoop 群集中的任何节点上安装关系数据库客户端和软件。

要使用 Sqoop，您必须在 JDBC 连接中配置 Sqoop 属性，并在 Hadoop 环境中运行映射。可为基于符合 JDBC 标准的数据库的关系数据对象、自定义数据对象和逻辑数据对象的 Sqoop 连接。例如，可为以下数据库配置 Sqoop 连接：

- Aurora
- IBM DB2
- IBM DB2 for z/OS
- Greenplum
- Microsoft SQL Server
- Netezza
- Oracle
- Teradata

您也可以在 Hadoop 运行时环境中对使用 Sqoop 的数据对象运行配置文件。

有关详细信息，请参阅《*Informatica 10.1 Big Data Management 用户指南*》。

Blaze 引擎上的转换支持

从版本 10.1 开始，Blaze 引擎支持以下转换：

- 地址验证器
- 大小写转换器
- 比较
- 合并
- 数据处理器
- 判定
- 键生成器
- 标签创建器
- 匹配
- 合并
- 规范器
- 解析器
- 序列生成器
- 标准创建器

- 加权平均值

地址验证器转换、合并转换、数据处理器转换、匹配转换和序列生成器转换均受支持，但有限制。

从版本 10.1 开始，Blaze 引擎还支持以下转换：

- 汇总器。支持传递端口。
- 查找支持未连接查找转换。

有关详细信息，请参阅《*Informatica Big Data Management 10.1 用户指南*》的“Hadoop 环境中的映射对象”一章。

Business Glossary

本节介绍版本 10.1 中的新 Business Glossary 功能。

将词汇表内容管理者继承给所有资产

从版本 10.1 开始，Analyst 工具会将您分配给词汇表的管理者和所有者分配给词汇表中的所有资产。

有关详细信息，请参阅《*Informatica 10.1 Business Glossary 指南*》中的“词汇表内容管理”一章。

双向自定义关系

从版本 10.1 开始，可以创建双向自定义关系。可以在关系视图图表中查看相关资产的方向。在双向自定义关系中，沿两个方向为关系提供名称。

有关详细信息，请参阅《*Informatica 10.1 Business Glossary 指南*》中的“查找词汇表内容”一章。

关系视图图表中的自定义颜色

从版本 10.1 开始，可以定义连接关系视图图表中相关资产的线条的颜色。

有关详细信息，请参阅《*Informatica 10.1 Business Glossary 指南*》中的“词汇表管理”一章。

连接

本节介绍版本 10.1 版本中新增的连接功能。

IBM DB2 连接中的架构名称

从版本 10.1 起，使用 IBM DB2 连接向 Developer tool 或 Analyst 工具导入表时，可以指定要从中导入表的一个或多个架构名称。在元数据连接字符串 URL 中使用 ischemaName 属性以指定架构名称。使用管道字符 (|) 分隔多个架构名称。

例如，在元数据连接字符串 URL 中输入以下语法：

```
jdbc:informatica:db2://<主机名>:<端口>;DatabaseName=<数据库名称>;ischemaName=<schema_name1>|<schema_name2>|<schema_name3>
```


此功能在 9.6.1 HotFix 4 中也可用，但在 10.0 中不可用。

有关详细信息，请参阅《*Informatica 10.1 Developer tool 指南*》和《*Informatica 10.1 Analyst 工具指南*》。

命令行程序

本节介绍版本 10.1 中的新命令。

infacmd bg 命令

下表介绍了新的 infacmd bg 命令：

命令	说明
listGlossary	在 Analyst 工具中列出业务词汇表。
exportGlossary	导出 Analyst 工具中可用的业务词汇表。
importGlossary	从 Analyst 工具中导出的 .xlsx 或 .zip 文件导入业务词汇表。

infacmd dis 命令

下表介绍了新的 infacmd dis 命令：

命令	说明
ListApplicationPermissions	列出用户或组对应用程序拥有的权限。
ListApplicationObjectPermissions	列出用户或组对映射或工作流等应用程序对象拥有的权限。
SetApplicationPermissions	将对应用程序的权限分配给用户或组。
SetApplicationObjectPermissions	将对映射或工作流等应用程序对象的权限分配给用户或组。

有关详细信息，请参阅《*Informatica 10.1 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：

命令	说明
BackupData	将内部 Hadoop 群集中的 HDFS 数据备份到 .zip 文件。
UpgradeClusterService	升级 Informatica 群集服务配置。
removeSnapshot	删除现有 HDFS 快照，以便可以成功运行 infacmd ihs BackupData 命令来备份 HDFS 数据。

有关详细信息，请参阅《*Informatica 10.1 命令引用*》中的“infacmd ihs 命令引用”一章。

infacmd isp 命令

下表介绍了新的 infacmd isp 命令：

命令	说明
AssignDefaultOSProfile	向用户或组分配默认操作系统配置文件。
ListDefaultOSProfiles	列出用户或组的默认操作系统配置文件。
ListDomainCiphers	<p>显示 Informatica 域或网关节点使用的以下一个或多个密码套件列表：</p> <p>黑名单</p> <p>Informatica 域阻止的用户指定密码套件列表。</p> <p>默认列表</p> <p>默认情况下 Informatica 支持的密码套件列表。</p> <p>有效列表</p> <p>使用 infasetup updateDomainCiphers 命令对 Informatica 域进行配置后，该域使用的密码套件列表。有效列表支持默认列表和白名单中的密码套件，但是阻止黑名单中的密码套件。</p> <p>白名单</p> <p>除了默认列表外 Informatica 域还可以使用的用户指定密码套件列表。</p> <p>可以指定要显示哪些列表。</p>
UnassignDefaultOSProfile	删除已分配给用户或组的默认操作系统配置文件。

下表介绍了更新的 infacmd isp 命令选项：

命令	说明
CreateOSProfile	<p>添加了以下选项：</p> <ul style="list-style-type: none">-DISProcessVariables-DISEnvironmentVariables-HadoopImpersonationUser-HadoopImpersonationProperties-UseLoggedInUserAsProxy-ProductExtensionName-ProductOptions <p>使用这些选项为数据集成服务配置操作系统配置文件属性。</p>
UpdateOSProfile	<p>添加了以下选项：</p> <ul style="list-style-type: none">-DISProcessVariables-DISEnvironmentVariables-HadoopImpersonationUser-HadoopImpersonationProperties-UseLoggedInUserAsProxy-ProductExtensionName-ProductOptions <p>使用这些选项为数据集成服务配置操作系统配置文件属性。</p>

有关详细信息，请参阅《Informatica 10.1 命令引用》中的“infacmd isp 命令引用”一章。

infacmd ldm 命令

下表介绍了新的 infacmd ldm 命令：

命令	说明
backupData	在本地计算机中创建 HDFS 目录的快照并创建该快照的 .zip 文件。
restoreData	从本地系统检索 HDFS 数据备份 .zip 文件并还原 HDFS 目录中的数据。
removeSnapshot	从 HDFS 目录中删除快照。
升级	升级目录服务。

有关详细信息，请参阅《*Informatica 10.1 命令引用*》中的“infacmd ldm 命令引用”一章。

infacmd ms 命令

下表介绍了 infacmd ms 命令的新选项：

命令	说明
RunMapping	该命令包含以下新选项： - -osp。操作系统配置文件名称（如果数据集成服务已启用，可使用操作系统配置文件）。

有关详细信息，请参阅《*Informatica 10.1 命令引用*》中的“infacmd ms 命令引用”一章。

infacmd ps 命令

下表介绍了新的 infacmd ps 命令选项：

命令	说明
- Execute - executeProfile	这些命令包含以下新选项： - -ospn。操作系统配置文件名称（如果数据集成服务已启用，可使用操作系统配置文件）。

有关详细信息，请参阅《*Informatica 10.1 命令引用*》中的“infacmd ps 命令引用”一章。

infacmd sch 命令

下表介绍了更新的 infacmd sch 命令选项：

命令	说明
CreateSchedule	以下参数已添加到 -RunnableObjects 选项： - -osProfileName。操作系统配置文件名称（如果数据集成服务已启用，可使用操作系统配置文件）。
UpdateSchedule	以下参数已添加到 -AddRunnableObjects 选项： - -osProfileName。操作系统配置文件名称（如果数据集成服务已启用，可使用操作系统配置文件）。

有关详细信息，请参阅《*Informatica 10.1 命令引用*》中的“infacmd sch 命令引用”一章。

infasetup 命令

下表介绍了新增的 infasetup 命令：

命令	说明
ListDomainCiphers	<p>显示 Informatica 域或网关节点使用的以下一个或多个密码套件列表：</p> <p>黑名单</p> <p>Informatica 域阻止的用户指定密码套件列表。</p> <p>默认列表</p> <p>默认情况下 Informatica 支持的密码套件列表。</p> <p>有效列表</p> <p>使用 infasetup updateDomainCiphers 命令对 Informatica 域进行配置后，该域使用的密码套件列表。有效列表支持默认列表和白名单中的密码套件，但是阻止黑名单中的密码套件。</p> <p>白名单</p> <p>Informatica 域可以使用的用户指定密码套件列表。</p> <p>可以指定要显示哪些列表。</p>
updateDomainCiphers	使用新的有效列表更新 Informatica 域可以使用的密码套件。

下表介绍了更新的 infasetup 命令选项：

命令	说明
<ul style="list-style-type: none">- DefineDomain- DefineGatewayNode- DefineWorkerNode- UpdateGatewayNode- UpdateWorkerNode	<p>这些命令包含以下新选项：</p> <ul style="list-style-type: none">- cipherWhiteList -cwl- cipherWhiteListFile -cwlf- cipherBlackList -cbl- cipherBlackListFile -cblf <p>使用这些选项可以为 Informatica 域配置用于在域内安全通信以及与 Web 应用程序服务建立安全连接的密码套件。</p>

有关详细信息，请参阅《Informatica 10.1 命令引用》中的“infasetup 命令引用”一章。

pmrep 命令

下表介绍了新的 pmrep 命令：

命令	说明
AssignIntegrationService	将 PowerCenter 集成服务分配给指定的工作流。

下表介绍了更新的 pmrep 命令选项：

命令	说明
CreateConnection	该命令包含以下更新的选项： - -s。连接类型列表包括 FTP。

有关详细信息，请参阅《Informatica 10.1 命令引用》中的“pmrep 命令引用”一章。

文档

本节介绍 Informatica 10.1 版本的文档中新增或更新的指南。

Informatica 文档包含以下新的指南：

Metadata Manager 命令引用

从版本 10.1 开始，《Metadata Manager 命令引用》包含有关所有 Metadata Manager 命令程序的信息。《Metadata Manager 命令引用》包含在 Metadata Manager 的联机帮助中。在以前的版本中，有关 Metadata Manager 命令程序的信息包含在《Metadata Manager 管理员指南》中。

有关详细信息，请参阅《Informatica 10.1 Metadata Manager 命令引用》。

适用于 Live Data Map® 的 Informatica Administrator 参考

从 Live Data Map 版本 2.0 开始，《适用于 Live Data Map 的 Informatica Administrator 参考》包含有关在 Live Data Map 中执行所需的 Informatica Administrator 任务的基本参考信息。《适用于 Live Data Map 的 Informatica Administrator 参考》包含在 Informatica Administrator 的联机帮助中。

有关详细信息，请参阅《适用于 Live Data Map 的 Informatica 2.0 Administrator 参考》。

异常管理

本节介绍版本 10.1 中新增的异常管理功能。

按数据类型搜索和替换数据值

从版本 10.1 开始，可以在异常任务中配置选项，以便基于数据类型搜索和替换数据值。可以配置选项，在包含日期、字符串或数值数据的任何列中搜索和替换数据。

指定数据类型时，Analyst 工具会在使用该数据类型的任何列中搜索您输入的值。可以查找和替换字符串数据列中包含的任何值。可以对字符串数据执行区分大小写的搜索。可以搜索字符串数据列中某个字段内容和搜索值之间的部分匹配项或完整匹配项。

此功能在 9.6.1 HotFix 4 中也可用，但在 10.0 中不可用。

有关详细信息，请参阅《Informatica 10.1 异常管理指南》中的“异常记录”一章。

Informatica Administrator

本节介绍版本 10.1 中的新 Administrator 工具功能。

域视图

从 10.1 开始，可以查看域中 CPU 使用量和内存使用量的历史统计信息。

可以查看过去 60 分钟的 CPU 和内存使用量统计信息。可以在当前统计信息和过去 60 分钟之间切换。在域视图中，选择操作 > 当前，或在 CPU 使用量面板或内存使用量面板中选择操作 > 最近一小时趋势。

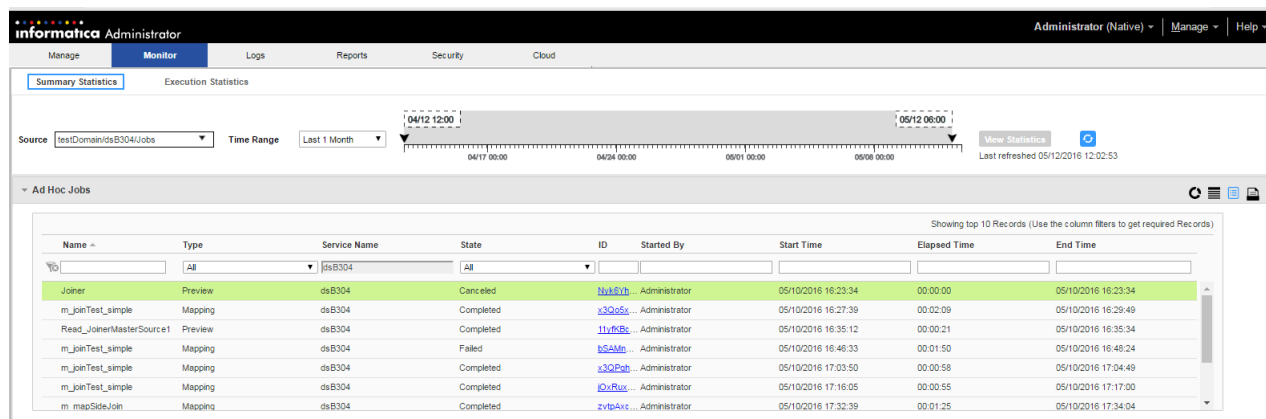
监视

从版本 10.1 开始，Administrator 工具中的“监视”选项卡包含以下功能：

“摘要统计信息”视图上的“详细信息”视图

摘要统计信息视图包含了详细信息视图。您可以查看有关作业的信息，将列表导出为 .csv 文件，以及链接到执行统计信息视图中的作业。要访问“详细信息”视图，请单击[查看详细信息](#)。

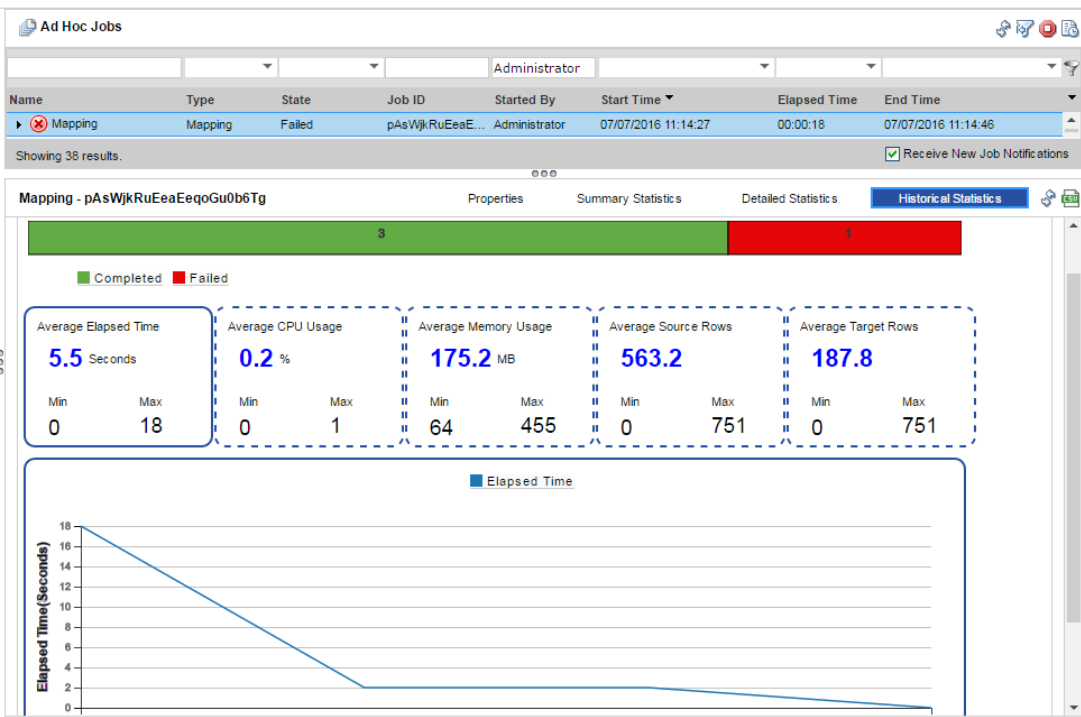
下图显示了详细信息视图：



“历史统计信息”视图。

在监视选项卡的内容面板中选择 Ad Hoc 或部署的映射作业时，详细信息面板会包含历史统计信息视图。历史统计信息视图会显示多次运行特定作业的平均数据。例如，您可以查看映射作业的最小、最大和平均持续时间。您可以查看该作业运行时使用的平均 CPU 量。

下图显示了**历史统计信息**视图：



Informatica Analyst

本节介绍版本 10.1 中的新 Analyst 工具功能。

配置文件

本节介绍针对配置文件和结果卡新增的 Analyst 工具功能。

遵从性条件

从版本 10.1 开始，您可以选择最少数量的遵从行数作为数据域发现的遵从性条件。

有关遵从性条件的详细信息，请参阅《Informatica 10.1 Data Discovery 指南》中的“Informatica Analyst 中的数据域发现”一章。

针对数据域发现排除空值

从版本 10.1 开始，对数据源执行数据域发现时，可以从数据集中排除空值。使用排除空值选项选择最小行数百分比时，遵从性百分比等于匹配行数除以总行数（列中为空值的排除在外）的比率。

有关从数据域发现选项中排除空值的详细信息，请参阅《Informatica 10.1 Data Discovery 指南》的“Informatica Analyst 中的数据域发现”一章。

运行时环境

从版本 10.1 开始，创建或编辑列配置文件、数据域发现配置文件、企业发现配置文件或结果卡时，您可以选择 Hadoop 选项作为运行时环境。选择 Hadoop 选项时，数据集成服务会将配置文件逻辑推入 Hadoop 群集中的 Blaze 引擎，以运行配置文件。

有关运行时环境的详细信息，请参阅《*Informatica 10.1 Data Discovery 指南*》中的“数据对象配置文件”一章。

结果卡仪表板

从版本 10.1 开始，您可以在结果卡仪表板中查看以下结果卡详细信息：

- 项目中结果卡的总数
- 过去六个月结果卡的运行趋势
- 数据对象总数和拥有结果卡的数据对象的数量
- 过去六个月的累计度量趋势

有关结果卡仪表板的详细信息，请参阅《*Informatica 10.1 Data Discovery 指南*》的“Informatica Analyst 中的结果卡”一章。

Informatica Developer

本节介绍 10.1 版本中的新 Informatica Developer 功能。

生成源文件名

从版本 10.1 开始，您可以使用文件名列选项返回源文件名。可以对映射进行配置，使其将源文件名写入到每个源行。

有关详细信息，请参阅《*Informatica 10.1 Developer tool 指南*》。

从 PowerCenter 导入

从版本 10.1 开始，您可以将包含 Netezza 和 Teradata 对象的映射从 PowerCenter 导入到 Developer Tool，并在本地或 Hadoop 运行时环境中运行映射。

有关详细信息，请参阅《*Informatica 10.1 Developer 映射指南*》。

在 Excel 和 Developer Tool 之间复制文本

从版本 10.1 开始，可以将文本从 Excel 复制到 Developer tool 或从 Developer tool 复制到 Excel。将文本从 Excel 复制到 Developer tool 以为转换提供元数据。例如，已在 Excel 中设计包括所有转换及其端口名称、数据类型和转换逻辑的映射。在 Developer tool 中，可以将字段从 Excel 复制到空转换的端口。同样，也可以将转换端口从 Developer tool 复制到 Excel。

逻辑数据对象读取和写入映射编辑

从 Informatica 10.1 开始，您可以使用逻辑数据对象编辑器编辑和更改逻辑数据对象读取和写入映射中的元数据。有关详细信息，请参阅《*Informatica 10.1 Developer Tool 指南*》中的“数据的逻辑视图”一章。

DDL 查询

从版本 10.1 开始，如果选择在运行时创建或替换目标，则您可以定义 DDL 查询，数据集成服务必须在运行时基于该查询创建或替换目标表。可以为关系目标和 Hive 目标定义 DDL 查询。

可以在 DDL 查询中输入占位符。数据集成服务会在运行时将占位符替换为实际值。例如，如果表包含 50 个列，则您可以输入占位符，而不必在 DDL 查询中输入所有列名称。

可以在 DDL 查询中输入以下占位符：

- INFA_TABLE_NAME
- INFA_COLUMN_LIST
- INFA_PORT_SELECTOR

还可以在 DDL 查询中输入参数。

有关详细信息，请参阅《*Informatica 10.1 Developer 映射指南*》。

配置文件

本节介绍为配置文件和结果卡新增的 Developer Tool 功能。

使用 Avro 和 Parquet 数据源的列配置文件

从版本 10.1 开始，您可以在 HDFS 中的 Avro 或 Parquet 数据源上创建列配置文件。

有关 Avro 和 Parquet 数据源上的列配置文件的详细信息，请参阅《*Informatica 10.1 Data Discovery 指南*》中的“半结构化数据源上的列配置文件”。

遵从性条件

从版本 10.1 开始，您可以选择最少数量的遵从行数作为数据域发现的遵从性条件。

有关遵从性条件的详细信息，请参阅《*Informatica 10.1 Data Discovery 指南*》的“Informatica Developer 中的数据域发现”。

针对数据域发现排除空值

从版本 10.1 开始，对数据源执行数据域发现时，可以从数据集中排除空值。使用排除空值选项选择最小行数百分比时，遵从性百分比等于匹配行数除以总行数（列中为空值的排除在外）的比率。

有关从数据域发现选项中排除空值的详细信息，请参阅《*Informatica 10.1 Data Discovery 指南*》的“Informatica Developer 中的数据域发现”一章。

运行时环境

从版本 10.1 开始，创建或编辑列配置文件、数据域发现配置文件、企业发现配置文件或结果卡时，您可以选择 Hadoop 选项作为运行时环境。选择 Hadoop 选项时，数据集成服务会将配置文件逻辑推入 Hadoop 群集中的 Blaze 引擎，以运行配置文件。

有关运行时环境的详细信息，请参阅《*Informatica 10.1 Data Discovery 指南*》中的“数据对象配置文件”一章。

Informatica Development Platform

本节介绍 Informatica Development Platform 的新功能和增强功能。

Informatica 连接器工具包

从版本 10.1 开始，可以使用 Informatica 连接器工具包中的以下功能：

预定义类型系统

创建使用 REST API 连接到数据源的连接器时，可以使用预定义的数据类型。您可以使用以下 Informatica Platform 数据类型：

- 字符串
- 整数
- 长整数
- 小数
- 双精度
- 二进制
- 日期

过程模式

为 Informatica Cloud 创建连接器时，可以为数据源中的过程定义本地元数据对象。您可以使用以下选项为过程定义本地元数据对象：

手动创建本地元数据对象

手动定义本地元数据对象时，可以指定以下详细信息：

元数据组件	说明
过程扩展	可为过程指定的其他元数据信息。
参数扩展	可为参数指定的其他元数据信息。
调用功能属性	可以指定的用以创建过程的读取或写入调用的其他元数据信息。

使用 swagger 规范

使用 swagger 规范定义本地元数据对象时，可以使用现有的 swagger 规范，也可以通过对 REST 端点进行采样来生成 swagger 规范。

编辑通用元数据

您可以为 Informatica Cloud 连接器指定通用元数据信息，例如架构名称和外键名称。

导出 Informatica Cloud 的连接器文件

设计和实施连接器组件后，可以通过指定插件 ID 和插件版本导出 Informatica Cloud 的连接器文件。

导出 PowerCenter 的连接器文件

设计和实施连接器组件后，可以通过指定 PowerCenter 版本导出 PowerCenter 的连接器文件。

Live Data Map

本节介绍版本 10.1 中的新 Live Data Map 功能。

电子邮件通知

从版本 10.1 开始，可以配置和接收目录服务状态上的电子邮件通知以密切监视应用程序服务问题并进行故障排除。可使用电子邮件服务和关联的模型存储库服务来发送电子邮件通知。

有关详细信息，请参阅《*适用于 Live Data Map 的 Informatica 10.1 Administrator 参考*》。

关键字搜索

从版本 10.1 开始，您可以使用以下关键字将搜索结果限制为特定资产类型：

- 表
- 列
- 文件
- 报告

例如，如果要搜索包含术语“customer”的所有表，请在“搜索”框中键入“tables with customer”。Enterprise Information Catalog 会列出表名称包含搜索术语“customer”的所有表。

有关详细信息，请参阅《*Informatica 10.1 Enterprise Information Catalog 用户指南*》。

剖析

从版本 10.1 开始，Live Data Map 可在 Hadoop 环境中运行配置文件。选择 Hadoop 连接时，数据集成服务会将配置文件逻辑推入 Hadoop 群集中的 Blaze 引擎，以运行配置文件。

有关详细信息，请参阅《*Informatica 10.1 Live Data Map 管理员指南*》。

扫描器

从版本 10.1 开始，可以从以下来源提取元数据：

- Amazon Redshift
- Amazon S3
- 自定义沿袭
- HDFS
- Hive
- Informatica Cloud
- MicroStrategy

有关详细信息，请参阅《*Informatica 10.1 Live Data Map 管理员指南*》。

映射

本节介绍版本 10.1 中的新映射功能。

Informatica 映射

本节介绍版本 10.1 中的 Informatica 映射的新功能。

从已连接转换生成 Mapplet

从版本 10.1 开始，可以从映射中的已连接转换组生成 Mapplet。使用 Mapplet 作为模板来添加到与不同源和目标连接的多个映射。

从 SQL 查询生成映射或逻辑数据对象

从版本 10.1 开始，可以在 Developer tool 中从 SQL 查询生成映射或逻辑数据对象。

要从 SQL 查询生成映射或逻辑数据对象，请单击**文件 > 新建 > 从 SQL 查询映射**。输入 SQL 查询或使用 SQL 查询选择要转换为映射的文本文件的位置。还可以从 SQL 查询生成仅包含 SELECT 语句的逻辑数据对象。

有关从 SQL 查询生成映射或逻辑数据对象的详细信息，请参阅《*Informatica 10.1 Developer 映射指南*》。

Metadata Manager

本节介绍版本 10.1 中新增的 Metadata Manager 功能。

通用资源

从版本 10.1 开始，您可以创建通用资源，以从 Metadata Manager 未打包模型的元数据源中提取元数据。例如，您可以创建通用资源，以从 Apache Hadoop Hive Server、QlikView 或 Talend 元数据源中提取元数据。

要从这些源中提取元数据，请首先创建表示元数据源类型的 XConnect。XConnect 包括元数据源的模型。然后，创建一个或多个基于模型的资源。您创建的通用资源的行为类似于 Metadata Manager 中的打包资源。

有关通用资源的详细信息，请参阅《*Informatica 10.1 Metadata Manager 管理员指南*》中的“通用资源”一章。

适用于 Oracle 资源和 Teradata 资源的增量加载

从版本 10.1 开始，您可以为 Oracle 资源和 Teradata 资源启用增量加载。增量加载使 Metadata Manager 加载最近对元数据的更改，而不是加载完整的元数据。增量加载可以缩短加载资源所需的时长。

要为 Oracle 资源或 Teradata 资源启用增量加载，请在资源配置属性中启用**增量加载**选项。默认情况下，此选项处于禁用状态。

有关适用于 Oracle 资源和 Teradata 资源的增量加载的详细信息，请参阅《*Informatica 10.1 Metadata Manager 管理员指南*》中的“数据库管理资源”一章。

隐藏摘要视图中的资源

从版本 10.1 开始，可以防止资源及其子对象显示在数据沿袭图的摘要视图中。要隐藏资源，请启用资源配置属性的**属性**页上的**在摘要沿袭中隐藏**选项。此选项适用于所有资源类型。默认处于禁用状态。

可以隐藏数据沿袭图表中的对象，如暂存数据库。如果要查看隐藏的对象，可以通过任务栏从摘要视图切换到详细信息视图。

有关数据沿袭图表的摘要视图的详细信息，请参阅《*Informatica 10.1 Metadata Manager 用户指南*》中的“使用数据沿袭”一章。

从多个包文件创建 SQL Server Integration Services 资源

从版本 10.1 开始，可以创建从不同的包文件 (.dtsx) 中的包提取元数据的 Microsoft SQL Server Integration Services 资源。包文件必须位于同一个目录中。

要创建从不同的包文件中的包提取元数据的资源，请在**目录**资源配置属性中指定包含这些包文件的目录。

有关创建和配置 Microsoft SQL Server Integration Services 资源的详细信息，请参阅《*Informatica 10.1.1 Metadata Manager 管理员指南*》中的“数据库管理资源”一章。

Metadata Manager 命令行程序

从版本 10.1 开始，Metadata Manager 有了新的命令行程序。mmXConPluginUtil 命令行程序会生成用于通用 XConnect 的图像映射信息或插件。

下表介绍了 mmXConPluginUtil 命令：

命令名称	说明
generateImageMapping	生成用于通用 XConnect 的图像映射信息。
generatePlugin	生成用于通用 XConnect 的插件。

有关 mmXConPluginUtil 命令行程序的详细信息，请参阅《*Informatica 10.1 Metadata Manager 命令参考*》中的“mmXConPluginUtil”一章。

应用程序属性

从版本 10.1 开始，可以在 Metadata Manager imm.properties 文件中配置新的应用程序属性。此功能在 9.6.1 HotFix 4 中也可用，但在 10.0 中不可用。

下表描述了 imm.properties 中的新 Metadata Manager 应用程序属性：

属性	说明
xconnect.custom.failLoadOnErrorCount	自定义资源加载失败之前允许 Metadata Manager 服务遇到的错误的最大数目。
xconnect.io.print.batch.errors	在您加载自定义资源时，Metadata Manager 服务在一个批处理中写入内存中缓存和 mm.log 文件的错误的数目。

有关 imm.properties 文件的详细信息，请参阅《*Informatica 10.1 Metadata Manager 管理员指南*》中的“Metadata Manager 属性文件”附录。

将 Business Glossary 审计跟踪的历史记录和链接迁移到技术元数据

从版本 10.1 开始，可以在导出业务词汇表时将审计跟踪的历史记录和链接迁移到技术元数据。可以在 Analyst 工具中导入审计跟踪的历史记录和链接。

此功能在 9.6.1 HotFix 4 中也可用，但在 10.0 中不可用。

有关详细信息，请参阅《*从 Informatica 版本 9.5.1 升级到 10.1 指南*》。

PowerCenter

本节介绍版本 10.1 中新增的 PowerCenter 功能。

从目标定义创建源定义

从版本 10.1 开始，您可以从目标定义创建源定义。将目标定义拖曳到 Source Analyzer 即可创建源定义。

有关详细信息，请参阅《*Informatica 10.1 PowerCenter Designer 指南*》。

从命令行创建 FTP 连接类型

从版本 10.1 开始，您可以使用 `pmrep CreateConnection` 命令创建 FTP 连接。

有关详细信息，请参阅《*Informatica 10.1 命令参考*》中的“pmrep 命令参考”一章。

针对 Greenplum 的下推优化

从版本 10.1 开始，如果连接类型为 ODBC，PowerCenter 集成服务可以将转换逻辑推入 Greenplum 源和目标。

有关详细信息，请参阅《*Informatica PowerCenter 10.1 高级工作流指南*》。

PowerExchange 适配器

本节介绍版本 10.1 中新增的 PowerExchange 适配器功能。

Informatica PowerExchange 适配器

本节介绍版本 10.1 中新增的 Informatica 适配器功能。

PowerExchange for HDFS

从版本 10.1 开始，无需通过数据处理器转换，使用 PowerExchange for HDFS 即可在 HDFS 和本地文件系统中读取和写入 Avro 和 Parquet 数据文件。

有关详细信息，请参阅《*Informatica PowerExchange for HDFS 10.1 用户指南*》。

PowerExchange for Hive

从版本 10.1 开始，您可以在映射中使用 char 和 varchar 数据类型。还可以在创建数据对象和映射时选择不同的 Hive 数据库。

有关详细信息，请参阅《*Informatica PowerExchange for Hive 10.1 用户指南*》。

PowerExchange for Teradata Parallel Transporter API

从版本 10.1 开始，您可以让 Teradata Connector for Hadoop (TDCH) 在 Blaze 引擎上运行 Teradata 映射。运行映射时，数据集成服务会将映射推入 Hadoop 群集并在 Blaze 引擎上处理映射，这将大幅提升性能。

有关详细信息，请参阅《*Informatica PowerExchange for Teradata Parallel Transporter API 10.1 用户指南*》。

PowerCenter PowerExchange 适配器

本节介绍版本 10.1 中新增的 PowerCenter 适配器功能。

PowerExchange for Greenplum

从版本 10.1 开始，可以为本地 Greenplum 连接配置 Kerberos 身份验证。

此功能在 9.6.1 HotFix 4 中也可用，但在 10.0 中不可用。

有关详细信息，请参阅《适用于 PowerCenter 的 Informatica 10.1 PowerExchange for Greenplum 用户指南》中的“Greenplum 会话和工作流”一章。

安全性

本节介绍版本 10.1 中的新安全功能。

自定义密码套件

从版本 10.1 开始，可以自定义 Informatica 域用于在域内安全通信以及与 Web 应用程序服务建立安全连接的密码套件。可以创建白名单和黑名单来启用或阻止特定的密码套件。此功能在 9.6.1 HotFix 4 中也可用，但在 10.0 中不可用。

Informatica 域使用密码套件的有效列表，该列表允许使用默认列表和白名单中的密码套件，但是阻止黑名单中的密码套件。

有关详细信息，请参阅《Informatica 10.1 安全指南》中的“域安全”一章。

操作系统配置文件

从版本 10.1 开始，如果数据集成服务在 UNIX 或 Linux 上运行，您可以创建操作系统配置文件，并配置数据集成服务以使用操作系统配置文件。使用操作系统配置文件可以提高安全性以及在 Big Data Management、Data Quality 和 Intelligent Data Lake 等 Informatica 产品中隔离运行时用户环境。

数据集成服务使用操作系统配置文件运行映射、配置文件、结果卡和工作流。操作系统配置文件包含操作系统用户名、服务进程变量、Hadoop 模拟属性、分析服务属性、环境变量和权限。数据集成服务使用操作系统用户的系统权限和操作系统配置文件中定义的属性来运行映射、配置文件、结果卡或工作流。

有关操作系统配置文件的详细信息，请参阅《Informatica 10.1 安全指南》中的“用户和组”一章。

应用程序和应用程序对象权限

从版本 10.1 起，可以通过分配权限来控制用户或组对应用程序和应用程序对象（如映射和工作流）的访问级别。

有关应用程序和应用程序对象权限的详细信息，请参阅《Informatica 10.1 安全指南》中的“权限”一章。

转换

本节介绍版本 10.1 中的新转换功能。

Informatica 转换

本节介绍版本 10.1 中的 Informatica 转换的新功能。

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：

爱尔兰

从版本 10.1 开始，可以返回爱尔兰地址对应的 eircode。eircode 是唯一标识爱尔兰地址的七字符代码。eircode 体系涵盖了所有的住宅、公共建筑物和经营场所，并包括公寓地址和村镇地址。

要返回某个地址的 eircode，请选择“邮政编码”端口或“邮政编码填写”端口。

法国

从版本 10.1 开始，地址验证使用国家地址管理服务的 Hexaligne 3 存储库根据 SNA 标准验证法国地址。

Hexaligne 3 数据集包含有关收件人地点地址的附加信息，包括子建筑物详细信息，如建筑物名称和住宅名称。

德国

从版本 10.1 开始，可以检索 *Frachtleitcode* 或货物代码的三位街道代码部分，以作为有效的德国地址的补充信息。街道代码标识地址所在的街道。

要检索作为已验证德国地址的补充信息提供的街道代码，请选择“街道代码德国”端口。在“德国补充”端口组中查找此端口。

Informatica 在版本 10.1 中添加了“街道代码德国”端口。

韩国

从版本 10.1 开始，可以验证韩国早期的基于地段的地址和包含早期六位邮政编码的地址。可以验证并更新使用当前格式、早期格式以及组合使用当前格式和早期格式的地址。当前的韩国地址采用基于街道的格式并包含五位数字的邮政编码。非当前地址采用基于地段的格式并包含六位数字的邮政编码。

要验证旧格式的韩国地址并将信息更改为其他格式，请使用“地址标识符韩国”端口。可以在两个阶段更新地址信息。首先，在批处理或交互模式下运行地址验证映射，然后选择“地址标识符韩国”输出端口。然后，在地址代码查找模式下运行地址验证映射，然后选择“地址标识符韩国”输入端口。在“离散值”端口组中查找“地址标识符韩国”输入端口。在“韩国补充”端口组中查找“地址标识符韩国”输出端口。

要验证地址验证器转换是否能够读取和写入地址数据，请将“补充韩国状态”端口添加到该转换中。

Informatica 在版本 10.1 中添加了“地址标识符韩国”端口、“补充韩国状态”端口和“韩国补充”端口组。

从版本 10.1 开始，可以在韩语脚本和拉丁语脚本中检索韩国地址数据。

英国

从版本 10.1 开始，可以检索英国地址的收件人地点类型数据和组织键数据。收件人地点类型是单字符代码，表示地址是指向住宅、小型组织还是大型组织。组织关键字是英国皇家邮政分配给小型组织的八位数字代码。

要将收件人地点类型添加到英国地址，请使用“收件人地点类型英国”端口。要将组织关键字添加到英国地址，请使用“组织关键字英国”端口。在“英国补充”端口组中查找此端口。要验证地址验证器转换是否能够读取和写入数据，请将“补充英国状态”端口添加到该转换中。

Informatica 在版本 10.1 中添加了“收件人地点类型英国”端口和“组织键英国”端口。

这些功能在 9.6.1 HotFix 4 中也有提供。但在 10.0 中不可用。

有关详细信息，请参阅《Informatica 10.1 地址验证器端口参考》。

数据处理器转换

本节介绍新增的数据处理器转换功能。

REST API

应用程序可以调用 Data Transformation REST API 来运行 Data Transformation 服务。

有关详细信息，请参阅《*Informatica 10.1 Data Transformation REST API 用户指南*》。

XmlToDocument_45 文档处理器

XmlToDocument_45 文档处理器可将 XML 数据转换为文档格式，例如 PDF 或 Excel。此组件使用 **Business Intelligence and Reporting Tool (BIRT)** 版本 4.5 Eclipse 加载项。较旧的 BIRT 版本的文档处理器也可用。

有关详细信息，请参阅《*Informatica 10.1 Data Transformation 用户指南*》。

关系到分层的转换

本节介绍您在 Developer tool 中创建的关系到分层转换。

关系到分层转换是版本 10.1 中引入的优化转换，可将关系输入转换为分层输出。

有关详细信息，请参阅《*Informatica 10.1 Developer 转换指南*》。

工作流

本节介绍版本 10.1 中新增的工作流功能。

PowerCenter 工作流

本节介绍版本 10.1 中的 PowerCenter 工作流的新功能。

将工作流分配给 PowerCenter 集成服务

从版本 10.1 开始，可以使用 `pmrep AssignIntegrationService` 命令将工作流分配给 PowerCenter 集成服务。

有关详细信息，请参阅《*Informatica 10.1 命令参考*》中的“pmrep 命令参考”一章。

第 37 章

更改 (10.1)

本章包括以下主题：

- [支持更改, 373](#)
- [应用程序服务, 374](#)
- [大数据, 374](#)
- [Business Glossary, 375](#)
- [命令行程序, 376](#)
- [异常管理, 376](#)
- [Informatica Developer, 377](#)
- [Live Data Map, 377](#)
- [Metadata Manager, 378](#)
- [PowerCenter, 378](#)
- [安全, 379](#)
- [转换, 379](#)
- [工作流, 380](#)

支持更改

从版本 10.1 开始，Informatica 宣布了以下支持更改：

Informatica 安装

从版本 10.1 开始，Informatica 对操作系统实施了以下更改：

支持更改	支持级别	注释
SUSE 11	添加了支持	从版本 10.1 开始，Informatica 添加了对 SUSE Linux Enterprise Server 11 的支持。

报告服务（已弃用）

从版本 10.1 开始，Informatica 弃用了报告服务。在将来的版本中，Informatica 将取消对报告服务的支持。报告服务自定义角色已弃用。

如果升级到版本 10.1，您可以继续使用报告服务。您可以继续使用 Data Analyzer。Informatica 建议在 Informatica 取消支持之前开始使用第三方报告工具。您可以使用建议的 SQL 查询构建 PowerCenter 早期版本随附的所有报告。

如果安装版本 10.1，则无法创建报告服务。您不能使用 Data Analyzer。必须使用第三方报告工具来运行 PowerCenter 和 Metadata Manager 报告。

有关 PowerCenter 报告的信息，请参阅《*Informatica PowerCenter 使用 PowerCenter 报告指南*》。有关 PowerCenter 存储库视图的信息，请参阅《*Informatica PowerCenter 存储库指南*》。有关 Metadata Manager 存储库视图的信息，请参阅《*Informatica Metadata Manager 视图参考*》。

报告和仪表板服务（已弃用）

自版本 10.1 开始，Informatica 已弃用报告和仪表板服务。在将来的版本中，Informatica 将取消对报告和仪表板服务的支持。

如果升级到 10.1，可以继续使用报告和仪表板服务。Informatica 建议在 Informatica 取消支持之前开始使用第三方报告工具。您可以使用建议的 SQL 查询构建 PowerCenter 早期版本随附的所有报告。

如果安装版本 10.1，则无法创建报告和仪表板服务。必须使用第三方报告工具来运行 PowerCenter 和 Metadata Manager 报告。

有关 PowerCenter 报告的信息，请参阅《*Informatica PowerCenter 使用 PowerCenter 报告指南*》。有关 PowerCenter 存储库视图的信息，请参阅《*Informatica PowerCenter 存储库指南*》。有关 Metadata Manager 存储库视图的信息，请参阅《*Informatica Metadata Manager 视图参考*》。

应用程序服务

本节介绍版本 10.1 中对应用程序服务的更改。

系统服务

本节介绍版本 10.1 中对系统服务的更改。

结果卡通知使用的电子邮件服务

从版本 10.1 起，结果卡通知使用在电子邮件服务中配置的电子邮件服务器。

以前，结果卡通知使用的是在域中配置的电子邮件服务器。

有关电子邮件服务的详细信息，请参阅《*Informatica 10.1 应用程序服务指南*》中的“系统服务”一章。

大数据

本节介绍对 Big Data 功能的更改。

JCE 策略文件安装

从版本 10.1 开始，Informatica Big Data Management 提供了 JCE 策略文件，并且会在用户运行安装程序时安装该策略文件。

在以前的版本中，您必须下载适用于 AES 加密的 JCE 策略文件并手动安装该策略文件。

Business Glossary

本节介绍版本 10.1 中对 Business Glossary 的更改。

自定义关系

从版本 10.1 开始，可以在**管理词汇表关系**工作区中创建自定义关系。在**管理**下，单击**词汇表关系**以打开**管理词汇表关系**工作区。

在以前的版本中，必须编辑词汇表模板才能创建自定义关系。

有关详细信息，请参阅《*Informatica 10.1 Business Glossary 指南*》中的“词汇表管理”一章。

双向默认关系

从版本 10.1 开始，默认业务术语关系是双向的。

在以前的版本中，默认关系是单向的。

有关详细信息，请参阅《*Informatica 10.1 Business Glossary 指南*》中的“查找词汇表内容”一章。

控制者关系

从版本 10.1 开始，您将无法在各个术语之间创建“控制者”关系。“控制者”关系只能在策略和术语之间使用。

在以前的版本中，您可以在各个术语之间创建“控制者”关系。

有关详细信息，请参阅《*Informatica 10.1 Business Glossary 指南*》。

词汇表工作区

从版本 10.1 开始，在**词汇表**工作区中，Analyst 工具将在单独的选项卡上显示多个词汇表资产。

在以前的版本中，Analyst 工具仅在**词汇表**工作区中显示一个词汇表资产。

有关详细信息，请参阅《*Informatica 10.1 Business Glossary 指南*》中的“查找词汇表内容”一章。

Business Glossary Desktop

从版本 10.1 开始，可以在 OS X 操作系统上安装 Business Glossary Desktop。

在以前的版本中，Business Glossary Desktop 仅适用于 Windows。

有关详细信息，请参阅《*Informatica 10.1 Business Glossary Desktop 安装和配置指南*》。

Business Glossary 命令程序的 Kerberos 身份验证

从版本 10.1 开始，使用 Kerberos 身份验证的域支持 Business Glossary 命令程序。

在以前的版本中，使用 Kerberos 身份验证的域不支持 Business Glossary 命令程序。

有关详细信息，请参阅《*Informatica 10.1 命令引用*》中的“infacmd bg 命令参考”一章。

命令行程序

本节介绍版本 10.1 中对命令的更改。

infacmd isp 命令

下表介绍了已弃用的 infacmd isp 命令：

命令	说明
BackupDARepositoryContents	将 Data Analyzer 存储库的内容备份到二进制文件。备份内容时，报告服务会保存 Data Analyzer 存储库，其中包括存储库对象、连接信息和代码页信息。
CreateDARepositoryContents	升级 Data Analyzer 存储库的内容。可以在创建报告服务或删除存储库目录时添加存储库目录。无法为已包含内容的存储库创建内容。
CreateReportingService	在域中创建报告服务。
DeleteDARepositoryContents	从 Data Analyzer 存储库中删除存储库目录。删除存储库目录时，也删除了分配给报告服务的用户的所有特权和角色。
RestoreDARepositoryContents	从二进制文件还原 Data Analyzer 存储库的内容。可以从存储库备份文件将元数据还原到数据库。如果在现有数据库上还原备份文件，则会覆盖现有内容。
UpdateReportingService	更新或创建报告服务的服务和沿袭选项。
UpgradeDARepositoryContents	升级 Data Analyzer 存储库的内容。
UpgradeDARepositoryUsers	升级 Data Analyzer 存储库中的用户和组。升级 Data Analyzer 存储库中的用户和组时，服务管理器会将其移动到 Informatica 域中。

有关详细信息，请参阅《Informatica 10.1 命令引用》中的“infacmd isp 命令引用”一章。

异常管理

本节介绍对版本 10.1 中的异常管理的更改。

异常任务中的默认搜索和替换操作

从版本 10.1 开始，可以配置异常任务中的选项以在一列或多列中查找和替换数据值。可以指定一列，也可以指定使用字符串、日期或数值数据类型的任何列。默认情况下，查找和替换操作适用于包含字符串数据的所有列。

在以前的版本中，默认情况下会对任务中的所有数据运行查找和替换操作。在版本 10.1 中，无法将查找和替换操作配置为在任务中的所有数据上运行。

有关详细信息，请参阅《Informatica 10.1 异常管理指南》中的“异常记录”一章。

Informatica Developer

本节介绍版本 10.1 中对 Developer tool 的更改。

键盘快捷方式

从版本 10.1 开始，用于选择下一个区域的快捷键是先按 **Ctrl+Tab**，然后按 **Tab** 键三次。

在以前的版本中，该快捷键是先后按 Ctrl+Tab 两次。

有关详细信息，请参阅《*Informatica 10.1.1 Developer tool 指南*》中的“键盘快捷键”附录。

Live Data Map

本节介绍版本 10.1 中 Live Data Map 的变化。

Enterprise Information Catalog

本节介绍对 Enterprise Information Catalog 的更改。

主页

从版本 10.1 开始，主页将显示趋势搜索、前 50 个资产和最近查看的资产。趋势搜索是指上周在目录中搜索最多的术语。前 50 个资产是指目录中与其他资产存在最多关系的资产。

在以前的版本中，Enterprise Information Catalog 主页显示了搜索字段、Live Data Map 从中扫描元数据的资源数以及目录中的总资产数。

有关 Enterprise Information Catalog 主页的详细信息，请参阅《*Informatica 10.1 Enterprise Information Catalog 用户指南*》中的“Informatica Enterprise Information Catalog 入门”一章。

资产概览

从版本 10.1 开始，可以在概览选项卡中查看与资产关联的架构名称。

在以前的版本中，资产的“概览”选项卡未显示关联的架构名称。

有关 Enterprise Information Catalog 中的资产的详细信息，请参阅《*Informatica 10.1 Enterprise Information Catalog 用户指南*》。

“Live Data Map 管理员”主页

从版本 10.1 开始，“启动”工作区会显示目录中的资产总数、未使用的资源和未分配的连接，以及许多其他监视统计信息。

在以前的版本中，“Live Data Map 管理员”主页显示多项监视统计信息，例如每种资源的资源数量、任务分布和预测的作业加载。

有关“Live Data Map 管理员”主页的详细信息，请参阅《*Informatica 10.1 Live Data Map 管理员指南*》中的“使用 Live Data Map 管理员”一章。

Metadata Manager

本节介绍版本 10.1 中对 Metadata Manager 的更改。

Microsoft SQL Server Integration Services 资源

从版本 10.1 开始，Metadata Manager 会根据使用 SQL Server Integration Services 对象的连接在元数据目录中组织这些对象。元数据目录不包含用于每个包的单独文件夹。要在元数据目录中选择表或列等对象，请通过使用该对象的源或目标连接导航到该对象。

在以前的版本中，Metadata Manager 会按照连接和按照包组织 SQL Server Integration Services 对象。元数据目录除了包含用于每个包的文件夹，还包含一个 Connections 文件夹。

有关 SQL Server Integration Services 资源的详细信息，请参阅《*Informatica 10.1 Metadata Manager 管理员指南*》中的“数据集成资源”一章。

命令程序的证书验证

从版本 10.1 开始，为 Metadata Manager Web 应用程序配置安全连接时，Metadata Manager 命令程序将不接受存在错误的安全证书。控制命令程序是否接受存在错误的安全证书的属性已删除。此功能在 9.6.1 HotFix 4 中也可用，但在 10.0 中不可用。

此前，MMCcmdConfig.properties 文件中的 Security.Authentication.Level 属性控制 mmcmm 或 mmRepoCmd 的证书验证。可以将属性配置为接受所有证书或只接受不存在错误的证书。

由于命令程序不再接受有错误的安全证书，因此 Security.Authentication.Level 属性已过时。该属性不再显示在 mmcmm 或 mmRepoCmd 的 MMCcmdConfig.properties 文件中。

有关 mmcmm 和 mmRepoCmd 的证书验证的详细信息，请参阅《*Informatica 10.1 Metadata Manager 管理员指南*》中的“Metadata Manager 命令程序”一章。

PowerCenter

本节介绍版本 10.1 中对 PowerCenter 的更改。

操作系统配置文件

从版本 10.1 开始，Administrator 工具“安全”页面中的“操作系统配置文件”选项卡被重命名为**操作系统配置文件**选项卡。要创建操作系统配置文件，请转至“安全操作”菜单，并单击**创建操作系统配置文件**。您还可以在创建操作系统配置文件时向用户和组分配默认的操作系统配置文件。在以前的版本中，“安全操作”菜单包含“操作系统配置文件配置”选项。

有关管理操作系统配置文件的详细信息，请参阅《*Informatica 10.1 安全指南*》中的“用户和组”一章。

安全

本节介绍版本 10.1 中对安全的更改。

传输层安全 (TLS)

从版本 10.1 开始，Informatica 使用 TLS v1.1 和 v1.2 对通信进行加密。此外，Informatica 已禁用对 TLS v1.0 和更低版本的支持。

这些更改会影响 Informatica 域内的安全通信、与 Web 应用程序服务的安全连接以及 Informatica 域到外部目标的连接。

此功能在 9.6.1 HotFix 4 中也可用，但在 10.0 中不可用。

权限

从版本 10.1 开始，以下模型存储库对象具有权限更改：

- 应用程序、映射和工作流。向域中的所有用户授予全部权限。
- SQL 数据服务和 Web 服务。向具有有效权限的用户分配直接权限。

这些更改会影响用户和组对这些对象的访问权限级别。

升级后，可能需要查看和更改权限来确保用户对对象具有相应的权限。

有关详细信息，请参阅《*Informatica 10.1 安全指南*》中的“权限”章节。

转换

本节介绍版本 10.1 中已更改的转换行为。

Informatica 转换

本节介绍版本 10.1 中对 Informatica 转换的更改。

地址验证器转换

本节介绍地址验证器转换的更改。

地址验证器转换对地址功能做了以下更新：

地址验证引擎升级

从版本 10.1 开始，地址验证器转换使用 5.8.1 版本的 Informatica Address Verification 软件引擎。该引擎支持 Informatica 在版本 10.1 中向地址验证器转换添加的功能。

以前，该转换使用 5.7.0 版本的 Informatica AddressDoctor 软件引擎。

产品名称更改

Informatica Address Verification 是 Informatica AddressDoctor 的新名称。Informatica AddressDoctor 在版本 5.8.0 中变为 Informatica Address Verification。

对英国地址地理编码选项的更改

从版本 10.1 开始，可以选择“屋顶”地理编码数据属性来检索英国地址的屋顶级地理编码。

以前，要检索英国地址的屋顶级地理编码，需要选择“到达点”地理编码数据属性。

如果升级包括地址验证器转换的存储库，则不需要重新配置转换来指定“屋顶”地理编码属性。如果指定了屋顶级地理编码，但地址验证器转换无法为地址返回地理编码，则转换不会返回任何地理编码数据。

支持在英国输入数据中使用唯一属性参考编号

从版本 10.1 开始，地址验证器转换具有 UPRN GB 输入端口和 UPRN GB 输出端口。

以前，该转换只有 UPRN GB 输出端口。

使用输入端口可检索您输入的唯一属性参考编号对应的英国地址。使用 UPRN GB 输出端口可检索英国地址对应的唯一属性参考编号。

这些功能在 9.6.1 HotFix 4 中也有提供。但在 10.0 中不可用。

有关详细信息，请参阅《*Informatica 10.1 地址验证器端口参考*》。

数据处理器转换

本节介绍对数据处理器转换的更改。

Excel 2013

从版本 10.1 开始，ExcelToXml_03_07_10 文档处理器可以处理 Excel 2013 文件。您可将数据处理器转换中的文档处理器用作预处理器，用于在转换之前转换源文档的格式。

有关详细信息，请参阅《*Informatica 10.1 Data Transformation 用户指南*》。

Avro 或 Parquet 输入的性能提高

数据处理器转换接收来自复杂文件读取器对象的 Avro 或 Parquet 数据输入。从版本 10.1 开始，可以配置复杂文件读取器设置来优化 Avro 或 Parquet 输入的性能。

有关详细信息，请参阅《*Informatica 10.1 Data Transformation 用户指南*》。

Hadoop 环境中 COBOL 输入的性能提高

从版本 10.1 开始，可以配置复杂文件读取器设置来优化处理 Hadoop 环境中大型 COBOL 文件时的性能。使用正则表达式定义如何为相应的 COBOL 输入文件拆分记录处理。

有关详细信息，请参阅《*Informatica 10.1 Data Transformation 用户指南*》。

异常转换

从版本 10.1 开始，可以配置离散记录异常转换和重复记录异常转换以在非默认数据库架构中创建异常表。

在以前的版本中，您将转换配置为在数据库上的默认架构中创建异常表。

此功能在 9.6.1 HotFix 4 中也可用，但在 10.0 中不可用。

有关详细信息，请参阅《*Informatica 10.1 Developer 转换指南*》。

工作流

本节介绍版本 10.1 中已更改的工作流行为。

Informatica 工作流

本节介绍版本 10.1 中对 Informatica 工作流行为的更改。

并行执行人工任务

从版本 10.1 开始，数据集成服务可以针对工作流中的多个序列流并行运行人工任务。要创建并行序列流，请在 Developer Tool 中向工作流添加非独占网关。向两个非独占网关之间的每个序列流添加一个或多个人工任务。

在以前的版本中，可以向两个非独占网关之间的序列流添加一个或多个人工任务。

有关详细信息，请参阅《*Informatica 10.1 Developer 工作流指南*》。

第 38 章

发布任务 (10.1)

本章包括以下主题：

- [Metadata Manager , 382](#)
- [安全性, 383](#)

Metadata Manager

本节介绍版本 10.1 中 Metadata Manager 的发布任务。

Informatica Platform 资源

从版本 10.1 起，要从部署到数据集成服务的 Informatica 10.0 应用程序提取元数据，必须安装版本 10.0 命令行实用程序。将此实用程序安装到 10.1 Metadata Manager 服务可以访问的目录。为使性能达到最佳，请将文件提取到运行 Metadata Manager 服务的计算机上的目录。

配置资源时，还必须在 **10.0 命令行实用程序目录** 属性中输入 10.0 Informatica 命令行实用程序安装目录的文件路径。

有关 Informatica Platform 资源的详细信息，请参阅《*Informatica 10.1 Metadata Manager 管理员指南*》中的“数据集成资源”一章。

验证命令程序的信任库文件

从版本 10.1 开始，为 Metadata Manager Web 应用程序配置安全连接时，Metadata Manager 命令程序将不接受存在错误的安全证书。控制命令程序是否接受存在错误的安全证书的属性已删除。此功能在 9.6.1 HotFix 4 中也可用，但在 10.0 中不可用。

MMCcmdConfig.properties 文件中的 Security.Authentication.Level 属性控制 mmc cmd 或 mmRepoCmd 的证书验证。该属性可以设置为以下值之一：

- NO_AUTH。命令程序接受数字证书，即使证书有错误也是如此。
- FULL_AUTH。命令程序不接受存在错误的安全证书。

NO_AUTH 设置不再有效。命令程序现在只接受不含错误的安全证书。

如果为 Metadata Manager Web 应用程序配置了安全连接，并且之前将 Security.Authentication.Level 属性设置为 NO_AUTH，则必须立即配置信任库文件。要配置 mmc cmd 或 mmRepoCmd 以使用信任库文件，请编辑与 mmc cmd 或 mmRepoCmd 相关联的 MMCcmdConfig.properties 文件。将 TrustStore.Path 属性设置为信任库文件的路径和文件名。

有关 mmcCmd 和 mmRepoCmd 的 MMCmdConfig.properties 文件的详细信息，请参阅《*Informatica 10.1 Metadata Manager 管理员指南*》中的“Metadata Manager 命令行程序”一章。

安全性

本节介绍版本 10.1 中安全功能的发布任务。

权限

升级到 10.1 之后，以下模型存储库对象具有权限更改：

- 应用程序、映射和工作流。向域中的所有用户授予全部权限。
- SQL 数据服务和 Web 服务。向具有有效权限的用户分配直接权限。

这些更改会影响用户和组对这些对象的访问权限级别。

升级后，查看并更改对应用程序、映射、工作流、SQL 数据服务和 Web 服务的权限，以确保用户对此类对象拥有合适的权限。

有关详细信息，请参阅《*Informatica 10.1 安全指南*》中的“权限”章节。