



Informatica®
10.4.1

New Features Guide

This software and documentation are provided only under a separate license agreement containing restrictions on use and disclosure. No part of this document may be reproduced or transmitted in any form, by any means (electronic, photocopying, recording or otherwise) without prior consent of Informatica LLC.

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

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

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

The information in this documentation is subject to change without notice. If you find any problems in this documentation, report them to us at infa_documentation@informatica.com.

Informatica products are warranted according to the terms and conditions of the agreements under which they are provided. INFORMATICA PROVIDES THE INFORMATION IN THIS DOCUMENT "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

Publication Date: 2020-07-12

Table of Contents

Preface	10
Informatica Resources.	10
Informatica Network.	10
Informatica Knowledge Base.	10
Informatica Documentation.	10
Informatica Product Availability Matrices.	11
Informatica Velocity.	11
Informatica Marketplace.	11
Informatica Global Customer Support.	11
 Chapter 1: New Features (10.4.1)	12
Connectivity (10.4.1).	12
DataDirect ODBC and JDBC Drivers.	12
Oracle Connection Manager.	12
Data Engineering Integration.	12
Hierarchical Data on Hive Sources and Targets.	13
SAML Authentication for Mass Ingestion.	13
Data Engineering Streaming.	13
FileName Port for ADLS Gen2.	13
Ingest CDC Data from Multiple Kafka Topics.	13
JDBC V2 Lookup Transformation.	13
Parquet Data Format for Complex Targets.	14
Rollover Parameters in Amazon S3 and ADLS Gen2 Targets.	14
Sources and Targets in Databricks.	14
Streaming Mappings in AWS Databricks.	14
Data Privacy Management.	15
Data Stores.	15
Jobs.	15
Privacy Dashboard.	16
Subject Registry Data Encryption.	16
Unstructured Data Stores.	16
Enterprise Data Catalog.	16
Business Term Association.	16
Configure Conflict Resolution for Data Rule and Column Name Rule.	17
Context Lineage Information.	17
Data Asset Analytics.	17
Data Curation through Export and Import.	17
Data Discovery.	17
Enterprise Data Catalog Walkthroughs.	18
Extracting Transformation Endpoints.	18

Support for Field-Level Lineage for Flat Files.	18
File Lineage for Cloud Storage.	19
Hive Resource.	19
Informatica MDM Resource.	19
Microsoft SQL Server Resource.	20
MicroStrategy Resource.	20
Partitioned File Discovery.	20
Reference Resources and Reference Assets.	20
Resource Level Permissions.	20
REST APIs.	21
SAP BW and SAP BW/4HANA Resources.	21
SSIS Resource.	21
Informatica Transformations.	21
Macro Transformation.	22
Intelligent Structure Model.	22
Change the Root Node in an Array Structure.	22
Metadata Manager.	22
Configure Java Location and Heap Size for Business Object Resources.	22
PowerExchange Adapters.	22
PowerExchange Adapters for Informatica.	23
PowerExchange Adapters for PowerCenter.	24
Chapter 2: New Features (10.4.0).....	27
CI/CD.	27
Data Integration Service REST API.	27
infacmd dis Commands.	28
Reverse Proxy Server.	29
infacmd roh Commands.	29
Application Patch Deployment.	30
Connect to a Run-time Application.	31
Object Explorer View.	31
Tags.	32
Command Line Programs.	32
infacmd isp Commands (New Features 10.4.0).	32
Data Engineering Integration.	32
New Data Types Support	33
AWS Databricks Integration.	33
Cluster Workflows for HDInsight Access to ALDS Gen2 Resources.	34
Databricks Delta Lake Storage Access.	34
Display Nodes Used in Mapping.	34
Log Aggregation	34
Parsing Hierarchical Data on the Spark Engine.	34
Profiles and Sampling Options on the Spark Engine.	35

Python Transformation.	35
Sqoop.	35
Data Engineering Streaming	36
Confluent Schema Registry in Streaming Mappings	36
Data Quality Transformations in Streaming Mappings.	36
Ephemeral Cluster in Streaming Mappings.	36
FileName Port in Amazon S3	37
Microsoft Azure Data Lake Storage Gen2.	37
Streaming Mappings in Azure Databricks.	37
Dynamic Mappings in Data Engineering Streaming	38
Enterprise Data Catalog.	38
Assigning Custom Attributes to Resources and Classes.	38
New Resources.	38
Reference Resources and Reference Assets.	39
Export Assets from Search Results.	39
Lineage and Impact Filters.	39
Asset Control Summary.	40
Rules and Scorecards.	40
Unique Key Inference.	40
Data Domain Discovery on the CLOB File Type.	40
Data Discovery and Sampling Options on the Spark Engine.	40
Tracking Technical Preview.	41
Data Preview and Provisioning.	41
Supported Resource Types for Standalone Scanner Utility.	41
REST APIs.	42
Enterprise Data Preparation.	42
Data Lake Access Management.	43
Microsoft Azure Data Lake Storage as a Data Source.	43
Publish Files to the Data Lake.	43
Upload Files to the Data Lake.	43
Informatica Mappings	44
Binding Mapping Outputs to Mapping Parameters.	44
CLAIRE Recommendations and Insights.	44
Update Mapping Optimizer Level.	45
Informatica Transformations.	45
Address Validator Transformation.	45
Informatica Workflows.	45
Amazon EMR Create Cluster Task Advanced Properties.	45
Informatica Installation.	46
PostgreSQL.	46
Pre-installation (i10Pi) System Check Tool in Silent Mode.	46
Encrypt Passwords in the Silent Installation Properties File.	46

Intelligent Structure Model.	46
Additional Input Types.	46
Create Model from Sample at Design Time.	46
Handling of Unidentified Data.	47
PowerCenter.	47
Connectivity.	47
PowerExchange Adapters.	48
PowerExchange Adapters for Informatica.	48
PowerExchange Adapters for PowerCenter.	50
Security.	52
Configure Web Applications to Use Different SAML Identity Providers.	52
Chapter 3: New Features (10.2.2 HotFix 1).....	53
Command Line Programs	53
infacmd Idm Commands.	53
Enterprise Data Catalog.	53
Azure Data Lake Storage Gen2.	53
Case Insensitive Linking	54
Enterprise Data Catalog Tableau Extension	54
New Resources.	54
Profiling Warehouse Scanner.	54
REST APIs.	54
Search Enhancements	55
Search Tabs.	55
Supported Resource Types for Standalone Scanner Utility.	55
Technical Preview.	56
Chapter 4: New Features (10.2.2 Service Pack 1).....	57
Big Data Management.	57
Sqoop.	57
Big Data Streaming.	58
Amazon S3 Target.	58
TIME_RANGE Function.	58
Writing Data to Multiple HDFS Files.	58
Enterprise Data Catalog.	58
Supported Resource Types for Standalone Scanner Utility.	58
Enterprise Data Preparation.	59
Revert All Inferred Data Types.	59
Prepare Avro and Parquet Files.	59
PowerExchange Adapters for Informatica.	59
PowerExchange for Hive.	59

Chapter 5: New Features (10.2.2)	60
Application Services	60
Mass Ingestion Service	60
Metadata Access Service	60
REST Operations Hub Service	61
Big Data Management	61
Azure Databricks Integration	61
Data Preview on the Spark Engine	62
Hierarchical Data	63
High Availability	63
Intelligent Structure Model	63
Mass Ingestion	64
Monitoring	64
Security	65
Targets	65
Big Data Streaming	65
Azure Event Hubs Data Objects	65
Cross-account IAM Role in Amazon Kinesis Connection	66
Intelligent Structure Model	66
Header Ports for Big Data Streaming Data Objects	66
AWS Credential Profile in Amazon Kinesis Connection	66
Spark Structured Streaming	66
Window Transformation	66
Command Line Programs	67
infacmd dis Commands	67
infacmd ihs Commands	68
infacmd ipc Commands	68
infacmd Idm Commands	68
infacmd mi Commands	69
infacmd ms Commands	70
infacmd oie Commands	70
infacmd tools Commands	70
infasetup Commands	71
Enterprise Data Catalog	71
Automatically Assign Business Title to a Column	71
User Collaboration on Assets	71
Create Enterprise Data Catalog Application Services Using the Installer	72
Custom Metadata Validation Utility	72
Change Notifications	72
Business Glossary Assignment Report	72
Operating System Profiles	73

REST APIs.	73
Source Metadata and Data Profile Filter.	73
Scanner Utility.	73
Resource Types.	73
Enterprise Data Lake.	74
Apply Active Rules.	74
Delete Duplicate Rows.	74
Cluster and Categorize Column Data	74
CLAIRE-based Recommendations.	75
Conditional Aggregation.	75
Data Masking.	75
Localization.	75
Partitioned Sources and Targets.	76
Add Comments to Recipe Steps.	76
Save a Recipe as a Mapping.	76
Amazon S3, ADLS, WASB, MapR-FS as Data Sources.	76
Statistical Functions.	76
Date and Time Functions.	77
Math Functions.	77
Text Functions.	78
Window Functions.	78
Purge Audit Events.	79
Spark Execution Engine.	79
Informatica Developer.	79
Applications.	79
Informatica Mappings.	79
Data Types.	79
Mapping Outputs.	79
Mapping Parameters.	80
Optimizer Levels.	80
Sqoop.	80
Informatica Transformations.	81
Address Validator Transformation.	81
Update Strategy Transformation.	84
PowerExchange Adapters for Informatica.	84
PowerExchange for Amazon Redshift.	84
PowerExchange for Amazon S3.	85
PowerExchange for Google BigQuery.	85
PowerExchange for HBase.	86
PowerExchange for HDFS.	86
PowerExchange for Hive.	87
PowerExchange for MapR-DB.	87

PowerExchange for Microsoft Azure Blob Storage.	87
PowerExchange for Microsoft Azure Cosmos DB SQL API.	88
PowerExchange for Microsoft Azure Data Lake Store.	88
PowerExchange for Microsoft Azure SQL Data Warehouse.	89
PowerExchange for Salesforce.	89
PowerExchange for Snowflake.	90
PowerExchange for Teradata Parallel Transporter API.	90

Preface

Read the *Informatica® New Features and Enhancements Guide* to learn about the new features and enhancements in Informatica products for this release.

Informatica Resources

Informatica provides you with a range of product resources through the Informatica Network and other online portals. Use the resources to get the most from your Informatica products and solutions and to learn from other Informatica users and subject matter experts.

Informatica Network

The Informatica Network is the gateway to many resources, including the Informatica Knowledge Base and Informatica Global Customer Support. To enter the Informatica Network, visit <https://network.informatica.com>.

As an Informatica Network member, you have the following options:

- Search the Knowledge Base for product resources.
- View product availability information.
- Create and review your support cases.
- Find your local Informatica User Group Network and collaborate with your peers.

Informatica Knowledge Base

Use the Informatica Knowledge Base to find product resources such as how-to articles, best practices, video tutorials, and answers to frequently asked questions.

To search the Knowledge Base, visit <https://search.informatica.com>. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at KB_Feedback@informatica.com.

Informatica Documentation

Use the Informatica Documentation Portal to explore an extensive library of documentation for current and recent product releases. To explore the Documentation Portal, visit <https://docs.informatica.com>.

If you have questions, comments, or ideas about the product documentation, contact the Informatica Documentation team at infa_documentation@informatica.com.

Informatica Product Availability Matrices

Product Availability Matrices (PAMs) indicate the versions of the operating systems, databases, and types of data sources and targets that a product release supports. You can browse the Informatica PAMs at <https://network.informatica.com/community/informatica-network/product-availability-matrices>.

Informatica Velocity

Informatica Velocity is a collection of tips and best practices developed by Informatica Professional Services and based on real-world experiences from hundreds of data management projects. Informatica Velocity represents the collective knowledge of Informatica consultants who work with organizations around the world to plan, develop, deploy, and maintain successful data management solutions.

You can find Informatica Velocity resources at <http://velocity.informatica.com>. If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at ips@informatica.com.

Informatica Marketplace

The Informatica Marketplace is a forum where you can find solutions that extend and enhance your Informatica implementations. Leverage any of the hundreds of solutions from Informatica developers and partners on the Marketplace to improve your productivity and speed up time to implementation on your projects. You can find the Informatica Marketplace at <https://marketplace.informatica.com>.

Informatica Global Customer Support

You can contact a Global Support Center by telephone or through the Informatica Network.

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link:

<https://www.informatica.com/services-and-training/customer-success-services/contact-us.html>.

To find online support resources on the Informatica Network, visit <https://network.informatica.com> and select the eSupport option.

CHAPTER 1

New Features (10.4.1)

This chapter includes the following topics:

- [Connectivity \(10.4.1\), 12](#)
- [Data Engineering Integration, 12](#)
- [Data Engineering Streaming, 13](#)
- [Data Privacy Management, 15](#)
- [Enterprise Data Catalog, 16](#)
- [Informatica Transformations, 21](#)
- [Intelligent Structure Model, 22](#)
- [Metadata Manager, 22](#)
- [PowerExchange Adapters, 22](#)

Connectivity (10.4.1)

This section describes new connectivity features in version 10.4.1.

DataDirect ODBC and JDBC Drivers

Effective in version 10.4.1, the DataDirect ODBC and JDBC drivers shipped with Informatica are of the latest available versions.

Oracle Connection Manager

Effective in version 10.4.1, you can use the Oracle Connection Manager to connect to Oracle. Specify the connection string for the Oracle connection in PowerCenter or Developer tool to connect to the Oracle database.

For more information, see the *Informatica 10.4.1 Administrator Guide*.

Data Engineering Integration

This section describes new Data Engineering Integration features in version 10.4.1.

Hierarchical Data on Hive Sources and Targets

Effective in version 10.4.1, you can read from and write to Hive tables that contain hierarchical data in a mapping that runs on the Spark engine.

For more information, see the *Data Engineering Integration 10.4.1 User Guide*.

SAML Authentication for Mass Ingestion

Effective in version 10.4.1, you can configure SAML authentication to log in to the Mass Ingestion tool.

SAML authentication allows you to use a third-party identity provider to manage user credentials and authorization.

For more information, see the following guides:

Data Engineering Integration 10.4.1 Mass Ingestion Guide

Informatica® 10.4.1 Application Service Guide

Informatica® 10.4.1 Security Guide

Data Engineering Streaming

This section describes new Data Engineering Streaming features in version 10.4.1.

FileName Port for ADLS Gen2

Effective in version 10.4.1, when you create a data object write operation for ADLS Gen2, the FileName port appears by default.

At run time, the Data Integration Service creates separate directories for each value in the FileName port and adds the target files within the directories. You can use the file name port in ADLS Gen2 target to ingest CDC data from PWX CDC Publisher.

For more information, see the *Data Engineering Streaming 10.4.1 User Guide*.

Ingest CDC Data from Multiple Kafka Topics

Effective in 10.4.1, you can ingest CDC data from the PWX CDC Publisher from multiple Kafka topics onto Data Engineering systems in one or more mappings.

For more information, see the *Data Engineering Streaming 10.4.1 User Guide*.

JDBC V2 Lookup Transformation

Effective in version 10.4.1, you can use a JDBC data object read operation to look up data in a JDBC V2 table.

You can add a JDBC V2 data object read operation as a lookup in a mapping. You can then configure a lookup condition to look up data from the JDBC V2 table. You can run this mapping on a Databricks engine.

For more information, see the *Data Engineering Streaming 10.4.1 User Guide*.

Parquet Data Format for Complex Targets

Effective in version 10.4.1, you can use the Parquet data format for complex targets.

You can use the Parquet data format for complex targets such as HDFS, ADLS Gen2, and Amazon S3 in the streaming mappings.

For more information, see the *Data Engineering Streaming 10.4.1 User Guide*.

Rollover Parameters in Amazon S3 and ADLS Gen2 Targets

Effective in 10.4.1, you can use different rollover parameters for Amazon S3 and ADLS Gen2 targets to decide the rollover time or size for each target.

For more information, see the *Data Engineering Streaming 10.4.1 User Guide*.

Sources and Targets in Databricks

Effective in version 10.4.1, you can use the Kafka and Confluent Kafka as sources and targets in streaming mappings in a Databricks environment.

You can run the streaming mappings in Databricks environment in both AWS cloud ecosystems and Microsoft Azure cloud services.

For more information, see the *Data Engineering Streaming 10.4.1 User Guide*.

Streaming Mappings in AWS Databricks

Effective in version 10.4.1, you can run streaming mappings in AWS Databricks service in AWS cloud ecosystems.

You can use AWS Databricks to run mappings with the following functionality:

Sources and Targets

You can run streaming mappings against the following sources and targets within the Databricks environment:

- Amazon S3
- Kinesis Streams
- Kinesis Firehose

Transformations

You can add the following transformations to a Databricks streaming mapping in AWS:

- Aggregator
- Expression
- Filter
- Joiner
- Normalizer
- Rank
- Router
- Union

Window

Data Types

AWS Databricks supports the same data types as Azure Databricks.

The following data types are supported:

Array

Bigint

Date/time

Decimal

Double

Integer

Map

Struct

Text

String

For more information, see the following *Data Engineering Streaming 10.4.1 User Guide*.

Data Privacy Management

This section describes new Data Privacy Management features in version 10.4.1.

Data Stores

This section describes new data store properties and sources.

Effective in version 10.4.1, you can scan unstructured data stores with a subject registry remote agent to discover sensitive data. You can configure a scan to read all folders, read specific folders from a list, or read folders based on regular expressions.

Version 10.4.1 supports Google Drive, SAP HANA, and Snowflake data sources.

For more information, see the "Data Store Properties" and "Scans" chapters in the *Informatica Data Privacy Management 10.4.1 User Guide*.

Jobs

This section describes new job types.

Version 10.4.1 includes the following job types:

- The Incremental Scan job scans only the records in a data store that changed since the last scan job ran.
- The Import Catalog Results job imports resources from Enterprise Data Catalog.
- The Sync Catalog Updates job synchronizes imported information from Enterprise Data Catalog with data stores in Data Privacy Management.

For more information, see the "Jobs" chapter in the *Informatica Data Privacy Management 10.4.1 User Guide*.

Privacy Dashboard

Effective in version 10.4.1, Data Privacy Management includes a Privacy Dashboard to view and monitor data in the Subject Registry.

Use the summary indicators and workspace indicators on the dashboard to view information on total subject counts, legal hold counts, subject types, and subject requests. You can track subject requests and take further action on requests from the details pages that you can open on the dashboard.

You can also view information on data stores that contain subjects and the location of data stores that contain subjects.

For more information about the Privacy Dashboard, see the *Informatica Data Privacy Management 10.4.1 User Guide*.

Subject Registry Data Encryption

Effective in version 10.4.1, Data Privacy Management encrypts golden record data identified in subject scans and stores encrypted data in the HBase repository.

The data is encrypted each time a subject scan runs and the data in the Subject Registry repository updates.

Unstructured Data Stores

This section describes changes to the way that Data Privacy Management discovers sensitive and personal data for unstructured data stores.

Version 10.4.1 includes the following new features for unstructured data stores:

- You can specify a global setting for the maximum file size to include in scans of unstructured data stores that use the remote agent scanner.
- You can use the data domains included in the Data Privacy Management installation that match pre-configured classification policies to scan unstructured data stores. For data domains that are not included in the installation, you can configure data domains with regular expressions in a multiplexing proxy agent (MPA).
- You can add a custom match condition in classification policies to match file tags in Google Drive and Microsoft SharePoint data stores.
- You can download scan job reports for unstructured data store scans.

For more information, see the *Informatica Data Privacy Management 10.4.1 User Guide*.

Enterprise Data Catalog

This section describes new Enterprise Data Catalog features in version 10.4.1.

Business Term Association

With effective in version 10.4.1, you can associate business terms to assets from Azure Data Lake Store Gen2 resource in the catalog.

For more information, see the *Enterprise Data Catalog Concepts* chapter in the *Informatica 10.4.1 Catalog Administrator Guide*.

Configure Conflict Resolution for Data Rule and Column Name Rule

Effective in version 10.4.1, you can choose to accept data rule result, column name rule result, or both the data rule and column name rule results if the rules conflict during auto acceptance of data domains.

For more information, see the *Enterprise Data Catalog Concepts* chapter in the *Informatica 10.4.1 Catalog Administrator Guide*.

Context Lineage Information

Effective in version 10.4.1, you can use the **Lineage and Impact** tab to view the context lineage information of the assets. Create a custom resource in Catalog Administrator to extract the context information for assets that include a process definition or mapping execution.

For more information, see the "Ingesting Custom Metadata into the Catalog" chapter in the *Informatica 10.4.1 Catalog Administrator Guide* and see the "View Lineage and Impact" chapter in the *Informatica 10.4.1 Enterprise Data Catalog User Guide*.

Data Asset Analytics

Effective in version 10.4.1, you can use Data Asset Analytics to view analytical insights about the catalog in the form of reports. Analytical insights include information about users configured to access the catalog, assets and resources, asset usage, and enrichment and collaboration details associated with assets.

For more information, see the *Informatica 10.4.1 Enterprise Data Catalog User Guide*

Data Curation through Export and Import

Effective in version 10.4.1, you can accept or reject the multiple inferred assets for data domains. You can export the assets to the comma-separated (CSV) file and import the updated CSV file into the Enterprise Data Catalog.

You can remove the data domains from the assets by deleting the data domains from the accepted or inferred data domain columns in the exported CSV file.

For more information, see the *View Assets* chapter in the *Informatica 10.4.1 Enterprise Data Catalog User Guide*.

Data Discovery

Effective in version 10.4.1, you can run profiles to discover data domains on the following resources and file types:

- Azure Data Lake Store Gen2 resource.
You can run profiles on the structured, unstructured, and extended unstructured file types.
- Cassandra resource.
You can perform data discovery on the Native engine.
- Parquet file for the Amazon S3 and Azure Data Lake Store resources.

For more information, see the *Enterprise Data Catalog Concepts* chapter in the *Informatica 10.4.1 Catalog Administrator Guide*.

Enterprise Data Catalog Walkthroughs

Effective in version 10.4.1, you can access walkthroughs to help you understand and quickly start using Enterprise Data Catalog. Walkthroughs are guided tutorials within Enterprise Data Catalog that explain how you can use a specific functionality.

The following walkthroughs are available in Enterprise Data Catalog:

- Getting Started and Introducing Catalog Homepage
- Introduction to Search Results
- Introduction to Table Overview
- Introduction to Lineage
- Introduction to Column Overview
- Enhance Asset Credibility

To launch a walkthrough, click the ? help icon in the toolbar, and then select and click the Walkthrough link.

Extracting Transformation Endpoints

Effective in version 10.4.1, you can extract the transformation endpoints for a PowerCenter resource that has web service transformation mappings. In Enterprise Data Catalog, the transformation endpoints appear as connections, which links the data sources or targets.

Support for Field-Level Lineage for Flat Files

Effective in version 10.4.1, you can view the field-level lineage for a source or target flat file in the Catalog.

You can configure the following resources to view the field-level lineage for a source or target flat file:

- PowerCenter
- SQL Server Integration Service
- Informatica Cloud Service
- IBM InfoSphere DataStage
- Informatica Platform

To view the field-level lineage for flat files without headers, you must select the **Enable Reference Resources**, and **Retain Unresolved Reference Assets** properties during resource configuration. You do not require these properties to view the field-level lineage for flat files with headers.

The following table includes the file systems supported by the resources:

Resources	Supported File Systems
PowerCenter	Linux Windows
SQL Server Integration Service	Windows Azure Data Lake Store Gen2

Resources	Supported File Systems
Informatica Cloud Service	Azure Data Lake Store Gen2 Azure Data Lake Store Gen1 Amazon S3 Azure Blob Storage Google Cloud Storage Linux Windows
IBM InfoSphere DataStage	Linux Windows
Informatica Platform	Linux Windows

File Lineage for Cloud Storage

Effective in version 10.4.1, you can use the Informatica Cloud Service resource to view the file lineage for Cloud Storage.

You can view the file lineage for the Informatica Cloud Service mappings that use data sources such as Amazon S3, Microsoft Azure Blob Storage, and Azure Data Lake Store Gen2.

Hive Resource

Effective in version 10.4.1, the Hive resource includes the following enhancements:

Extract connection metadata

You can use the Hive resource to extract connection details for views of different schemas.

Automatically assign connections

When you create the resource, you can choose to automatically assign the database schemas to the Hive resource. You can view the list of automatically assigned schemas and their connections for the resource. You can assign or unassign schemas in the auto-assigned connections.

For more information, see the *Informatica 10.4.1 Enterprise Data Catalog Scanner Configuration Guide*.

Informatica MDM Resource

Effective in version 10.4.1, the Informatica MDM resource extracts metadata from the MDM data source through APIs.

The Informatica MDM resource includes the following enhancements:

Metadata extraction

You can configure the Informatica MDM resource to extract metadata such as base objects, landing, and staging tables along with the field and attribute information.

Lineage enhancements

You can view the lineage between landing tables up to business entities. You can also view detailed lineage from the Informatica Platform applications.

For more information, see the *Informatica 10.4.1 Enterprise Data Catalog Scanner Guide*.

Microsoft SQL Server Resource

Effective in version 10.4.1, when you create the Microsoft SQL Server resource, you can choose to automatically assign the database schemas to the resource. You can view the list of automatically assigned schemas and their connections for the resource. You can assign or unassign schemas in the auto-assigned connections.

For more information, see the *Informatica 10.4.1 Enterprise Data Catalog Scanner Configuration Guide*.

MicroStrategy Resource

Effective in version 10.4.1, the MicroStrategy resource includes the following enhancements:

View lineage at the report level

You can view the lineage at the report level for a MicroStrategy resource. The report level lineage does not include containers such as page header, page footer, page body, and grouping elements such as columns, rows, and metrics.

Support for incremental loading

You can enable incremental loading for a MicroStrategy resource. An incremental load causes the data source to load recent changes to the metadata instead of loading complete metadata. Incremental loading reduces the amount of time it takes to load the resource.

For more information, see the *Informatica 10.4.1 Enterprise Data Catalog Scanner Configuration Guide*.

Partitioned File Discovery

Effective in version 10.4.1, you can use the Amazon S3 and Azure Data Lake Store resources to identify and publish horizontally partitioned files under the same directory, and files organized in hierarchical Hive-style directory structures as a single partitioned file.

Reference Resources and Reference Assets

Effective in version 10.4.1, you can configure the following resources to extract metadata about data sources or other resources in the catalog referenced by the resource:

- Hive
- Microsoft SQL Server

For more information, see the *Informatica 10.4.1 Catalog Administrator Guide* and the *Informatica 10.4.1 Enterprise Data Catalog User Guide*.

Resource Level Permissions

Effective in version 10.4.1, you can assign configuration permissions and apply restrictions on resources for users or user groups.

Note: You can also apply restrictions on resources during resource creation.

For more information, see the *Informatica 10.4.1 Catalog Administrator Guide*.

REST APIs

Effective in version 10.4.1, you can use the following Informatica Enterprise Data Catalog REST APIs:

- **Catalog Events REST APIs.** In addition to the existing REST APIs, you can list the class types of events for the objects that the user subscribed.
- **Lineage Filter REST APIs.** In addition to the existing REST APIs, you can create a default lineage filter.
- **Monitoring Info REST APIs.** In addition to the existing REST APIs, you can perform the following tasks:
 - List the logs for a job.
 - Submit and list backup jobs, resource scan log jobs, and service log jobs.
 - Download a specific log for a job and ZIP files for a backup job, resource scan log job, and service log job.

For more information about the REST APIs, see the *Informatica 10.4.1 Enterprise Data Catalog REST API Reference*.

SAP BW and SAP BW/4HANA Resources

Effective in version 10.4.1, the SAP BW and SAP BW/4HANA resources extract metadata from the InfoObject as InfoProvider asset. You can view the lineage and the relationship information for the InfoObject as InfoProvider asset in the Catalog.

For more information, see the *Informatica 10.4.1 Enterprise Data Catalog Scanner Configuration Guide*.

SSIS Resource

Effective in version 10.4.1, the SSIS resource includes the following enhancements:

View detailed lineage for transformations

You can view the detailed lineage for transformations such as aggregate, audit, and character map in the Catalog.

Metadata extraction from SSIS database

You can configure the SSIS resource to extract metadata from the SSIS database.

Field level and control lineage support

You can view the field-level lineage for flat files, and the control summary of SSIS assets in the Catalog.

For more information, see the *Informatica 10.4.1 Enterprise Data Catalog Scanner Guide*.

Informatica Transformations

This section describes new Informatica transformation features in version 10.4.1.

Macro Transformation

Effective in version 10.4.1, you can use the Macro transformation in Informatica mappings. The Macro transformation adds dynamic functionality to transformation logic contained in a mapplet.

Before you use a Macro transformation, create a mapplet with the transformation logic that you want to accept dynamic input. Then add the Macro transformation to a dynamic mapping and use the mapplet you created as the macro instruction.

For example, you want to read data from a dynamic source, perform data encryption, and then write the encrypted data to a target. First create a mapplet that performs the encryption using a Data Masking transformation. Then create a mapping that reads from and writes to your data sources. Add a Macro transformation to the dynamic mapping and select the Data Masking mapplet as the macro instruction.

For more information, refer to the "Macro Transformation" chapter in the *Informatica 10.4.1 Developer Transformation Guide*.

Intelligent Structure Model

This section describes new intelligent structure model features in version 10.4.1.

Change the Root Node in an Array Structure

Effective in version 10.4.1, you can determine where Intelligent Structure Discovery flattens a structure by selecting which element in the structure is the root node. For example, if an intelligent structure model in a midstream mapping contains complex hierarchical inputs, select the top-level node as the root to avoid data loss.

For more information, refer to the *Data Engineering Integration 10.4.1 User Guide*.

Metadata Manager

This section describes new Metadata Manager features in version 10.4.1.

Configure Java Location and Heap Size for Business Object Resources

Effective in version 10.4.1, you can configure the Java location and heap size in the Miscellaneous property for Business Object resources. You can configure these options based on the Java SE Runtime Environment.

For more information, see the "Business Intelligence Resources" chapter in the *Informatica 10.4.1 Metadata Manager Administrator Guide*.

PowerExchange Adapters

This section describes new PowerExchange adapter features in version 10.4.1.

PowerExchange Adapters for Informatica

This section describes new PowerExchange adapter features in version 10.4.1.

PowerExchange for Amazon Redshift

Effective in version 10.4.1, PowerExchange for Amazon Redshift includes the following features:

- The Amazon S3 bucket that you specify to create staging files can be in a different region than the Amazon Redshift cluster.
- You can use KMS Customer managed keys from an external account to create encrypted resources by providing the Amazon Resource Name (ARN) of the external account in the connection properties.

For more information, see the *Informatica 10.4.1 PowerExchange for Amazon Redshift User Guide*.

PowerExchange for Amazon S3

Effective in version 10.4.1, PowerExchange for Amazon S3 includes the following features:

- You can read multiple files of the flat format type from Amazon S3 and write data to a target using a `.manifest` file.
- You can use the AWS ap-east-1 (Hong Kong) region in the native environment and on the Spark engine.
- You can use KMS Customer managed keys from an external account to create encrypted resources by providing the Amazon Resource Name (ARN) of the external account in the connection properties.
- You can run a mapping to read an Amazon S3 binary file using the `FileName` port in native mode.
- You can read an Amazon S3 binary file from a directory in native mode.

For more information, see the *Informatica 10.4.1 PowerExchange for Amazon S3 User Guide*.

PowerExchange for Google BigQuery

Effective in version 10.4.1, PowerExchange for Google BigQuery includes the following features:

- You can configure a cached lookup operation on a Google BigQuery table.
- You can read data from and write data to a table in a Google BigQuery dataset available in a specific region.

For more information, see the *Informatica 10.4.0 PowerExchange for Google BigQuery User Guide*.

PowerExchange for Hive

Effective in version 10.4.1, you can use complex data types to read and write hierarchical data from Hive tables in a mapping that runs on the Spark engine.

When you read from and write to Hive tables with hierarchical data, you can perform data preview and schema synchronization on the mapping.

You can use Hive tables with hierarchical data in HDP 3.1 distribution only.

For more information, see the *Informatica 10.4.1 PowerExchange for Hive User Guide*.

PowerExchange for Microsoft Azure Data Lake Storage Gen2

Effective in version 10.4.1, PowerExchange for Microsoft Azure Data Lake Storage Gen2 includes the following features:

- The following run-time engines can read and write ORC files:
 - Data Integration Service
 - Spark
 - Databricks Spark
- You can run mappings to read and write JSON files in the native environment.
- You can configure the Azure Government endpoints in mappings in the native environment and on Spark engine.
- You can configure the authenticated proxy server settings for the Data Integration Service to connect to Microsoft Azure Data Lake Storage Gen2.

For more information, see the *Informatica 10.4.1 PowerExchange for Microsoft Azure Data Lake Storage Gen2 User Guide*.

PowerExchange for Salesforce

Effective in version 10.4.1, PowerExchange for Salesforce includes the following features:

- You can use version 48.0 of Salesforce API to create a Salesforce connection and access Salesforce objects.
- You can configure dynamic mappings to include the frequent changes to sources, targets, and transformation logic at run time based on parameters and rules that you define.

For more information, see the *Informatica 10.4.1 PowerExchange for Salesforce User Guide*.

PowerExchange for SAP NetWeaver

Effective in version 10.4.1, when you use the HTTP streaming data transfer mode in SAP Table Reader, you can specify a transfer packet size in MB to read data from SAP tables.

After the upgrade, existing mappings that use the HTTP streaming data transfer mode in SAP Table Reader might show a performance improvement. To further tune the performance, specify an appropriate transfer packet size based on your requirement.

For more information, see the *Informatica 10.4.1 PowerExchange for SAP NetWeaver User Guide*.

PowerExchange for Snowflake

Effective in version 10.4.1, you can use external tables and materialized views as Snowflake sources and perform all the read operations.

For more information, see the *Informatica 10.4.1 PowerExchange for Snowflake User Guide*.

PowerExchange Adapters for PowerCenter

This section describes new PowerExchange adapter features in version 10.4.1.

PowerExchange for Db2 Warehouse

Effective in version 10.4.1, PowerExchange for Db2 Warehouse includes the following features:

- You can read and write data for the VarcharForBitData and CharForBitData data types in Db2 Warehouse.
- You can use PowerExchange for Db2 Warehouse on the Windows and AIX platforms.

For more information, see the *Informatica PowerExchange for Db2 Warehouse 10.4.1 User Guide for PowerCenter*.

PowerExchange for Greenplum

Effective in version 10.4.1, you can use PowerExchange for Greenplum on the Windows platform.

For more information, see the *Informatica 10.4.1 PowerExchange for Greenplum User Guide for PowerCenter*.

PowerExchange for Google BigQuery

Effective in version 10.4.1, PowerExchange for Google BigQuery includes the following features:

- You can create a pipeline Lookup transformation to perform a lookup on a Google BigQuery table.
- You can create a session to read the real-time or changed data from a Change Data Capture (CDC) source and load data into Google BigQuery.
- You can read data from and write data to a table in a Google BigQuery dataset available in a specific region.

For more information, see the *Informatica 10.4.1 PowerExchange for Google BigQuery User Guide for PowerCenter*.

PowerExchange for Google Cloud Storage

Effective in version 10.4.1, when you import a Google Cloud Storage target definition, you can remove the header row in a Google Cloud Storage flat file.

For more information, see the *Informatica 10.4.1 PowerExchange for Google Cloud Storage User Guide for PowerCenter*.

PowerExchange for Kafka

Effective in version 10.4.1, PowerExchange for Kafka includes the following features:

- When you write data to a Kafka target, you can configure file-based recovery in a real-time session.
- You can configure additional configuration properties to connect to a Kafka broker over SSL.
- When you specify configuration properties in Additional Security Properties field, the value that you specify is masked.
- You can configure PLAIN security for a Kafka broker.

For more information, see the *Informatica PowerExchange for Kafka 10.4.1 User Guide for PowerCenter*.

PowerExchange for Microsoft Azure SQL Data Warehouse V3

Effective in version 10.4.1, you can read data from or write data to a Microsoft Azure SQL Data Warehouse endpoint that resides in a virtual network (VNet).

For more information, see the *Informatica 10.4.1 PowerExchange for Microsoft Azure SQL Data Warehouse V3 User Guide for PowerCenter*.

PowerExchange for Salesforce

Effective in version 10.4.1, you can use version 48.0 of Salesforce API to create a Salesforce connection and access Salesforce objects.

For more information, see the *Informatica 10.4.1 PowerExchange for Salesforce User Guide for PowerCenter*.

PowerExchange for SAP NetWeaver

Effective in version 10.4.1, when you use the HTTP streaming data transfer mode in PowerExchange for SAP Dynamic ABAP Table Extractor, you can specify a transfer packet size in MB to optimize the performance of reading data from SAP tables.

For more information, see the *Informatica 10.4.1 PowerExchange for SAP NetWeaver User Guide for PowerCenter*.

PowerExchange for Snowflake

Effective in version 10.4.1, PowerExchange for Snowflake includes the following features:

- You can use external tables and materialized views as Snowflake sources and perform all the read operations.
- You can override the default update query that the PowerCenter Integration Service generates with the update query that you specify.

For more information, see the *Informatica 10.4.1 PowerExchange for Snowflake User Guide for PowerCenter*.

CHAPTER 2

New Features (10.4.0)

This chapter includes the following topics:

- [CI/CD, 27](#)
- [Command Line Programs, 32](#)
- [Data Engineering Integration, 32](#)
- [Data Engineering Streaming , 36](#)
- [Enterprise Data Catalog, 38](#)
- [Enterprise Data Preparation, 42](#)
- [Informatica Mappings , 44](#)
- [Informatica Transformations, 45](#)
- [Informatica Workflows, 45](#)
- [Informatica Installation, 46](#)
- [Intelligent Structure Model, 46](#)
- [PowerCenter, 47](#)
- [PowerExchange Adapters, 48](#)
- [Security, 52](#)

CI/CD

This section describes enhancements to CI/CD in version 10.4.0.

CI/CD, or continuous integration and continuous delivery, is a practice that automates the integration and delivery operations in a CI/CD pipeline. In version 10.4.0, you can incorporate the enhancements into your CI/CD pipeline to improve how you deploy, test, and deliver objects to the production environment.

Data Integration Service REST API

Effective in version 10.4.0, you can use the Data Integration Service REST API to automate tasks in a CI/CD pipeline.

Some of the tasks that the REST API can automate include the following tasks:

Query objects.

Query objects, including design-time objects in a Model repository and run-time objects that are deployed to a Data Integration Service.

You can pass the query to other REST API requests. For example, you can pass a query to a version control operation to perform version control on a specific set of objects. You can also pass a query to deploy specific design-time objects to an application patch archive file.

Perform version control operations.

Perform version control operations to check in, check out, undo a check out, or reassign a checked-out design-time object to another developer.

Manage tags.

Manage the tags that are assigned to design-time objects. You can assign a new tag or replace tags for an object. You can also untag an object.

Update applications.

Deploy design-time objects to an application patch archive file and deploy the file to a Data Integration Service to update a deployed incremental application.

Manage applications.

Start, undeploy, or stop an application based on project or organizational requirements.

Run deployed mappings.

Run a deployed mapping to test application output.

Compare mappings.

Compare two mappings in the same domain.

For example, you can compare two design-time mappings or a design-time mapping to a run-time mapping.

To view the REST API requests that you can use and the parameters for each request, access the REST API documentation through the Data Integration Service process properties or the REST Operations Hub Service properties in the Administrator tool.

Compared to the `infacmd` command line programs, the REST API does not have any setup requirements and you can run the REST API in environments that do not have Informatica services installed on the client machine.

For information about the REST API, see the "Data Integration Service REST API" chapter in the *Informatica 10.4.0 Application Service Guide*.

infacmd dis Commands

Effective in version 10.4.0, you can use the `infacmd dis` commands to perform operations on design-time objects, run-time objects in an application, and deployed mappings. You can also use the `infacmd dis` commands to run Data Integration Service utilities.

The following table describes the new `infacmd dis` commands:

Command	Description
<code>compareMapping</code>	Compares two queried mappings. Query the mappings to compare mapping properties, transformation properties, and ports within transformations. To query design-time mappings, specify the design-time Model repository. To query run-time mappings, do not specify a Model repository. The query will use the Data Integration Service that you specify to run the command.
<code>deployObjectsToFile</code>	Deploys design-time objects to an application patch archive file.

Command	Description
queryDesignTimeObjects	Queries design-time objects from a Model Repository service.
queryRunTimeObjects	Queries run-time objects that are deployed to a Data Integration Service and returns a list of objects.
replaceAllTag	Replaces tags with the specified tags on the queried objects in Model Repository service.
tag	Assigns a tag on the queried objects in Model Repository service.
untag	Removes a tag on the queried objects in Model Repository service.
listPatchNames	Lists all patches that have been applied to an incremental application.

For more information, see the "infacmd dis Command Reference" chapter in the *Informatica® 10.4.0 Command Reference*.

Reverse Proxy Server

Effective in version 10.4.0, you can use the REST Operations Hub to launch the reverse proxy server that performs the load balancing for the Data Integration Service API requests across nodes in a Data Integration Service grid.

The REST Operations Hub is grid enabled.

For information about the reverse proxy server, see the "System Services" chapter in the Informatica 10.4.0 Application Service Guide.

infacmd roh Commands

Effective in version 10.4.0, you can use the following infacmd roh commands to update REST Operations Hub Service process.

The following table describes the new infacmd roh commands:

Commands	Description
listReverseProxyServerOptions	Lists the reverse proxy server properties.
listServiceProcessOptions	Lists the REST Operations Hub Service process properties.
updateServiceOptions	Updates the REST Operations Hub Service properties.
updateReverseProxyServerOptions	Updates the reverse proxy server properties in the domain.

Effective in version 10.4.0, the following infacmd roh commands are renamed:

- listROHProperties to listProcessProperties.
- updateROHService to updateServiceProcessOptions.

Note: Update any scripts that you use with the previous command name.

For more information, see the "infacmd roh Command Reference" chapter in the *Informatica 10.4.0 Command Reference*.

Application Patch Deployment

Effective in version 10.4.0, application patch deployment has new functionality in the following areas:

Incremental Deployment Wizard

Effective in version 10.4.0, you can perform the following tasks in the Incremental Deployment wizard:

- Optionally, you can enter a patch description.
- Optionally, you can choose to retain or discard state information.
State information refers to mapping properties and the properties of run-time objects such as mapping outputs or the Sequence Generator transformation.

For more information about the Incremental Deployment wizard, see the "Application Patch Deployment" chapter in the *Informatica 10.4.0 Developer Tool Guide*.

For more information about state information, see the "Application Deployment" chapter in the *Informatica 10.4.0 Developer Tool Guide*.

Patch History

Effective in version 10.4.0, the patch history in the Incremental Deployment wizard shows both the patch name and the patch description of the patches that were deployed to update the incremental application. The time that the patch was created is appended to the beginning of the patch description.

Additionally, you can use the Administrator tool to view the patch history for a deployed incremental application.

For more information about the patch history, see the "Application Patch Deployment" chapter in the *Informatica 10.4.0 Developer Tool Guide*.

For more information about deployed applications, see the "Data Integration Service Applications" chapter in the *Informatica 10.4.0 Application Service Guide*.

Application Patch Archive Files

Effective in version 10.4.0, application patch archive files additionally store the patch description and indicate whether state information is retained.

For more information about patch history, see the "Application Patch Deployment" chapter in the *Informatica 10.4.0 Developer Tool Guide*.

infacmd tools Commands

Effective in version 10.4.0, the infacmd tools patchApplication command has the following new option:

Option	Argument	Description
-RetainStateInformation -rsi	True False	Optional. Indicates whether state information is retained or discarded. State information refers to mapping properties and the properties of run-time objects such as mapping outputs or the Sequence Generator transformation.

For more information, see the "infacmd tools Command Reference" chapter in the *Informatica 10.4.0 Command Reference*.

Connect to a Run-time Application

Effective in version 10.4.0, you can connect to a run-time application in the Developer tool. After you connect to a run-time application, you can expand the application in the **Object Explorer** view and open read-only copies of the run-time objects in the **Editor**.

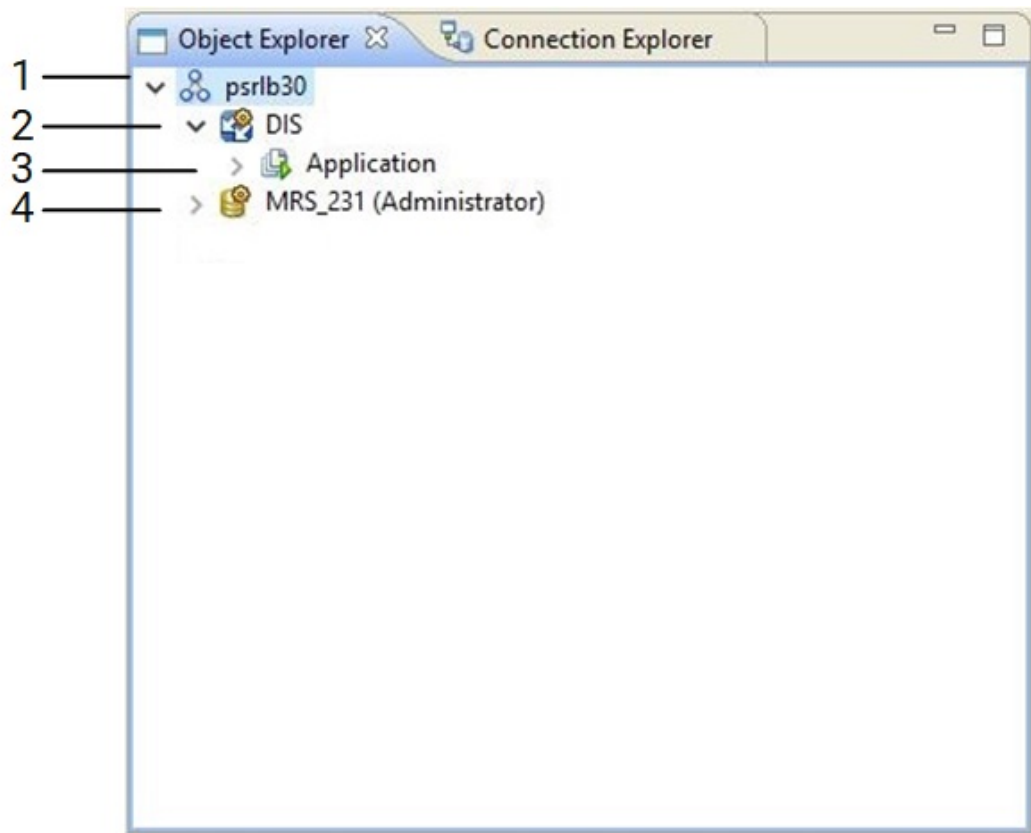
After you connect to a run-time application, the searches that you perform in the Developer tool can locate run-time objects in the application.

For more information about connecting to a run-time application and viewing the run-time objects, see the "Application Deployment" chapter in the *Informatica 10.4.0 Developer Tool Guide*.

Object Explorer View

Effective in version 10.4.0, the **Object Explorer** view in the Developer tool shows the domain after you connect to a Model repository or to a run-time application. You can expand the domain to view the design-time objects in the Model repository or the run-time objects in the run-time application.

The following image shows the **Object Explorer** view:



1. Domain
2. Data Integration Service
3. Run-time application
4. Model repository

For more information about the user interface in the Developer tool, see the "Informatica Developer" chapter in the *Informatica 10.4.0 Developer Tool Guide*.

Tags

Effective in version 10.4.0, tags have the following functionality:

- When you deploy a mapping that is associated with a tag, the tag is propagated to the run-time version of the mapping on the Data Integration Service.
- If you update the deployed mapping using an application patch, the name of the patch is associated as a tag with the run-time version of the mapping.

For more information about tags, see the "Informatica Developer" chapter in the *Informatica 10.4.0 Developer Tool Guide*.

Command Line Programs

This section describes new commands in version 10.4.0.

infacmd isp Commands (New Features 10.4.0)

The following table describes new infacmd isp commands:

Command	Description
addCustomLDAPType	Adds a custom LDAP type defining an LDAP directory service.
listAllCustomLDAPTypes	Lists the configuration information for all custom LDAP types used by the specified domain.
listAllLDAPConnectivity	Lists the configuration information for all LDAP configurations used by the specified domain.
listCustomLDAPType	Lists the configuration information for a custom LDAP type.
removeCustomLDAPType	Removes the specified custom LDAP type from the specified domain.
removeLDAPConnectivity	Removes the specified LDAP configuration from the specified domain.
updateCustomLDAPType	Updates the specified custom LDAP type.
updateLDAPConnectivity	Updates the specified LDAP configuration.

For more information, see the *Informatica 10.4.0 Command Reference*.

Data Engineering Integration

This section describes new Data Engineering Integration features in version 10.4.0.

New Data Types Support

Effective in version 10.4.0, you can use the following new data types for complex files:

- When you run a mapping that reads or writes to Avro and Parquet complex file objects in the native environment or on the Hadoop environment, you can use the following data types:
 - Date
 - Decimal
 - Timestamp
- You can use Time data type to read and write Avro or Parquet complex file objects in the native environment or on the Blaze engine.
- You can use Date, Time, Timestamp, and Decimal data types are applicable when you run a mapping on the Databricks Spark engine.

The new data types are applicable to the following adapters:

- PowerExchange for HDFS
- PowerExchange for Amazon S3
- PowerExchange for Google Cloud Storage
- PowerExchange for Microsoft Azure Blob Storage
- PowerExchange for Microsoft Azure Data Lake Storage Gen1
- PowerExchange for Microsoft Azure Data Lake Storage Gen2

For more information about data types, see the "Data Type Reference" chapter in the *Data Engineering Integration 10.4.0 User Guide*.

AWS Databricks Integration

Effective in version 10.4.0, you can integrate the Informatica domain with Databricks on AWS.

You can use AWS Databricks to run mappings with the following functionality:

- You can run mappings against Amazon Simple Storage Service (S3) and Amazon Redshift sources and targets within the Databricks environment.
- You can develop cluster workflows to create ephemeral clusters using Databricks on AWS.
- You can add the Python transformation to a mapping configured to run on the Databricks Spark engine. The Python transformation is supported for technical preview only.

AWS Databricks supports the same data types as Azure Databricks.

For more information, refer to the following guides:

Data Engineering 10.4.0 Integration Guide

Data Engineering 10.4.0 Administrator Guide

Data Engineering Integration 10.4.0 User Guide

Informatica 10.4.0 Developer Workflow Guide

Cluster Workflows for HDInsight Access to ALDS Gen2 Resources

Effective in version 10.4.0, you can create a cluster workflow to run on an Azure HDInsight cluster to access ADLS Gen2 resources.

For more information about cluster workflows, see the *Informatica Data Engineering Integration 10.4.0 User Guide*.

Databricks Delta Lake Storage Access

Effective in version 10.4.0, you can access Databricks Delta Lake storage as sources and targets.

Mappings can access Delta Lake resources on the AWS and Azure platforms.

For information about configuring access to Delta Lake tables, see the *Data Engineering Integration Guide*.

For information about creating mappings to access Delta Lake tables, see the *Data Engineering Integration User Guide*.

Display Nodes Used in Mapping

Effective in version 10.4.0, you can view the maximum number of cluster nodes used by a mapping in a given time duration.

You can use the REST Operations Hub API `ClusterStats(startTimeInmillis=[value], endTimeInmillis=[value])` to view the maximum number of Hadoop nodes for a cluster configuration used by a mapping in a given time duration.

For more information about the REST API, see the "Monitoring REST API Reference" chapter in the *Data Engineering 10.4.0 Administrator Guide*

Log Aggregation

Effective in version 10.4.0, you can get aggregated logs for deployed mappings that run in the Hadoop environment.

You can collect the aggregated cluster logs for a mapping based on the job ID in the Monitoring tool or by using the `infacmd ms fetchAggregatedClusterLogs` command. You can get a .zip or tar.gz file of the aggregated cluster logs for a mapping based on the job ID and write the compressed aggregated log file to a target directory.

For more information, see the *Informatica 10.4.0 Administrator Guide*.

Parsing Hierarchical Data on the Spark Engine

Effective in 10.4.0, you can use complex functions to parse up to 5 MB of data midstream in a mapping.

The Spark engine can parse raw string source data using the following complex functions:

- PARSE_JSON
- PARSE_XML

The complex functions parse JSON or XML data in the source string and generate struct target data.

For more information, see the "Hierarchical Data Processing" chapter in the *Informatica Data Engineering Integration 10.4.0 User Guide*.

For more information about the complex functions, see the "Functions" chapter in the *Informatica 10.4.0 Developer Transformation Language Reference*.

Profiles and Sampling Options on the Spark Engine

Effective in version 10.4.0, you can run profiles and choose sampling options on the Spark engine.

Profiling on the Spark engine

You can create and run profiles on the Spark engine in the Informatica Developer and Informatica Analyst tools. You can perform data domain discovery and create scorecards on the Spark engine.

Sampling options on the Spark engine

You can choose following sampling options to run profiles on the Spark engine:

- **Limit n** sampling option runs a profile based on the number of the rows in the data object. When you choose to run a profile in the Hadoop environment, the Spark engine collects samples from multiple partitions of the data object and pushes the samples to a single node to compute sample size. You can not apply limit n sample options on the profiles with advance filter. Supported on Oracle database through Sqoop connection.
- **Random percentage** sampling option runs a profile on a percentage of rows in the data object.

For information about the profiles and sampling options on the Spark engine, see *Informatica 10.4.0 Data Discovery Guide*.

Python Transformation

Effective in version 10.4.0, the Python transformation has the following functionality:

Active Mode

You can create an active Python transformation. As an active transformation, the Python transformation can change the number of rows that pass through it. For example, the Python transformation can generate multiple output rows from a single input row or the transformation can generate a single output row from multiple input rows.

For more information, see the "Python Transformation" chapter in the *Informatica Data Engineering Integration 10.4.0 User Guide*.

Partitioned Data

You can run Python code to process incoming data based on the data's default partitioning scheme, or you can repartition the data before the Python code runs. To repartition the data before the Python code runs, select one or more input ports as a partition key.

For more information, see the "Python Transformation" chapter in the *Informatica Data Engineering Integration 10.4.0 User Guide*.

Sqoop

Effective in version 10.4.0, you can configure the following Sqoop arguments in the JDBC connection:

- --update-key
- --update-mode
- --validate
- --validation-failurehandler
- --validation-threshold
- --validator
- --mapreduce-job-name

- --bindir
- --class-name
- --jar-file
- --outdir
- --package-name

For more information about configuring these Sqoop arguments, see the Sqoop documentation.

Data Engineering Streaming

This section describes new Data Engineering Streaming features in version 10.4.0.

Confluent Schema Registry in Streaming Mappings

Effective in version 10.4.0, you can use Confluent Kafka as sources and targets in streaming mappings using schema registry.

You can use Confluent Kafka to store and retrieve Apache Avro schemas in streaming mappings. Schema registry uses Kafka as its underlying storage mechanism.

For more information, see the *Data Engineering Streaming 10.4.0 User Guide*.

Data Quality Transformations in Streaming Mappings

Effective in version 10.4.0, you can use data quality transformations in streaming mappings.

You can use the following data quality transformations in streaming mappings to apply the data quality process on the streaming data:

- Address Validator transformation
- Classifier transformation
- Parser transformation
- Standardizer transformation

For more information, see the *Data Engineering Streaming 10.4.0 User Guide*.

Ephemeral Cluster in Streaming Mappings

Effective in version 10.4.0, you can run a workflow to create an ephemeral cluster that runs mapping and other tasks on a cloud platform cluster.

To resume data process from the point in which a cluster is deleted, you can run streaming mappings on ephemeral cluster by specifying an external storage and a checkpoint directory.

For more information, see the *Data Engineering Streaming 10.4.0 User Guide*.

FileName Port in Amazon S3

Effective in version 10.4.0, when you create a data object write operation for Amazon S3 files, the FileName port appears by default.

At run time, the Data Integration Service creates separate directories for each value in the FileName port and adds the target files within the directories.

For more information, see the *Data Engineering Streaming 10.4.0 User Guide*.

Microsoft Azure Data Lake Storage Gen2

Effective in version 10.4.0, you can use Microsoft Azure Data Lake Storage Gen2 as a target in streaming mappings.

Azure Data Lake Storage Gen2 is built on Azure Blob Storage. Azure Data Lake Storage Gen2 has the capabilities of both Azure Data Lake Storage Gen1 and Azure Blob Storage. You can use Azure Databricks version 5.4 or Azure HDInsight version 4.0 to access the data stored in Azure Data Lake Storage Gen2.

For more information, see the *Data Engineering Streaming 10.4.0 User Guide*.

Streaming Mappings in Azure Databricks

Effective in version 10.4.0, you can run streaming mappings in Azure Databricks service in Microsoft Azure cloud services.

Sources and Targets

You can run streaming mappings against the following sources and targets within the Databricks environment:

- Microsoft Azure Event Hubs
- Azure Data Lake Storage Gen2 (ADLS Gen2)

Transformations

You can add the following transformations to a Databricks streaming mapping:

- Aggregator
- Expression
- Filter
- Joiner
- Normalizer
- Rank
- Router
- Union
- Window

Data Types

The following data types are supported:

- Array
- Bigint
- Date/time

Decimal
Double
Integer
Map
Struct
Text
String

Workflows

You can develop cluster workflows to create ephemeral clusters in the Databricks environment. Use Azure Data Lake Storage Gen1 (ADLS Gen1) and Azure Data Lake Storage Gen2 (ADLS Gen2) to create ephemeral clusters in the Databricks environment.

For more information about streaming mappings in Azure Databricks, see the *Data Engineering Streaming 10.4.0 User Guide*.

Dynamic Mappings in Data Engineering Streaming

Effective in version 10.4.0, dynamic mapping support in Data Engineering Streaming is available for technical preview.

You can use Confluent Kafka data objects as dynamic sources and targets in a streaming mapping.

Technical preview functionality is supported for evaluation purposes but is unwarranted and is not production-ready. Informatica recommends that you use in non-production environments only. Informatica intends to include the preview functionality in an upcoming release for production use, but might choose not to in accordance with changing market or technical circumstances. For more information, contact Informatica Global Customer Support.

Enterprise Data Catalog

This section describes the new Enterprise Data Catalog features in version 10.4.0.

Assigning Custom Attributes to Resources and Classes

Effective in version 10.4.0, you can assign custom attributes to specific resources and classes in the catalog or to all resources and classes in the catalog. You can perform this task when you create or modify a custom attribute.

For more information, see the *Informatica 10.4.0 Catalog Administrator Guide*.

New Resources

Effective in version 10.4.0, the following new resources are added in Enterprise Data Catalog:

- AWS Glue
- Microsoft Power BI
- Apache Cassandra

- Snowflake
- Google Cloud Storage

You can extract metadata, relationship and lineage information from all the above resources. For more information, see the *Informatica 10.4.0 Enterprise Data Catalog Scanner Configuration Guide*.

Reference Resources and Reference Assets

Effective in version 10.4.0, you can configure resources to extract metadata about data sources or other resources in the catalog referenced by the resource. Examples include source and target tables in PowerCenter mappings, and source tables and files from Tableau reports. These referenced data sources are referred to as Reference Resources and the assets included in the data sources are referred to as reference assets. You can view the reference resources and reference assets in Enterprise Data Catalog applications. To view the complete lineage in Enterprise Data Catalog, you can perform connection assignments for the reference resources. You can perform connection assignments between the reference resources and the resources you configure in Enterprise Data Catalog for the reference resource data sources.

You can configure the following resources to extract metadata about data sources or other resources in the catalog referenced by the resource:

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

For more information, see the *Informatica 10.4.0 Catalog Administrator Guide* and the *Informatica 10.4.0 Enterprise Data Catalog User Guide*.

Export Assets from Search Results

Effective in version 10.4.0, you can export all the assets that appear in the search results to the comma-separated values (CSV) file and import the CSV file back into Enterprise Data Catalog. Export assets after you refine the search results on any search tab.

For more information, see the *Asset Tasks* chapter in the *Informatica 10.4.0 Enterprise Data Catalog User Guide*.

Lineage and Impact Filters

Effective in version 10.4.0, you can create and apply filters to the Lineage and Impact view. Use filters to narrow down the results in the Lineage and Impact view. Create customized filters using different combinations of filter options based on the asset types available in the Lineage and Impact view.

For more information, see the *View Lineage and Impact* chapter in the *Informatica 10.4.0 Enterprise Data Catalog User Guide*.

Asset Control Summary

Effective in version 10.4.0, you can view control flow assets in the **Asset Control Summary** tab for table and file asset types. A control flow places constraints or conditions on the flow of data. For example, a SQL query can contain constraints in the WHERE clause or mappings can contain transformations.

For more information, see the *View Lineage and Impact* chapter in the *Informatica 10.4.0 Enterprise Data Catalog User Guide*.

Rules and Scorecards

Effective in version 10.4.0, you can extract rule and scorecard results from the profiling warehouse and view them in Enterprise Data Catalog. You can view the column results for a rule and scorecard to measure data quality of the data source.

For more information, see the *View Assets* chapter in the *Informatica 10.4.0 Enterprise Data Catalog User Guide*.

Unique Key Inference

Effective in version 10.4.0, you can generate unique key candidates from the columns in the source data object. You can view the unique key inference results to understand data quality issues by identifying the columns that have duplicate values.

You can accept or reject the inferred unique key inference results. After you accept or reject an inferred unique key inference, you can reset the unique key inference to restore the inferred status.

For more information, see the *View Assets* chapter in the *Informatica 10.4.0 Enterprise Data Catalog User Guide*.

Data Domain Discovery on the CLOB File Type

Effective in version 10.4.0, you can perform data domain discovery on the CLOB file type.

For more information, see the *Enterprise Data Catalog Concepts* chapter in the *Informatica 10.4.0 Enterprise Catalog Administrator Guide*.

Data Discovery and Sampling Options on the Spark Engine

Effective in version 10.4.0, you can run profiles to discover data domains and choose sampling options on the Spark Engine.

Data Domain Discovery on the Spark Engine

You can perform data domain discovery on the Spark engine.

Sampling Options on the Spark Engine

You can choose the following sampling options to discover data domains on the Spark engine:

- **Limit n** sampling option runs a profile based on the number of the rows in the data object. When you choose to discover data domains in the Hadoop environment, the Spark engine collects samples from multiple partitions of the data object and pushes the samples to a single node to compute sample size.
- **Random percentage** sampling option runs a profile on a percentage of rows in the data object.

For more information, see the *Enterprise Data Catalog Concepts* chapter in the *Informatica 10.4.0 Enterprise Catalog Administrator Guide*.

Tracking Technical Preview

Enterprise Data Catalog version 10.4.0 includes functionality that is available for technical preview.

Technical preview functionality is supported but is unwarranted and is not production-ready. Informatica recommends that you use in non-production environments only. Informatica intends to include the preview functionality in an upcoming GA release for production use, but might choose not to in accordance with changing market or technical circumstances. For more information, contact Informatica Global Customer Support.

- Effective in version 10.4.0, you can choose to display the compact view of the lineage and Impact view. The compact lineage and impact view displays the lineage and impact diagram summarized at the resource level.
For more information, see the *View Lineage and Impact* chapter in the *Informatica 10.4.0 Enterprise Data Catalog User Guide*.
- Effective in version 10.4.0, you can extract metadata from the SAP Business Warehouse, SAP BW/4HANA, IBM InfoSphere DataStage, and Oracle Data Integrator sources when they are inaccessible at runtime or offline.
For more information, see the *Informatica 10.4.0 Catalog Administrator Guide*.
- Effective in version 10.4.0, you can extract metadata from the SAP Business Warehouse and SAP BW/4HANA data sources.
For more information, see the *Informatica 10.4.0 Enterprise Data Catalog Scanner Configuration Guide*.

Data Preview and Provisioning

Effective in version 10.4.0, you can perform data provisioning after you complete data discovery in the catalog. Data provisioning helps you to move data to a target for further analysis. You can preview data for relational sources to assess the data before you move it to the target.

For more information about previewing and provisioning data, see the *Informatica 10.4.0 Catalog Administrator Guide* and *Informatica 10.4.0 Enterprise Data Catalog User Guide*.

Supported Resource Types for Standalone Scanner Utility

Effective in version 10.4, you can extract metadata from the following external sources when they are inaccessible at runtime or offline:

- Amazon Redshift
- Amazon S3
- Apache Cassandra
- Axon
- Azure Data Lake Store
- Azure Microsoft SQL Data Warehouse
- Azure Microsoft SQL Server
- Business Glossary
- Custom Lineage
- Database Scripts
- Erwin
- Glue

- Google BigQuery
- Google Cloud Storage
- Informatica Cloud Service
- IBM Cognos
- Microsoft Azure Blob Storage
- Microsoft SQL Server Integration Services
- PowerBI
- QlikView Business Intelligence
- Salesforce
- SAP HANA
- SAP PowerDesigner
- Snowflake
- Workday

For more information, see the "Metadata Extraction from Offline and Inaccessible Resources" chapter in the *Informatica 10.4 Enterprise Data Catalog Administrator Guide*.

REST APIs

Effective in version 10.4, you can use the following Informatica Enterprise Data Catalog REST APIs:

- Data Provision REST APIs. In addition to the existing REST APIs, you can view whether data provisioning is available to the user and list the resources that support data provisioning.
- Lineage Filter REST APIs. You can create, update, list, or delete a lineage filter.
- Model Information REST APIs. In addition to the existing REST APIs, you can list the predefined slider facets, slider facet definition, and lineage filter definitions.
- Model Modification REST API. In addition to the existing REST APIs, you can create, update, and delete a slider facet definition.
- Monitoring Info REST APIs. You can submit or list jobs which includes jobs of object export type, object import type, resource export type, and search export type.
- Object Child Count REST API. You can list the total number of child assets for an object.
- Product Info REST API. You can list the details about Enterprise Data Catalog which includes the release version, build version, and build date.

For more information about the REST APIs, see the *Informatica 10.4 Enterprise Data Catalog REST API Reference*.

Enterprise Data Preparation

This section describes new Enterprise Data Preparation features in version 10.4.0.

Data Lake Access Management

Effective in version 10.4.0, you can manage access to the data lake by assigning permissions on the Enterprise Data Catalog resources that represent Hive schemas and HDFS locations in the data lake to Enterprise Data Preparation users and groups of users.

When you grant permissions on specific schemas or locations to a user or group of users, the application displays only those schemas and locations to which a user has permissions when the user performs an import, publish, or upload operation.

For more information, see the *Enterprise Data Preparation 10.4.0 Administrator Guide*.

Microsoft Azure Data Lake Storage as a Data Source

Effective in version 10.4.0, you can use a Microsoft Azure Data Lake Storage Gen2 location as a data source.

Publish Files to the Data Lake

Effective in version 10.4.0, you can publish prepared data in a file to an HDFS location in the data lake.

When you publish data, you can select the file type to write the data to in the data lake. For example, if choose to publish data as a comma separated value file, the application writes the data to the data lake as a .csv file.

For more information, see the *Enterprise Data Preparation 10.4.0 User Guide*.

Upload Files to the Data Lake

Effective in version 10.4.0, you can upload several file formats from your local drive to the data lake.

You can choose from the following upload options:

Upload a file directly to the data lake.

You can upload a comma delimited file, an Avro file, a JSON file, or a Parquet file in UTF-8 format directly from your local drive to the data lake, without previewing the data. You might choose this option if you want to upload a file without previewing the data.

Allow CLAIRE to determine the file structure, and then upload the file to the data lake.

You can upload the data in a comma delimited file or a Microsoft Excel spreadsheet to the data lake. When you upload the file, Enterprise Data Preparation uses the CLAIRE embedded discovery engine to determine the structure of the file and display a preview of the data.

When you use this option to upload an Excel spreadsheet, the CLAIRE engine discovers the sheets and tables in the spreadsheet. You can select the sheet and table you want to preview.

Note: Uploading an Excel spreadsheet is a preview feature.

Define the file structure, and then upload the file to the data lake.

You can upload the data in a comma delimited file from your local drive to the data lake. When you upload the file, you can preview the data, specify the structure of the file, and configure column attributes to meet your requirements. You might choose this option if you need to modify column attributes before you upload the file.

For more information, see the *Enterprise Data Preparation 10.4.0 User Guide*.

Informatica Mappings

This section describes new features for Informatica mappings in version 10.4.0.

Binding Mapping Outputs to Mapping Parameters

Effective in version 10.4.0, you can persist and bind mapping outputs in mappings that are deployed to run in the native environment or on the Spark engine.

Create a mapping output. Bind the output to a mapping parameter to use the value in subsequent runs of the mapping. When you run the mapping, the Data Integration Service passes the value of the mapping output to the mapping parameter. To persist mapping outputs, you must specify a run-time instance name using the `- RuntimeInstanceName` option for the `infacmd ms runMapping` command.

The Developer tool now includes a **Binding** column in the mapping **Properties** view to bind a mapping output to a parameter.

For information about mapping outputs in deployed mappings, see the "Mapping Outputs" chapter in the *Informatica 10.4.0 Developer Mapping Guide*.

infacmd ms Commands

The following table describes new and updated `infacmd ms` commands:

Command	Description
<code>deleteMappingPersistedOutputs</code>	New command that deletes all persisted mapping outputs for a deployed mapping. Specify the outputs to delete using the name of the application and the run-time instance name of the mapping. To delete specific outputs, use the option <code>- OutputNamesToDelete</code> .
<code>getMappingStatus</code>	Updated command that now returns the job name. If you defined a run-time instance name in <code>infacmd ms runMapping</code> , the job name is the run-time instance name.
<code>listMappingPersistedOutputs</code>	New command that lists the persisted mapping outputs for a deployed mapping. The outputs are listed based on the name of the application and the run-time instance name of the mapping.

For more information, see the "infacmd ms Command Reference" chapter in the *Informatica 10.4.0 Command Reference*.

CLAIRE Recommendations and Insights

Effective in version 10.4.0, you can use CLAIRE artificial intelligence to tune and correct mappings during development.

When you enable recommendations, CLAIRE runs automatically on mappings as you develop them, and displays recommendations that enable you to correct or tune the mapping.

You can also run CLAIRE analysis on the mappings within a project or project folder. When you analyze a group of mappings, CLAIRE displays insights about similarities between mappings.

For more information about recommendations and insights, see the *Data Engineering Integration User Guide*.

Update Mapping Optimizer Level

Effective in version 10.4.0, you can use the `infacmd ms UpdateOptimizationDefaultLevel` command to set the optimizer level of the mappings in an application to the default level, Auto.

When you run the command, you must specify an application name. `UpdateOptimizationDefaultLevel` sets the optimizer level for all mappings in the application with optimization level Normal. The command does not affect mappings in the application with optimization level other than Normal.

For more information, see the *Informatica 10.4.0 Command Reference* and the *Informatica 10.4.0 Developer Mapping Guide*.

Informatica Transformations

This section describes new features in Informatica transformations in version 10.4.0.

Address Validator Transformation

This section describes new Address Validator transformation features.

The Address Validator transformation contains additional address functionality for the following countries:

United States

Effective in version 10.4, the Address Validator recognizes MC as an alternative version of MSC, or Mail Stop Code, in a United States address.

For comprehensive information about the features and operations of the address verification software engine in version 10.4, see the *Informatica Address Verification 5.15.0 Developer Guide*.

Informatica Workflows

This section describes new Informatica workflow features in version 10.4.0.

Amazon EMR Create Cluster Task Advanced Properties

Effective in version 10.4.0, new advanced properties are available when you configure a Create Cluster task with an Amazon EMR connection in a workflow.

You can configure the following functionality on a Create Cluster task:

- Root device EBS volume size. The number of GB of the EBS root device volume.
- Custom AMI ID. ID of a custom Amazon Linux Amazon Machine Image (AMI).
- Security Configuration. The name of a security configuration for authentication and encryption on the cluster.

For more information, see the *Data Engineering Integration 10.4.0 User Guide* and the *Informatica® 10.4.0 Developer Workflow Guide*.

Informatica Installation

This section describes new installation features in 10.4.0.

PostgreSQL

Effective in version 10.4.0, you can use the PostgreSQL database for the domain configuration repository, the Model repository, and the PowerCenter repository. For Enterprise Data Preparation, you can use PostgreSQL database only on the additional Model Repository Service.

You can also install psql client application version 10.6 for PostgreSQL to work on Linux or Windows.

For more information about PostgreSQL, refer the *Informatica 10.4.0* installation guides.

Pre-installation (i10Pi) System Check Tool in Silent Mode

Effective in version 10.4.0, you can run the Pre-installation (i10pi) System Check Tool in silent mode. You can run i10pi in silent mode to verify whether the machine meets the system requirements for installation without user interaction.

For more information about how to run i10Pi in silent mode, see the Informatica 10.4.0 installation guide.

Encrypt Passwords in the Silent Installation Properties File

Effective in version 10.4.0, the Informatica installer includes a utility that you can use to encrypt passwords that you set in the properties file containing the options required to install the services in silent mode.

When you run the installer in silent mode, the installation framework decrypts the encrypted passwords.

For more information, see the *Informatica Installation and Configuration Guide*.

Intelligent Structure Model

This section describes new intelligent structure model features in version 10.4.0.

Additional Input Types

Effective in version 10.4.0, Intelligent Structure Discovery can process ORC, Avro, and Parquet inputs.

For more information, refer to the *Data Engineering Integration 10.4.0 User Guide*.

Create Model from Sample at Design Time

Effective in version 10.4.0, Intelligent Structure Discovery can create a model based on a sample file that you select when you create a complex data object. You can then refine the model in Informatica Intelligent Cloud Services Data Integration.

This functionality is supported for XML, JSON, ORC, AVRO and Parquet sample files.

For more information, refer to the *Data Engineering Integration 10.4.0 User Guide*.

Handling of Unidentified Data

Effective in version 10.4.0, Intelligent Structure Discovery arranges the output of unidentified data in the model in structured JSON format.

For more information, refer to the *Data Engineering Integration 10.4.0 User Guide*.

PowerCenter

This section describes new PowerCenter features in version 10.4.0.

HTTP Transformation

Effective in version 10.4.0, the HTTP transformation also includes the following methods for the final URL construction: SIMPLE PATCH, SIMPLE PUT, and SIMPLE DELETE.

You can perform a partial update and the input data does not need to be a complete body with the SIMPLE PATCH method. You can use it to update data from input port as a patch to the resource.

You can perform a complete replacement of a document with the SIMPLE PUT method. You can create data from an input port as a single block of data to the HTTP server. If the data already exists, you can update data from an input port as a single block of data to the HTTP server.

You can delete data from the HTTP server with the SIMPLE DELETE method.

You can also parameterize the base URL for the HTTP transformation.

Previously, you could specify the final URL construction only for the following two methods: SIMPLE GET and SIMPLE POST. You also could not parameterize the final URL for the HTTP transformation.

For more information, see the "HTTP transformation" chapter in the *PowerCenter 10.4.0 Transformation Guide*.

Connectivity

This section describes new connectivity features in version 10.4.0.

User Impersonation for Oracle Connections

Effective in version 10.4.0, you can specify the impersonate user name in the Oracle connection in PowerCenter to connect to Oracle sources and targets. When you specify an impersonate user, you can connect to Oracle on behalf of that user and perform read or write operations.

For more information, see the *Informatica 10.4.0 PowerCenter Workflow Basics Guide*.

Support for SAP HANA Database Modelling Views

Effective in version 10.4.0, you can read data from the following types of SAP HANA database modelling views:

- Analytics Views
- Attribute Views
- Calculated Views

For more information, see the *Informatica 10.4.0 PowerCenter Designer Guide*.

PowerExchange Adapters

This section describes new PowerExchange adapter features in version 10.4.0.

PowerExchange Adapters for Informatica

This section describes new Informatica adapter features in version 10.4.0.

PowerExchange for Amazon Redshift

Effective in version 10.4.0, PowerExchange for Amazon Redshift includes the following features:

- You can run mappings in the AWS Databricks environment.
- You can select a cluster region name in the Cluster Region connection property, even though you specify the cluster region name in the JDBC URL connection property.
- You can retain null values when you read data from Amazon Redshift.
- You can specify the number of staging files per batch, when you write data to Amazon Redshift.
- You can preserve the record order, when you write data from a CDC source to an Amazon Redshift target.

For more information, see the *Informatica 10.4.0 PowerExchange for Amazon Redshift User Guide*.

PowerExchange for Amazon S3

Effective in version 10.4.0, PowerExchange for Amazon S3 includes the following features:

- You can run mappings in the AWS Databricks environment.
- You can use the temporary security credentials using AssumeRole to access the AWS resources.
- You can parameterize the data format type and schema in the read and write operation properties at run time.

For more information, see the *Informatica 10.4.0 PowerExchange for Amazon S3 User Guide*.

PowerExchange for Google BigQuery

Effective in version 10.4.0, PowerExchange for Google BigQuery includes the following features:

- You use a Google Dataproc cluster to run mappings on the Spark engine.
- You can increase the performance of the mapping by running the mapping in Optimized Spark mode. When you use the Optimized Spark mode to read data, you can specify the number of partitions to use. You can specify whether to run the mapping in **Generic** or **Optimized** mode in the advanced read and write operation properties. Optimized Spark mode increases the mapping performance.
- You can configure a SQL override to override the default SQL query used to extract data from the Google BigQuery source.
- You can read or write data of the NUMERIC data type to Google BigQuery. The NUMERIC data type is an exact numeric value with 38 digits of precision and 9 decimal digits of scale. When you read or write the NUMERIC data type, the Data Integration Service maps the NUMERIC data type to the Decimal transformation data type and the allowed precision is up to 38 and scale up to 9.

For more information, see the *Informatica 10.4.0 PowerExchange for Google BigQuery User Guide*.

PowerExchange for Google Cloud Storage

Effective in version 10.4.0, PowerExchange for Google Cloud Storage includes the following features:

- You use a Google Dataproc cluster to run mappings on the Spark engine.
- You can configure the following Google Cloud Storage data object read operation advanced properties when you read data from a Google Cloud Storage source:

Google Cloud Storage Path

Overrides the Google Cloud Storage path to the file that you selected in the Google Cloud Storage data object.

Use the following format:

```
gs://<bucket name> or gs://<bucket name>/<folder name>
```

Source File Name

Overrides the Google Cloud Storage source file name specified in the Google Cloud Storage data object.

Is Directory

Reads all the files available in the folder specified in the **Google Cloud Storage Path** data object read operation advanced property.

For more information, see the *Informatica 10.4.0 PowerExchange for Google Cloud Storage User Guide*.

PowerExchange for Microsoft Azure Blob Storage

Effective in version 10.4.0, PowerExchange for Microsoft Azure Blob Storage includes the following features:

- You can parameterize the data format type and schema in the read and write operation properties at run time.
- You can use shared access signatures authentication while creating a Microsoft Azure Blob Storage connection.

For more information, see the *Informatica 10.4.0 PowerExchange for Microsoft Azure Blob Storage User Guide*.

PowerExchange for Microsoft Azure SQL Data Warehouse

Effective in version 10.4.0, you can read data from or write data to a Microsoft Azure SQL Data Warehouse endpoint that reside in a virtual network (VNet).

For more information, see the *Informatica 10.4.0 PowerExchange for Microsoft Azure SQL Data Warehouse User Guide*.

PowerExchange for Salesforce

Effective in version 10.4.0, PowerExchange for Salesforce includes the following features:

- You can use version 45.0, 46.0, and 47.0 of Salesforce API to create a Salesforce connection and access Salesforce objects.
- You can enable primary key chunking for queries on a shared object that represents a sharing entry on the parent object. Primary key chunking is supported for shared objects only if the parent object is supported. For example, if you want to query on CaseHistory, primary key chunking must be supported for the parent object Case.
- You can create assignment rules to reassign attributes in records when you insert, update, or upsert records for Lead and Case target objects using the standard API.

For more information, see the *Informatica 10.4.0 PowerExchange for Salesforce User Guide*.

PowerExchange for SAP NetWeaver

Effective in version 10.4.0, PowerExchange for SAP NetWeaver includes the following features:

- You can configure HTTPS streaming for SAP table reader mappings.
- You can read data from ABAP CDS views using SAP Table Reader if your SAP NetWeaver system version is 7.50 or later.
- You can read data from SAP tables with fields that have the following data types:
 - DF16_DEC
 - DF32_DEC
 - DF16_RAW
 - DF34_RAW
 - INT8
 - RAWSTRING
 - SSTRING
 - STRING

For more information, see the *Informatica PowerExchange for SAP NetWeaver 10.4.0 User Guide*.

PowerExchange for Snowflake

Effective in version 10.4.0, PowerExchange for Snowflake includes the following features:

- You can run Snowflake mappings in the Databricks environment.
- You can use Snowflake objects as dynamic sources and targets in a mapping.
- You can create a Snowflake target using the **Create Target** option.
- You can configure a target schema strategy for a Snowflake target in a mapping. You can choose from the available options to either retain the existing target schema or create a target if it does not exist. You can also specify the target schema strategy options as a parameter value.
- You can specify a rejected file name and path in the Snowflake advanced target properties which the Data Integration Service uses to write records that are rejected while writing to the target.
- When the ODBC provider type in the Snowflake ODBC connection is Snowflake, you can configure pushdown optimization to push the transformation logic to the Snowflake database.
- You can read or write data of Decimal data type of 38-bit precision and scale by configuring the `EnableSDKDecimal38` custom flag in the Data Integration Service properties.

For more information, see the *Informatica 10.4.0 PowerExchange for Snowflake User Guide*.

PowerExchange for HDFS

Effective in version 10.4.0, PowerExchange for HDFS includes the following features:

- You can parameterize the data format type and schema in the read and write operation properties at run time.
- You can format the schema of a complex file data object for a read or write operation.

For more information, see the *Informatica 10.4.0 PowerExchange for HDFS User Guide*.

PowerExchange Adapters for PowerCenter

This section describes new PowerCenter adapter features in version 10.4.0.

PowerExchange for Google BigQuery

Effective in version 10.4.0, you can read or write data of the NUMERIC data type to Google BigQuery. The NUMERIC data type is an exact numeric value with 38 digits of precision and 9 decimal digits of scale. When you read or write the NUMERIC data type, the PowerCenter Integration Service maps the NUMERIC data type to the Decimal transformation data type and the allowed precision is up to 28 and scale up to 9.

For more information, see the *Informatica 10.4.0 PowerExchange for Google BigQuery User Guide for PowerCenter*.

PowerExchange for Google Cloud Storage

Effective in version 10.4.0, you can configure the following Google Cloud Storage data object read operation advanced properties when you read data from a Google Cloud Storage source:

Google Cloud Storage Path

Overrides the Google Cloud Storage path to the file that you selected in the Google Cloud Storage data object.

Use the following format:

```
gs://<bucket name> or gs://<bucket name>/<folder name>
```

Source File Name

Overrides the Google Cloud Storage source file name specified in the Google Cloud Storage data object.

Is Directory

Reads all the files available in the folder specified in the **Google Cloud Storage Path** data object read operation advanced property.

For more information, see the *Informatica 10.4.0 PowerExchange for Google Cloud Storage User Guide for PowerCenter*.

PowerExchange for Greenplum

Effective in version 10.4.0, you can use PowerExchange for Greenplum to read data from Greenplum. You can configure specific session properties for Greenplum sources to determine how to extract data from Greenplum.

When you run a Greenplum session to read data, the PowerCenter Integration Service invokes the Greenplum database parallel file server, gpfdist, which is Greenplum's file distribution program, to read data.

For more information, see the *Informatica 10.4.0 PowerExchange for Greenplum User Guide for PowerCenter*.

PowerExchange for JD Edwards EnterpriseOne

Effective in version 10.4.0, you can use version 9.2 of JD Edwards EnterpriseOne API to create a JD Edwards EnterpriseOne connection and access JD Edwards EnterpriseOne objects.

For more information, see the *Informatica 10.4.0 PowerExchange for JD Edwards EnterpriseOne User Guide for PowerCenter*.

PowerExchange for Kafka

Effective in version 10.4.0, you can use configure the following SSL properties to enable a secure connection to a Kafka broker:

- SSL Mode

- SSL TrustStore File Path
- SSL TrustStore Password
- SSL KeyStore File Path
- SSL KeyStore Password

You can configure the Kafka messaging broker to use Kafka broker version 0.10.1.1 and above.

For more information, see the *Informatica PowerExchange for Kafka 10.4.0 User Guide for PowerCenter*.

PowerExchange for Salesforce

Effective in version 10.4.0, you can use version 46.0 and 47.0 of Salesforce API to create a Salesforce connection and access Salesforce objects.

For more information, see the *Informatica 10.4.0 PowerExchange for Salesforce User Guide for PowerCenter*.

PowerExchange for SAP NetWeaver

Effective in version 10.4.0, you can use PowerExchange for SAP Dynamic ABAP Table Extractor to read data from SAP tables and ABAP Core Data Services (CDS) views through HTTP/HTTPS streaming. You can read data from ABAP CDS views using PowerExchange for SAP Dynamic ABAP Table Extractor if your SAP NetWeaver system version is 7.50 or later.

For more information, see the *Informatica PowerExchange for SAP NetWeaver 10.4.0 User Guide*.

Security

This section describes new security features in version 10.4.0.

Configure Web Applications to Use Different SAML Identity Providers

Effective in version 10.4.0, you can configure Informatica web applications that run in a domain to use different Security Assertion Markup Language (SAML) identity providers. For example, you might configure Informatica Administrator to use Active Directory Federation Services as the identity provider, and configure Informatica Analyst to use PingFederate as the identity provider.

For more information, see the *Informatica 10.4.0 Security Guide*.

CHAPTER 3

New Features (10.2.2 HotFix 1)

This chapter includes the following topics:

- [Command Line Programs , 53](#)
- [Enterprise Data Catalog, 53](#)

Command Line Programs

This section describes new commands in version 10.2.2 HotFix 1.

infacmd Idm Commands

The following table describes new infacmd Idm command options:

New Option	Description
-Force -fr	Optional. If you want to force backup when the backup mode is offline. Forcefully takes backup and overwrites the existing backup.
-Force -fr	Optional. If you want to clean the existing content from HDFS and Apache Zookeeper. Forcefully restores the backup data.

For more information, see the "infacmd Idm Command Reference" chapter in the *Informatica 10.2.2 HotFix 1 Command Reference*.

Enterprise Data Catalog

This section describes new Enterprise Data Catalog features in version 10.2.2 HotFix 1.

Azure Data Lake Storage Gen2

Effective in version 10.2.2 HotFix 1, you can extract metadata from an Azure Data Lake Storage Gen2 source. You can access and extract the metadata of files and folders in an Azure Data Lake Storage Gen2 storage account to Enterprise Data Catalog.

For more information, see the "Azure Data Lake Store" chapter in the *Informatica 10.2.2 HotFix1 Enterprise Data Catalog Scanner Configuration Guide*.

Case Insensitive Linking

Effective in version 10.2.2 HotFix 1, you can configure resources as case insensitive. This option enables you to view linked lineage between sources and targets irrespective of the case sensitivity on the resource names.

For information about case insensitive linking, see the "Managing Resources" chapter in the *Informatica 10.2.2 HotFix 1 Catalog Administrator Guide*.

Enterprise Data Catalog Tableau Extension

Effective in version 10.2.2 HotFix 1, Enterprise Data Catalog Tableau Extension is a native extension for Tableau dashboard that you can use within Tableau.

You can use Enterprise Data Catalog Tableau Extension in Tableau Desktop, Tableau Server, and all the web browsers that Tableau supports. Download the extension from Enterprise Data Catalog application, and add the extension to a dashboard in Tableau.

For more information about the extension, see the *Informatica 10.2.2 HotFix1 Enterprise Data Catalog Extension for Tableau* guide.

New Resources

Effective in version 10.2.2 HotFix 1, the following new resources are added in Enterprise Data Catalog:

- SAP PowerDesigner. You can extract metadata, relationship, and lineage information from a SAP PowerDesigner data source.
- SAP HANA. You can extract object and lineage metadata from a SAP HANA database.

For more information, see the *Informatica 10.2.2 HotFix1 Scanner Configuration Guide*.

Profiling Warehouse Scanner

Effective in version 10.2.2 HotFix 1, you can create one or more profiling warehouse resources. You can associate a profiling warehouse and choose schemas in the resource. When you run the resource, the scanner extracts and migrates the profile results for the selected schemas from the associated profiling warehouse to the catalog.

For more information, see the "Configuring Informatica Platform Scanners" chapter in the *Informatica 10.2.2 HotFix1 Enterprise Data Catalog Scanner Configuration Guide*.

REST APIs

Effective in version 10.2.2 HotFix 1, you can use the following Informatica Enterprise Data Catalog REST APIs:

- Data Provision REST APIs. You can return, update, or delete connections and resources.
- Catalog Model REST APIs. In addition to the existing REST APIs, you can access, update, or delete the field facets, query facets, and search tabs.
- Object APIs. In addition to the existing REST APIs, you can list the catalog search and suggestions.

For more information about the REST APIs, see the *Informatica 10.2 .2 HotFix 1 Enterprise Data Catalog REST API Reference*.

Search Enhancements

Effective in version 10.2.2 HotFix 1, you can perform better search in Enterprise Data Catalog based on the following enhancements:

Double Quotes Search

You can perform a search for assets using the double quotes (") to find assets that exactly match the asset name within the double quotes, but not the variations of the asset name in the catalog.

Search Operators

You can use newer search operators to make the search results accurate. The search operators are, AND, OR, NOT, title, and description.

Search Rank

Enterprise Data Catalog uses a ranking algorithm to rank data assets in the search results page. Search ranking refers to the precedence of an asset when compared to other assets that are part of specific search results.

Related Search

You can enable the **Show Related Search** option in the **Search Results** page to view related assets.

For more information about search enhancements, see the "Search for Assets" chapter in the *Informatica 10.2.2 HotFix 1 Enterprise Data Catalog User Guide*.

Search Tabs

Effective in version 10.2.2 HotFix 1, you can use the search tabs to search assets without having to repeatedly set the same search criteria when you perform a search for assets. The search tabs are predefined filters in the Catalog.

For more information about search tabs, see the "Customize Search" chapter in the *Informatica 10.2.2 HotFix 1 Enterprise Data Catalog User Guide*.

Supported Resource Types for Standalone Scanner Utility

Effective in version 10.2.2 HotFix 1, you can extract metadata from the following external sources when they are inaccessible at runtime or offline:

- Apache Atlas
- Cloudera Navigator
- File System
- HDFS
- Hive
- Informatica Platform
- MicroStrategy
- OneDrive
- Oracle Business Intelligence

- SharePoint
- Sybase
- Tableau

For more information, see the "Metadata Extraction from Offline and Inaccessible Resources" chapter in the *Informatica 10.2.2 HotFix1 Enterprise Data Catalog Administrator Guide*.

Technical Preview

Enterprise Data Catalog version 10.2.2 HotFix 1 includes functionality that is available for technical preview.

Technical preview functionality is supported but is unwarranted and is not production-ready. Informatica recommends that you use in non-production environments only. Informatica intends to include the preview functionality in an upcoming GA release for production use, but might choose not to in accordance with changing market or technical circumstances. For more information, contact Informatica Global Customer Support.

Effective in version 10.2.2 HotFix 1, the following features are available for technical preview:

- Effective in version 10.2.2 HotFix 1, you can extract metadata for data lineage at the column level including transformation logic from an Oracle Data Integrator data source.
- Effective in version 10.2.2 HotFix 1, you can extract metadata for data lineage at the column level including transformation logic from an IBM InfoSphere DataStage data source.
- Effective in version 10.2.2 HotFix 1, you can extract data lineage at the column level for stored procedures in Oracle and SQL Server.
- Effective in version 10.2.2 HotFix 1, you can perform data provisioning after you complete data discovery in the catalog. Data provisioning helps you to move data to a target for further analysis.
For more information about previewing data, see the *Informatica 10.2.2 HotFix 1 Catalog Administrator Guide* and *Informatica 10.2.2 Hotfix 1 Enterprise Data Catalog User Guide*.
- Effective in version 10.2.2 HotFix 1, you can preview data to assess the data before you move it to the target. You can preview data only for tabular assets in Oracle and Microsoft SQL Server resources.
For more information about previewing data, see the *Informatica 10.2.2 HotFix 1 Catalog Administrator Guide* and *Informatica 10.2.2 Hotfix 1 Enterprise Data Catalog User Guide*.

CHAPTER 4

New Features (10.2.2 Service Pack 1)

This chapter includes the following topics:

- [Big Data Management, 57](#)
- [Big Data Streaming, 58](#)
- [Enterprise Data Catalog, 58](#)
- [Enterprise Data Preparation, 59](#)
- [PowerExchange Adapters for Informatica, 59](#)

Big Data Management

This section describes new Big Data Management features in version 10.2.2 Service Pack 1.

Sqoop

Effective in version 10.2.2 Service Pack 1, you can use the following new Sqoop features.

[Connect to SSL-enabled Oracle and Microsoft SQL Server databases](#)

You can configure Sqoop to connect to secure relational databases such as Oracle and Microsoft SQL Server.

Use the appropriate JDBC connection string and the connect argument in the JDBC connection to connect to an SSL-enabled Oracle or Microsoft SQL Server database.

For more information, see the *Informatica Big Data Management 10.2.2 Service Pack 1 User Guide*.

[Support for connection-param-file Sqoop argument](#)

You can define the `connection-param-file` argument to connect to an SSL-enabled Oracle database. The `connection-param-file` is a property file that specifies the extra JDBC parameters that Sqoop must use to connect to the SSL-enabled database.

The contents of this file are parsed as standard Java properties and passed into the driver when you create a connection.

You can specify the `connection-param-file` argument in the **Sqoop Arguments** field in the JDBC connection.

Use the following syntax:

```
--connection-param-file <parameter_file_name>
```

For more information, see the *Informatica Big Data Management 10.2.2 Service Pack 1 User Guide*.

Big Data Streaming

This section describes the new Big Data Streaming features in version 10.2.2 Service Pack 1.

Amazon S3 Target

Effective in version 10.2.2 Service Pack 1, you can create a streaming mapping to write data to Amazon S3.

Create an Amazon S3 data object to write data to Amazon S3. You can create an Amazon S3 connection to use Amazon S3 as targets. You can create and manage an Amazon S3 connection in the Developer tool or through infacmd.

For more information, see the *Informatica Big Data Streaming 10.2.2 Service Pack 1 User Guide*.

TIME_RANGE Function

Effective in version 10.2.2 Service Pack 1, you can use the TIME_RANGE function in a Joiner transformation that determines the time range for the streaming events to be joined.

The TIME_RANGE function is applicable only for a Joiner transformation in a streaming mapping.

Syntax

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

For more information about the TIME_RANGE function, see the *Informatica 10.2.2 Service Pack 1 Transformation Language Reference* guide.

Writing Data to Multiple HDFS Files

Effective in version 10.2.2 Service Pack 1, you can use a complex file data object to write the source data to multiple HDFS files. Based on the FileName header port values of the complex file data object, the Data Integration Service creates multiple HDFS files at run time in the target location.

For more information, see the *Informatica Big Data Streaming 10.2.2 Service Pack 1 User Guide*.

Enterprise Data Catalog

This section describes the new Enterprise Data Catalog feature in version 10.2.2 Service Pack 1.

Supported Resource Types for Standalone Scanner Utility

Effective in version 10.2.2 Service Pack 1, you can extract metadata from the following external sources when they are inaccessible at runtime or offline:

- IBM DB2

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

For more information, see the "Metadata Extraction from Offline and Inaccessible Resources" chapter in the *Informatica 10.2.2 Service Pack 1 Enterprise Data Catalog Administrator Guide*.

Enterprise Data Preparation

This section describes new Enterprise Data Preparation features in version 10.2.2 Service Pack 1.

Revert All Inferred Data Types

Effective in version 10.2.2 Service Pack 1, you can revert all inferred types and data domains applied to source columns in a worksheet back to their original types.

For more information, see the "Prepare Data" chapter in the *Informatica 10.2.2 Service Pack 1 Enterprise Data Preparation User Guide*.

Prepare Avro and Parquet Files

Effective in version 10.2.2 Service Pack 1, you can sample the hierarchical data in Avro and Parquet files you add to your project as the first step in data preparation. Enterprise Data Preparation converts the Avro or Parquet file structure into a flat structure, and presents the data in a worksheet that you use to prepare the data.

For more information, see the "Prepare Data" chapter in the *Informatica 10.2.2 Service Pack 1 Enterprise Data Preparation User Guide*.

PowerExchange Adapters for Informatica

This section describes new Informatica adapter features in version 10.2.2 Service Pack 1.

PowerExchange for Hive

Effective in version 10.2.2 Service Pack 1, when you import a Hive table, You can use the **Show Default Schema Only** option to either clear search results and show only tables that use the default schema or search for tables that use a schema other than the default schema.

For more information, see the *Informatica PowerExchange for Hive 10.2.2 Service Pack 1 User Guide*.

CHAPTER 5

New Features (10.2.2)

This chapter includes the following topics:

- [Application Services, 60](#)
- [Big Data Management, 61](#)
- [Big Data Streaming , 65](#)
- [Command Line Programs, 67](#)
- [Enterprise Data Catalog, 71](#)
- [Enterprise Data Lake, 74](#)
- [Informatica Developer, 79](#)
- [Informatica Mappings, 79](#)
- [Informatica Transformations, 81](#)
- [PowerExchange Adapters for Informatica, 84](#)

Application Services

This section describes new application service features in version 10.2.2.

Mass Ingestion Service

Effective in version 10.2.2, you can enable secure communication between a Mass Ingestion Service process and external components. You can specify the HTTPS port number for the HTTPS protocol and a keystore file for the SSL certificate.

For more information, see the "Mass Ingestion Service" chapter in the *Informatica 10.2.2 Application Service Guide*.

Metadata Access Service

Effective in version 10.2.2, you can configure the Metadata Access Service to use operating system profiles. Use operating system profiles to increase security and to isolate the design-time user environment when you import and preview metadata from a Hadoop cluster.

For more information, see the "Users and Groups" chapter in the *Informatica 10.2.2 Security Guide*.

REST Operations Hub Service

Effective in version 10.2.2, you can configure a REST Operations Hub Service for REST applications. The REST Operations Hub Service is a REST system service in the Informatica domain that exposes Informatica product functionality to external clients through REST APIs.

You can configure the REST Operations Hub Service through the Administrator tool or through infacmd. You can use the REST Operations Hub Service to view mapping execution statistics for the deployed mapping jobs in the application.

You can use the REST Operations Hub Service to get mapping execution statistics for big data mappings that run on the Data Integration Service, or in the Hadoop environment.

For more information about the REST API, see the *Big Data Management 10.2.2 Administrator Guide*.

Big Data Management

This section describes new Big Data Management features in version 10.2.2.

Azure Databricks Integration

Effective in version 10.2.2, you can integrate the Informatica domain with the Azure Databricks environment.

Azure Databricks is an analytics cloud platform that is optimized for the Microsoft Azure cloud services. It incorporates the open source Apache Spark cluster technologies and capabilities.

The Informatica domain can be installed on an Azure VM or on-premises. The integration process is similar to the integration with the Hadoop environment. You perform integration tasks, including importing the cluster configuration from the Databricks environment. The Informatica domain uses token authentication to access the Databricks environment. The Databricks token ID is stored in the Databricks connection.

Sources and Targets

You can run mappings against the following sources and targets within the Databricks environment:

- Microsoft Azure Data Lake Store
- Microsoft Azure Blob Storage
- Microsoft Azure SQL Data Warehouse
- Microsoft Azure Cosmos DB

Transformations

You can add the following transformations to a Databricks mapping:

- Aggregator
- Expression
- Filter
- Joiner
- Lookup
- Normalizer
- Rank

Router
Sorter
Union

The Databricks Spark engine processes the transformation in much the same way as the Spark engine processes in the Hadoop environment.

Data Types

The following data types are supported:

Array
Bigint
Date/time
Decimal
Double
Integer
Map
Struct
Text
String

Mappings

When you configure a mapping, you can choose to validate and run the mapping in the Databricks environment. When you run the mapping, the Data Integration Service generates Scala code and passes it to the Databricks Spark engine.

Workflows

You can develop cluster workflows to create ephemeral clusters in the Databricks environment.

For more information, refer to the following guides:

Big Data Management 10.2.2 Integration Guide

Big Data Management 10.2.2 Administrator Guide

Big Data Management 10.2.2 User Guide

Data Preview on the Spark Engine

Effective in version 10.2.2, you can preview data within a mapping that runs on the Spark engine in the Developer tool. Previewing data helps to design and debug big data mappings.

You can choose sources and transformations as preview points in a mapping that contain the following hierarchical types:

- Array
- Struct
- Map

Data preview is available for technical preview. Technical preview functionality is supported for evaluation purposes but is unwarranted and is not production-ready. Informatica recommends that you use in non-production environments only. Informatica intends to include the preview functionality in an upcoming

release for production use, but might choose not to in accordance with changing market or technical circumstances. For more information, contact Informatica Global Customer Support.

For more information, see the *Informatica® Big Data Management 10.2.2 User Guide*.

Hierarchical Data

This section describes new features for hierarchical data in version 10.2.2.

Dynamic Complex Ports

Effective in version 10.2.2, you can add dynamic complex ports to a dynamic mapping that runs on the Spark engine. Use dynamic complex ports to manage frequent schema changes to hierarchical data in complex files.

A dynamic complex port receives new or changed elements of a complex port based on the schema changes at run time. The input rules determine the elements of a dynamic complex port. Based on the input rules, a dynamic complex port receives one or more elements of a complex port from the upstream transformation. You can use dynamic complex ports such as dynamic array, dynamic map, and dynamic struct in some transformations on the Spark engine.

For more information, see the "Processing Hierarchical Data with Schema Changes" chapter in the *Informatica Big Data Management 10.2.2 User Guide*.

High Availability

This section describes new high availability features in version 10.2.2.

Big Data Job Recovery

Effective in version 10.2.2, the Data Integration Service can recover a big data job configured to run on the Spark engine when a Data Integration Service node stops unexpectedly. When a Data Integration Service node fails before a job completes, the Data Integration Service sends the job to another node, which resumes processing job tasks from the point at which the node failure occurred.

To recover big data mappings, you must enable big data job recovery in Data Integration Service properties and run the job from infacmd.

For more information, see the "Data Integration Service Processing" chapter in the *Informatica Big Data Management 10.2.2 Administrator Guide*.

Distributed Data Integration Service Queues

Effective in version 10.2.2, the Data Integration Service uses a distributed queue to store job information when big data recovery is enabled for deployed big data jobs. The distributed queue is stored in the Model repository, and any available Data Integration Service can run jobs from the queue when resources are available.

For more information, see the "Data Integration Service Processing" chapter in the *Informatica Big Data Management 10.2.2 Administrator Guide*.

Intelligent Structure Model

This section describes new intelligent structure model features in version 10.2.2.

Aliases in XML Files

Effective in version 10.2.2, Intelligent Structure Discovery can process XML files that use different aliases to identify the same namespace, as used in the XML file with which an intelligent structure model was created.

Data Types

Effective in version 10.2.2, and starting with the Winter 2019 March release of Informatica Intelligent Cloud Services, when a complex file reader uses an intelligent structure model, Intelligent Structure Discovery passes the data types to the output data ports.

For example, when Intelligent Structure Discovery detects that a field contains a date, it passes the data to the output data ports as a date, not as a string.

Field Names

Effective in version 10.2.2, and starting with the Winter 2019 March release of Informatica Intelligent Cloud Services, field names in complex file data objects that you import from an intelligent structure model can begin with numbers and reserved words, and can contain the following special characters: \. [] {} () * + - ? . ^ \$ /

When a field begins with a number or reserved word, the Big Data Management mapping adds an underscore (_) to the beginning of the field name. For example, if a field in an intelligent structure model begins with OR, the mapping imports the field as _OR. When the field name contains a special character, the mapping converts the character to an underscore.

Processing Large XML Files

Effective in version 10.2.2, Intelligent Structure Discovery can stream XML files and process data for repeating elements in chunks. This makes the processing of large XML files more efficient.

Data Drift

Effective in version 10.2.2, and starting with the Winter 2019 March release of Informatica Intelligent Cloud Services, Intelligent Structure Discovery enhances the handling of data drifts.

In Intelligent Structure Discovery, data drifts occur when the input data contains fields that the sample file did not contain. In this case, Intelligent Structure Discovery passes the undefined data to an unassigned data port on the target, rather than discarding the data.

Mass Ingestion

Effective in version 10.2.2, you can run an incremental load to ingest incremental data. When you run an incremental load, the Spark engine fetches incremental data based on a timestamp or an ID column and loads the incremental data to the Hive or HDFS target. If you ingest the data to a Hive target, the Spark engine can also propagate the schema changes that have been made on the source tables.

If you ingest incremental data, the Mass Ingestion Service leverages Sqoop's incremental import mode.

For more information, see the *Informatica Big Data Management 10.2.2 Mass Ingestion Guide*.

Monitoring

This section describes the new features related to monitoring in Big Data Management in version 10.2.2.

Spark Monitoring

Effective in version 10.2.2, you can view both the pre-job and post-job tasks within the Summary Statistics pane for the Spark monitoring.

For more information about the pre-job and post-job tasks, see the *Informatica Big Data Management 10.2.2 User Guide*.

Security

This section describes the new features related to security in Big Data Management in version 10.2.2.

Enterprise Security Package

Effective in version 10.2.2, Informatica supports an Azure HDInsight cluster with Enterprise Security Package.

The Enterprise Security Package uses Kerberos for authentication and Apache Ranger for authorization.

For more information about Enterprise Security Package, see the *Informatica Big Data Management 10.2.2 Administrator Guide*.

Targets

This section describes new features for targets in version 10.2.2.

HDFS Flat File Targets

Effective in version 10.2.2, you can append output data to HDFS target files and reject files. To append output data, choose to append data if the HDFS target exists.

To help you manage the files that contain appended data, the Data Integration Service appends the mapping execution ID to the names of the target files and reject files.

For more information, see the "Targets" chapter in the *Informatica Big Data Management 10.2.2 User Guide*.

Big Data Streaming

This section describes new Big Data Streaming features in version 10.2.2.

Azure Event Hubs Data Objects

Effective in version 10.2.2, you can deploy a streaming mapping that has an event hub as a source in the following distributions:

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

Cross-account IAM Role in Amazon Kinesis Connection

Effective in version 10.2.2, you can use the cross-account IAM role to authenticate an Amazon Kinesis source.

Use the cross-account IAM role to share resources in one AWS account with users in a different AWS account without creating users in each account.

For more information, see the *Informatica Big Data Streaming 10.2.2 User Guide*.

Intelligent Structure Model

Effective in version 10.2.2, you can use intelligent structure models in Big Data Streaming.

You can incorporate an intelligent structure model in a Kafka, Kinesis, or Azure Event Hubs data object. When you add the data object to a mapping, you can process any input type that the model can parse.

The data object can accept input and parse PDF forms, JSON, Microsoft Excel, Microsoft Word tables, CSV, text, or XML input files, based on the file which you used to create the model.

For more information, see the *Informatica Big Data Streaming 10.2.2 User Guide*.

Header Ports for Big Data Streaming Data Objects

Effective in version 10.2.2, some data objects contain default header ports that represent metadata associated with events. For example, the timestamp port contains the time at which the event is generated. You can use the header ports to group and process the data.

For more information about the header ports, see the *Informatica Big Data Streaming 10.2.2 User Guide*.

AWS Credential Profile in Amazon Kinesis Connection

Effective in version 10.2.2, you can use AWS credential profile based authentication in Amazon Kinesis connection.

When you create an Amazon Kinesis connection, you can enter an AWS credential profile name. The mapping accesses the AWS credentials through the profile name listed in the AWS credentials file during run time.

For more information, see the *Informatica Big Data Streaming 10.2.2 User Guide*.

Spark Structured Streaming

Effective in version 10.2.2, Big Data Streaming uses Spark Structured Streaming to process streaming data.

Spark Structured Streaming is a scalable and fault-tolerant open source stream processing engine built on the Spark engine. It can handle late arrival of streaming events and process streaming data based on source timestamp.

The Spark engine runs the streaming mapping continuously. It reads the data, divides the data into micro batches, processes the micro batches, publishes the results, and then writes to a target.

For more information, see the *Informatica Big Data Streaming 10.2.2 User Guide*.

Window Transformation

Effective in version 10.2.2, you can use the following features when you create a Window transformation:

Watermark Delay

The watermark delay defines threshold time for a delayed event to be accumulated into a data group.

Watermark delay is a threshold where you can specify the duration at which late arriving data can be grouped and processed. If an event data arrives within the threshold time, the data is processed, and the data is accumulated into the corresponding data group.

Window Port

The window port specifies the column that contains the timestamp values based on which you can group the events. The accumulated data contains the timestamp value. Use the Window Port column to group the event time data that arrives late.

For more information, see *Informatica Big Data Streaming 10.2.2 User Guide*.

Command Line Programs

This section describes new commands in version 10.2.2.

infacmd dis Commands

This topic describes new infacmd dis command options and execution options.

Update Service Options

The Update Service Options command updates Data Integration Service properties.

The following table describes new infacmd dis updateServiceOptions command options:

New Option	Description
-RecycleMode -rm	Optional. Recycle mode restarts the service and applies the latest service and service process properties. Select Abort or Complete. <ul style="list-style-type: none">- Complete. Stops all applications and cancels all jobs within each application. Waits for all jobs to cancel before disabling the service.- Abort. Stops all applications and tries to cancel all jobs before aborting them and disabling the service. Default: Complete

The following table describes new infacmd dis updateServiceOptions command execution options:

New Option	Description
ExecutionOptions.BigDataJobRecovery	Enable big data job recovery. Set to "true" or "false." Default: false.

For more information, see the "infacmd dis Command Reference" chapter in the *Informatica 10.2.2 Command Reference*.

infacmd ihs Commands

The following table describes new infacmd ihs command options:

New Option	Description
-PrimaryNode -nm	Optional. Primary node on which the service runs.
-BackupNodes -bn	Optional. Nodes on which the service can run if the primary node is unavailable. You can configure backup nodes if you have high availability.

The following table describes new infacmd ihs commands:

Command	Description
cleanCluster	Cleans the Informatica Cluster Service.

For more information, see the "infacmd ihs Command Reference" chapter in the *Informatica 10.2.2 Command Reference*.

infacmd ipc Commands

The following table describes obsolete infacmd ipc commands:

Command	Description
ExportToPC	Exports objects from the Model repository or an export file and converts them to PowerCenter objects.

infacmd Idm Commands

The following table describes new infacmd Idm command options:

New Option	Description
-PrimaryNode -nm	Optional. If you want to configure high availability for Enterprise Data Catalog, specify the primary node name.
-BackupNodes -bn	Optional. If you want to configure high availability for Enterprise Data Catalog, specify a list of comma-separated backup node names.
-isNotifyChangeEmailEnabled -cne	Optional. Specify True if you want to enable asset change notifications. Default is False.

New Option	Description
-ExtraJarsPath -ejp	Optional. Path to the directory on the machine where you installed Informatica domain. The directory must include the JAR files required to deploy Enterprise Data Catalog on an existing cluster with WANDisco Fusion.
-ExtraJarsPath -ejp	Optional. Path to the directory on the machine where you installed Informatica domain. The directory must include the JAR files required to deploy Enterprise Data Catalog on an existing cluster with WANDisco Fusion.

The following table describes new infacmd Idm commands:

Command	Description
collectAppLogs	Collects log files for YARN applications that run to enable the Catalog Service.
publishArchive	Creates a resource in offline mode and runs the scan.

For more information, see the "infacmd Idm Command Reference" chapter in the *Informatica 10.2.2 Command Reference*.

infacmd mi Commands

The following table describe changes to infacmd mi commands:

Command	Change Description
createService	Effective in version 10.2.2, you can use the -HttpsPort, -KeystoreFile, and -KeystorePassword options to specify whether the Mass Ingestion Service processes use a secure connection to communicate with external components.
extendedRunStats	Effective in version 10.2.2, you must use the -RunID option to specify the RunID of the mass ingestion specification and the -SourceName option to specify the name of a source table to view the extended run statistics for the source table. If the source table was ingested using an incremental load, the run statistics show the incremental key and the start value. Previously, you specified the JobID for the ingestion mapping job that ingested the source table. If you upgrade to 10.2.2, you must update any scripts that run infacmd mi extendedRunStats to use the new options.
listSpecRuns	Effective in version 10.2.2, the command additionally returns the load type that the Spark engine uses to run a mass ingestion specification.
runSpec	Effective in version 10.2.2, you can use the -LoadType option to specify the load type to run a mass ingestion specification. The load type can be a full load or an incremental load.

For more information, see the "infacmd mi Command Reference" chapter in the *Informatica 10.2.2 Command Reference*.

infacmd ms Commands

The following table describes new infacmd ms commands:

Command	Description
abortAllJobs	Aborts all deployed mapping jobs that are configured to run on the Spark engine. You can choose to abort queued jobs, running jobs, or both.
createConfigurationWithParams	Creates a cluster configuration through cluster parameters that you specify in the command line.
listMappingOptions	Lists mapping options in an application.
purgeDatabaseWorkTables	Purges all job information from the queue when you enable big data recovery for the Data Integration Service.
updateMappingOptions	Updates mapping options in an application.
updateOptimizationLevel	Updates optimization level for multiple mappings in an application.

For more information, see the "infacmd ms Command Reference" chapter in the *Informatica Command Reference*.

infacmd oie Commands

Effective in version 10.2.2, the oie plugin is deprecated and support for the plugin will be removed in a future release. The infacmd oie commands have migrated to the tools plugin. For details, see ["infacmd tools Commands" on page 70](#).

infacmd tools Commands

Effective in version 10.2.2, the tools infacmd plugin replaces the oie plugin. The tools plugin performs the object import and export operations that the oie plugin performs, and you can use it to perform some additional operations.

The following table lists the infacmd oie commands that have migrated to the tools plugin:

Previous Command	Current Command
infacmd oie deployApplication	infacmd tools deployApplication
infacmd oie exportObjects	infacmd tools exportObjects
infacmd oie exportResources	infacmd tools exportResources
infacmd oie importObjects	infacmd tools importObjects

The following table describes new infacmd tools commands:

Command	Description
patchApplication	Deploys an application patch using a .piar file to a Data Integration Service.

For more information, see the "infacmd tools Command Reference" chapter in the *Informatica 10.2.2 Command Reference*.

infasetup Commands

The following table describes changed infasetup commands:

Command	Description
DefineDomain	Effective in 10.2.2, the -spid option is added to the DefineDomain command.
updateDomainSamlConfig	Effective in 10.2.2, the -spid option is added to the updateDomainSamlConfig command.

For more information, see the "infasetup Command Reference" chapter in the *Informatica 10.2.2 Command Reference*.

Enterprise Data Catalog

This section describes new Enterprise Data Catalog features in version 10.2.2.

Automatically Assign Business Title to a Column

Effective in version 10.2.2, Enterprise Data Catalog infers a business term as the business title for column assets. A business term with the confidence score of 80 percent or more is inferred as a business title of a column. The confidence score is calculated by an internal algorithm based on accepted business terms on data domains, column name similarity, and name similarity between the column and business term.

For more information, see the "Perform Asset Tasks" chapter in the *Informatica 10.2.2 Enterprise Data Catalog User Guide*.

User Collaboration on Assets

Effective in version 10.2.2, you can collaborate with other Enterprise Data Catalog users on assets. Collaboration on assets provides you the ability to interact with the other users, share insights about the assets, ask queries related to the assets, follow up on all the asset changes, and certify the assets.

You can collaborate on assets in the following ways:

Follow assets

You can follow assets to monitor asset changes in the catalog. Follow an asset to be informed about the changes that other users make to the asset, so that you can monitor the asset and take necessary actions.

Rate and review asset

You can rate and review assets based on a five-star scale in the catalog. Rate and review an asset to provide feedback about the asset based on different aspects of the asset, such as the quality, applicability, usability, and availability of the asset.

Asset queries

You can ask questions about an asset if you want a better understanding about the asset in the catalog. Ask questions that are descriptive, exploratory, predictive, or causal in nature.

Certify asset

You can certify an asset to endorse it so that other users can use the asset as a trustworthy one over the assets that are not certified.

For more information, see the "User Collaboration on Assets" chapter in the *Informatica 10.2.2 Enterprise Data Catalog Guide*.

Create Enterprise Data Catalog Application Services Using the Installer

Effective in version 10.2.2, you can use the installer to create the Enterprise Data Catalog application services after you install Enterprise Data Catalog. You can use the installer if you had installed Enterprise Data Catalog without creating the application services.

For more information about using the installer to create the application services, see the *Informatica Enterprise Data Catalog 10.2.2 Installation and Configuration Guide*.

Custom Metadata Validation Utility

Effective in version 10.2.2, you can use a stand-alone Java-based Validation Command Line Utility to validate the syntax and semantics of custom metadata that you want to ingest in the catalog. Custom metadata represents metadata that you want to ingest from custom data sources for which Enterprise Data Catalog does not provide a resource.

For more information about using the utility, see the KB article *How To: Validate Custom Metadata Before Ingesting it in the Catalog*. Contact Informatica Global Customer Support for instructions to download the utility.

Change Notifications

Effective in version 10.2.2, Enterprise Data Catalog shows notifications when changes are made to assets that you follow. The notification types include application notifications, change email notification, and digest email notification.

For more information, see the "User Collaboration on Assets" chapter in the *Informatica 10.2.2 Enterprise Data Catalog Guide*.

Business Glossary Assignment Report

Effective in version 10.2.2, you can accept or reject multiple recommended business terms for a resource.

For more information, see the "Perform Asset Tasks" chapter in the *Informatica 10.2.2 Enterprise Data Catalog Guide*.

Operating System Profiles

Effective in version 10.2.2, you can choose an operating system profile if you do not have a default operating system profile. The Data Integration Service uses the operating system profile user credentials to perform data discovery. Data discovery includes column profiles and data domain discovery profiles.

For more information about using the operating system profiles in Enterprise Data Catalog, see the "Enterprise Data Catalog Concepts" chapter in the *Informatica 10.2.2 Catalog Administrator Guide*.

REST APIs

Effective in version 10.2.2, you can use the following Informatica Enterprise Data Catalog REST APIs:

- Business Terms REST APIs. You can return, update, or delete an accepted, inferred, or rejected business term.
- Catalog Events REST APIs. You can access, update, or delete the user configuration, email configuration, and user subscriptions.
- Object Certification APIs. You can list, update, and delete the certification properties for an object.
- Object Comments APIs. You can list, create, update, and delete comments, replies, and votes for a data object.
- Object Reviews APIs. You can list, create, update, and delete reviews, ratings, and votes for a review.

For more information about the REST APIs, see the *Informatica 10.2.2 Enterprise Data Catalog REST API Reference*.

Source Metadata and Data Profile Filter

Effective in version 10.2.2, you can use the source metadata filter and data profile filter to specify source tables and views in a resource run. When you use these filters, Enterprise Data Catalog extracts source metadata and profile metadata from specific tables and views.

For more information about the source metadata and data profile filter, see the "Managing Resources" chapter in the *Informatica 10.2.2 Catalog Administrator Guide*.

Scanner Utility

Effective in version 10.2.2, Informatica provides a standalone scanner utility that you can use to extract metadata from offline and inaccessible resources. The utility contains a script that you need to run along with the associated commands in a sequence.

For more information about the standalone scanner utility, see the "Metadata Extraction from Offline and Inaccessible Resources" appendix in the *Informatica 10.2.2 Catalog Administrator Guide*.

Resource Types

Effective in version 10.2.2, you can create resources for the following data source types:

Google BigQuery

You can extract metadata, relationship, and lineage information from the following assets in a Google BigQuery data source:

- Project
- Dataset

- Table
- View

For more information about configuring a Google BigQuery data source, see the *Informatica 10.2.2 Catalog Administrator Guide*.

Workday

You can extract metadata, relationship, and lineage information from the following assets in a Workday data source:

- Service
- Entity
- Report
- Operation
- Data source
- Property
- Business objects

For more information about configuring a Workday data source, see the *Informatica 10.2.2 Catalog Administrator Guide*.

Enterprise Data Lake

This section describes new Enterprise Data Lake features in version 10.2.2.

Apply Active Rules

Effective in version 10.2.2, you can use active rules in projects.

Active rules are mapplets developed using the Developer tool. You can use active rules to apply complex transformations such as aggregator and Data Quality transformations to worksheets for matching and consolidation.

An active rule uses all rows within a data set as input. You can select multiple worksheets to use as inputs to the rule. The application adds a worksheet containing the rule output to the project.

For more information, see the "Prepare Data" chapter in the *Informatica 10.2.2 Enterprise Data Lake User Guide*.

Delete Duplicate Rows

Effective in version 10.2.2, you can delete rows containing duplicate values from a worksheet.

For more information, see the "Prepare Data" chapter in the *Informatica 10.2.2 Enterprise Data Lake User Guide*.

Cluster and Categorize Column Data

Effective in version 10.2.2, you can cluster similar values in a column, and then categorize the values based on recommendations from Enterprise Data Lake. The application uses a phonetic algorithm to cluster similar

values, and then suggests that you replace the less frequently occurring values with the most frequently occurring value.

For more information, see the "Prepare Data" chapter in the *Informatica 10.2.2 Enterprise Data Lake User Guide*.

CLAIRE-based Recommendations

Effective in version 10.2.2, the application uses the embedded CLAIRE machine learning discovery engine to provide recommendations when you prepare data.

When you view the Project page, the application displays alternate and additional recommendations derived from upstream data sources based on data lineage, as well as documented primary-foreign key relationships.

When you select a column in a worksheet during data preparation, the application displays suggestions to improve the data based on the column data type in the Column Overview panel.

When you perform a join operation on two worksheets, the application utilizes primary-foreign key relationships to indicate incompatible sampling when low overlap for desired key pairs occurs.

For more information, see the "Prepare Data" chapter in the *Informatica 10.2.2 Enterprise Data Lake User Guide*.

Conditional Aggregation

Effective in 10.2.2, you can use AND and OR logic to apply multiple conditions on IF calculations that you use when you create an aggregate worksheet in a project.

- Use AND with all operators to include more than one column in a condition.
- Use OR with the IS, IS NOT and IS BETWEEN operators to include more than one value within a column in a condition.

For more information, see the "Prepare Data" chapter in the *Informatica 10.2.2 Enterprise Data Lake User Guide*.

Data Masking

Effective in version 10.2.2, Enterprise Data Lake integrates with Informatica Dynamic Data Masking, a data security product, to enable masking of sensitive data in data assets.

To enable data masking in Enterprise Data Lake, you configure the Dynamic Data Masking Server to apply masking rules to data assets in the data lake. You also configure the Informatica domain to enable Enterprise Data Lake to connect to the Dynamic Data Masking Server.

Dynamic Data Masking intercepts requests sent to the data lake from Enterprise Data Lake, and applies the masking rules to columns in the requested asset. When Enterprise Data Lake users view or perform operations on columns containing masked data, the actual data is fully or partially obfuscated based on the masking rules applied.

For more information, see the "Masking Sensitive Data" chapter in the *Informatica 10.2.2 Enterprise Data Lake Administrator Guide*.

Localization

Effective in version 10.2.2, the user interface supports Japanese. You can also use non-Latin characters in project names and descriptions.

Partitioned Sources and Targets

Effective in version 10.2.2, Enterprise Data Lake can read data from partitioned sources during import, publish, or copy operations. The application can also append data to partitioned targets in the data lake during import, publish, copy, or upload operations.

Add Comments to Recipe Steps

Effective in version 10.2.2, you can add a comment to a recipe step. Use comments to improve collaboration and provide details to meet auditing requirements.

For more information, see the "Prepare Data" chapter in the *Informatica 10.2.2 Enterprise Data Lake User Guide*.

Save a Recipe as a Mapping

Effective in version 10.2.2, you can save a recipe as a mapping, instead of publishing the recipe and creating a new output table.

You can save the mapping to the Model repository associated with the Enterprise Data Lake Service, or you can save the mapping to an .xml file. Developers can use the Developer tool to review and modify the mapping, and then execute the mapping when appropriate based on system resource availability.

For more information, see the "Prepare Data" chapter in the *Informatica 10.2.2 Enterprise Data Lake User Guide*.

Amazon S3, ADLS, WASB, MapR-FS as Data Sources

Effective in version 10.2.2, you can prepare data in files stored in the following data sources:

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

You must create a resource in Enterprise Data Catalog for each data source containing data that you want to prepare. A resource is a repository object that represents an external data source or metadata repository. Scanners attached to a resource extract metadata from the resource and store the metadata in Enterprise Data Catalog.

For more information about creating resources in Enterprise Data Catalog, see the "Managing Resources" chapter in the *Informatica 10.2.2 Catalog Administrator Guide*.

Statistical Functions

Effective in version 10.2.2, you can apply the following statistical functions to columns in a worksheet when you prepare data:

- AVG
- AVGIF
- COUNT
- COUNTIF
- COUNTDISTINCT

- COUNTDISTINCTIF
- MAX
- MAXIF
- MIN
- MINIF
- STDDEV
- STDDEVIF
- SUM
- SUMIF
- VARIANCE
- VARIANCEIF

For more information, see the "Prepare Data" chapter in the *Informatica 10.2.2 Enterprise Data Lake User Guide*.

Date and Time Functions

Effective in version 10.2.2, you can apply the following date and time functions to columns in a worksheet when you prepare data:

- ADD_TO_DATE
- CURRENT_DATETIME
- DATETIME
- DATE_DIFF
- DATE_TO_UNIXTIME
- EXTRACT_MONTH_NAME
- UNIXTIME_TO_DATE
- Convert Date to Text
- Convert Text to Date

For more information, see the "Prepare Data" chapter in the *Informatica 10.2.2 Enterprise Data Lake User Guide*.

Math Functions

Effective in version 10.2.2, you can apply the following math functions to columns when you prepare data:

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

For more information, see the "Prepare Data" chapter in the *Informatica 10.2.2 Enterprise Data Lake User Guide*.

Text Functions

Effective in version 10.2.2, you can apply the following text functions to columns when you prepare data:

- 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
- Convert Date to Text
- Convert Number to Text
- Convert Text to Date
- Convert Text to Number

For more information, see the "Prepare Data" chapter in the *Informatica 10.2.2 Enterprise Data Lake User Guide*.

Window Functions

Effective in version 10.2.2, you can use window functions to perform operations on groups of rows within a worksheet. The group of rows on which a function operates is called a window, which you define with a partition key, an order by key, and optional offsets. A window function calculates a return value for every input row within the context of the window.

You can use window functions to perform the following tasks:

- Retrieve data from previous or subsequent rows.
- Calculate a cumulative sum or a cumulative average based on a group of rows.
- Assign a sequential row number to each row in a group of rows.
- Replace null values in rows with the preceding non-null value within a group of rows.
- Generate session identifiers that you can use to group rows based on a specific time period, such as web site visits recorded in a log file.

You can apply multiple window functions to a worksheet. For example, you might apply a function to calculate the sum of values for each row following the current row within a window, and then apply another function to calculate the average of the same values.

Enterprise Data Lake adds a column containing the results of each function you apply to the worksheet.

For more information, see the "Prepare Data" chapter in the *Informatica 10.2.2 Enterprise Data Lake User Guide*.

Purge Audit Events

Effective in version 10.2.2, you can run the `infacmd edl purgeevents` command to delete user activity events from the audit history database. You can optionally run the command to delete project history events from the database.

Spark Execution Engine

Effective in version 10.2.2, Enterprise Data Lake uses the Spark engine for high resource consumption activities such as asset publication, and to run active rule mapplets that use the Python transformation. Using the Spark engine for high resource consumption activities provides better performance, and enables an Enterprise Data Lake deployment on Amazon Elastic MapReduce (EMR) to take advantage of autoscaling.

Informatica Developer

This section describes new Developer tool features in version 10.2.2.

Applications

Effective in version 10.2.2, you can create incremental applications. An incremental application is an application that you can update by deploying an application patch to update a subset of application objects. The Data Integration Service updates the objects in the patch while other application objects continue running.

If you upgrade to version 10.2.2, existing applications are labeled "full applications." You can continue to create full applications in version 10.2.2, but you cannot convert a full application to an incremental application.

For more information, see the "Application Deployment" and the "Application Patch Deployment" chapters in the *Informatica 10.2.2 Developer Tool Guide*.

Informatica Mappings

This section describes new Informatica mapping features in version 10.2.2.

Data Types

Effective in version 10.2.2, you can enable high-precision mode in batch mappings that run on the Spark engine. The Spark engine can process decimal values with up to 38 digits of precision.

For more information, see the *Informatica Big Data Management 10.2.2 User Guide*.

Mapping Outputs

Effective in version 10.2.2, you can use mapping outputs in batch mappings that run as Mapping tasks in workflows on the Spark engine. You can persist the mapping outputs in the Model repository or bind the mapping outputs to workflow variables.

For more information, see the "Mapping Outputs" chapter in the *Informatica 10.2.2 Developer Mapping Guide* and the "Mapping Task" chapter in the *Informatica 10.2.2 Developer Workflow Guide*.

Mapping Parameters

Effective in version 10.2.2, you can assign expression parameters to port expressions in Aggregator, Expression, and Rank transformations that run in the native and non-native environments.

For more information, see the "Where to Assign Parameters" and "Dynamic Mappings" chapters in the *Informatica 10.2.2 Developer Mapping Guide*.

Optimizer Levels

Effective in version 10.2.2, you can configure the Auto optimizer level for mappings and mapping tasks. With the Auto optimization level, the Data Integration Service applies optimizations based on the execution mode and mapping contents.

The optimizer level default for new mappings is Auto.

When you upgrade to version 10.2.2, optimizer levels configured in mappings remain the same. To use the Auto optimizer level with upgraded mappings, you must manually change the optimizer level.

For more information, see the "Optimizer Levels" chapter in the *Informatica 10.2.2 Developer Mapping Guide*.

Sqoop

Effective in version 10.2.2, you can use the following new Sqoop features:

Incremental data extraction support

You can configure a Sqoop mapping to perform incremental data extraction based on an ID or timestamp. With incremental data extraction, Sqoop extracts only the data that changed since the last data extraction. Incremental data extraction increases the mapping performance.

Vertica connectivity support

You can configure Sqoop to read data from a Vertica source or write data to a Vertica target.

Spark engine optimization for Sqoop pass-through mappings

When you run a pass-through mapping with a Sqoop source on the Spark engine, the Data Integration Service optimizes mapping performance in the following scenarios:

- You write data to a Hive target that was created with a custom DDL query.
- You write data to an existing Hive target that is either partitioned with a custom DDL query or partitioned and bucketed with a custom DDL query.
- You write data to an existing Hive target that is both partitioned and bucketed.

--infoownername argument support

You can configure the --infoownername argument to indicate whether Sqoop must honor the owner name for a data object.

For more information, see the *Informatica Big Data Management 10.2.2 User Guide*.

Informatica Transformations

This section describes new features in Informatica transformations in version 10.2.2.

Address Validator Transformation

This section describes new Address Validator transformation features.

The Address Validator transformation contains additional address functionality for the following countries:

All Countries

Effective in version 10.2.2, the Address Validator transformation supports single-line address verification in every country for which Informatica provides reference address data.

In earlier versions, the transformation supported single-line address verification for 26 countries.

To verify a single-line address, enter the address in the Complete Address port. If the address identifies a country for which the default preferred script is not a Latin or Western script, use the default Preferred Script property on the transformation with the address.

Australia

Effective in version 10.2.2, you can configure the Address Validator transformation to add address enrichments to Australia addresses. You can use the enrichments to discover the geographic sectors and regions to which the Australia Bureau of Statistics assigns the addresses. The sectors and regions include census collection districts, mesh blocks, and statistical areas.

The transformation uses the following ports to deliver the enrichments:

- Census Collection District Code 2006
- Geocoded National Address File Identifier
- Greater Capital City Statistical Area 5-Digit
- Greater Capital City Statistical Area Name
- Level One Statistical Area 11-Digit
- Level One Statistical Area 7-Digit
- Level Two Statistical Area 9-Digit
- Level Two Statistical Area 5-Digit
- Level Two Statistical Area Name
- Level Three Statistical Area 5-Digit
- Level Three Statistical Area Name
- Level Four Statistical Area 3-Digit
- Level Four Statistical Area Name
- Mesh Block 11-Digit 2011
- Mesh Block 11-Digit 2016
- State or Territory Code
- State or Territory Name
- Supplementary AU Status

Find the ports in the AU Supplementary port group.

Bolivia

Effective in version 10.2.2, the Address Validator transformation improves the parsing and validation of Bolivia addresses. Additionally, Informatica updates the reference data for Bolivia.

The transformation also includes the following improvements for Bolivia:

- Address validation to street level.
- Geocoordinates at street mid-point level for addresses in major cities.

Canada

Informatica introduces the following features and enhancements for Canada:

Support for the Global Preferred Descriptor property in Canada Addresses

Effective in version 10.2.2, you can configure the Address Validator transformation to return the short or long form of an element descriptor.

The transformation can return the short or long form of the following descriptors:

- Street descriptors
- Directional values
- Building descriptors
- Sub-building descriptors

To specify the output format for the descriptors, configure the Global Preferred Descriptor property on the transformation. The property applies to English-language and French-language descriptors. By default, the transformation returns the descriptor in the format that the reference data specifies. If you select the PRESERVE INPUT option on the property, the Preferred Language property takes precedence over the Global Preferred Descriptor property.

Support for CH and CHAMBER as Sub-Building Descriptors

Effective in version 10.2.2, the Address Validator transformation recognizes CH and CHAMBER as sub-building descriptors in Canada addresses.

Colombia

Effective in version 10.2.2, the Address Validator transformation improves the processing of street data in Colombia addresses. Additionally, Informatica updates the reference data for Colombia.

France

Effective in version 10.2.2, Informatica introduces the following improvements for France addresses:

- Informatica improves the supplementary reference data for France.
- The Address Validator transformation assigns addresses to IRIS units in France with greater accuracy. The transformation uses the house number in the address to verify the IRIS unit to which the address belongs. The use of house numbers can improve the assignment accuracy when the address lies close to the border between different units.

Israel

Effective in version 10.2.2, Informatica introduces the following features and enhancements for Israel:

Multilanguage Support for Israel Addresses

You can configure the Address Validator transformation to return an Israel address in the English language or the Hebrew language.

Use the Preferred Language property to select the preferred language for the addresses that the transformation returns.

The default language for Israel addresses is Hebrew. To return address information in Hebrew, set the Preferred Language property to DATABASE or ALTERNATIVE_1. To return the address information in English, set the property to ENGLISH or ALTERNATIVE_2.

Support for Multiple Character Sets for Israel Addresses

The Address Validator transformation can read and write Israel addresses in Hebrew and Latin character sets.

Use the Preferred Script property to select the preferred character set for the address data.

The default character set for Israel addresses is Hebrew. When you set the Preferred Script property to Latin or Latin-1, the transformation transliterates Hebrew address data into Latin characters.

Peru

Effective in version 10.2.2, the Address Validator transformation validates a Peru address to house number level. Additionally, Informatica updates the reference data for Peru.

Sweden

Effective in version 10.2.2, the Address Validator transformation improves the verification of street names in Sweden addresses.

The transformation improves the verification of street names in the following ways:

- The transformation can recognize a street name that ends in the character G as an alias of the same name with the final characters GATAN.
- The transformation can recognize a street name that ends in the character V as an alias of the same name with the final characters VÄGEN.
- The Address Validator transformation can recognize and correct a street name with an incorrect descriptor when either the long form or the short form of the descriptor is used.

For example, The transformation can correct RUNIUSV or RUNIUSVÄGEN to RUNIUSGATAN in the following address:

RUNIUSGATAN 7

SE-112 55 STOCKHOLM

United States

Effective in version 10.2 HotFix 2, you can configure the Address Validator transformation to identify United States addresses that do not receive mail on one or more days of the week.

To identify the addresses, use the *Non-Delivery Days* port. The port contains a seven-digit string that represents the days of the week from Sunday through Saturday. Each position in the string represents a different day.

The Address Validator transformation returns the first letter of a weekday in the corresponding position on the port if the address does not receive mail on that day. The transformation returns a dash symbol in the corresponding position for other days of the week.

For example, a value of S----FS on the Non-Delivery Days port indicates that an address does not receive mail on Sunday, Friday, and Saturday.

Find the Non-Delivery Days port in the US Specific port group in the Basic model. To receive data on the Non-Delivery Days port, run the Address Validator transformation in certified mode. The transformation reads the port values from the USA5C129.MD and USA5C130.MD database files.

For comprehensive information about the features and operations of the address verification software engine in version 10.2.2, see the *Informatica Address Verification 5.14.0 Developer Guide*.

Update Strategy Transformation

Effective in version 10.2.2, you can use an Update Strategy transformation in a mapping that runs on the Spark engine to update relational targets.

Previously, you could use an Update Strategy transformation in a mapping that runs on the Spark engine only to update Hive targets.

For more information, see the Update Strategy transformation chapter in the *Developer Transformation Guide*.

PowerExchange Adapters for Informatica

This section describes new Informatica adapter features in version 10.2.2.

PowerExchange for Amazon Redshift

Effective in version 10.2.2, PowerExchange for Amazon Redshift includes the following features:

- You can read data from or write data to the following regions:
 - China(Ningxia)
 - EU(Paris)
- You can use Amazon Redshift objects as dynamic sources and target in a mapping.
- You can use octal values of printable and non-printable ASCII characters as a DELIMITER or QUOTE.
- You can enter pre-SQL and post-SQL commands to run queries for source and target objects in a mapping.
- You can define an SQL query for read data objects in a mapping to override the default query. You can enter an SQL statement supported by the Amazon Redshift database.
- You can specify the maximum size of an Amazon S3 object in bytes when you download large Amazon S3 objects in multiple parts.
- You can read unique values when you read data from an Amazon Redshift source.
- When you upload an object to Amazon S3, you can specify the minimum size of the object and the number of threads to upload the objects in parallel as a set of independent parts.
- You can choose to retain an existing target table, replace a target table at runtime, or create a new target table if the table does not exist in the target.
- You can configure the Update Strategy transformations for an Amazon Redshift target in the native environment.
- When you write data to Amazon Redshift, you can override the Amazon Redshift target table schema and the table name during run time.
- When the connection type is ODBC, the Data Integration Service can push transformation logic to Amazon Redshift sources and targets using source-side and full pushdown optimization.
- You can use Server-Side Encryption with AWS KMS (AWS Key Management Service) on Amazon EMR version 5.16 and Cloudera CDH version 5.15 and 5.16.
- PowerExchange for Amazon Redshift supports AWS SDK for Java version 1.11.354.

For more information, see the *Informatica PowerExchange for Amazon Redshift 10.2.2 User Guide*.

PowerExchange for Amazon S3

Effective in version 10.2.2, PowerExchange for Amazon S3 includes the following features:

- You can read data from or write data to the following regions:
 - China(Ningxia)
 - EU(Paris)
 - AWS GovCloud (US)
- You can use Amazon S3 objects as dynamic sources and target in a mapping.
- When you run a mapping in the native environment or on the Spark engine to read data from an Avro, flat, JSON, ORC, or Parquet file, you can use wildcard characters to specify the source directory name or the source file name.
- You can add a single or multiple tags to the objects stored on the Amazon S3 bucket to categorize the objects. Each tag contains a key value pair. You can either enter the key value pairs or specify the absolute file path that contains the key value pairs.
- You can specify the maximum threshold size to download an Amazon S3 object in multiple parts.
- When you upload an object to Amazon S3, you can specify the minimum size of the object and the number of threads to upload the objects in parallel as a set of independent parts.
- When you create a data object read or write operation, you can read data present in the FileName port that contains the endpoint name and source path of the file.
- You can add new columns or modify the columns in the Port tab directly when you create a data object read or write operation.
- You can copy the columns of the source transformations, target transformations, or any other transformations from the Port tab and paste the columns in the data object read or write operation directly when you create a mapping to read or write an Avro, JSON, ORC, or Parquet file.
- You can update the Amazon S3 file format without losing the column metadata in the Schema field of the column projection properties even after you configure the column projection properties for another Amazon S3 file format.
- You can use Server-Side Encryption with AWS KMS (AWS Key Management Service) on Amazon EMR version 5.16 and Cloudera CDH version 5.15 and 5.16.
- PowerExchange for Amazon S3 supports AWS SDK for Java version 1.11.354.

For more information, see the *Informatica PowerExchange for Amazon S3 10.2.2 User Guide*.

PowerExchange for Google BigQuery

Effective in version 10.2.2, you can create a Google BigQuery target using the right-click **Create Target** option.

For more information, see the *Informatica PowerExchange for Google BigQuery 10.2.2 User Guide*.

PowerExchange for HBase

Effective in version 10.2.2, PowerExchange for HBase includes the following new features:

- When you create an HBase data object, you can select an operating system profile to increase security and to isolate the design-time user environment when you import and preview metadata from a Hadoop cluster.

Note: You can choose an operating system profile if the Metadata Access Service is configured to use operating system profiles. The Metadata Access Service imports the metadata with the default operating system profile assigned to the user. You can change the operating system profile from the list of available operating system profiles.

- You can use the HBase objects as dynamic sources and targets in a mapping.
- You can run a mapping on the Spark engine to look up data in an HBase resource.

For more information, see the *Informatica PowerExchange for HBase 10.2.2 User Guide*.

PowerExchange for HDFS

Effective in version 10.2.2, PowerExchange for HDFS includes the following new features:

- When you create a complex file data object, you can select an operating system profile to increase security and to isolate the design-time user environment when you import and preview metadata from a Hadoop cluster.

Note: You can choose an operating system profile if the Metadata Access Service is configured to use operating system profiles. The Metadata Access Service imports the metadata with the default operating system profile assigned to the user. You can change the operating system profile from the list of available operating system profiles.

- When you run a mapping in the native environment or on the Spark engine to read data from a complex file data object, you can use wildcard characters to specify the source directory name or the source file name.

You can use the following wildcard characters:

? (Question mark)

The question mark character (?) allows one occurrence of any character.

* (Asterisk)

The asterisk mark character (*) allows zero or more than one occurrence of any character.

- You can use complex file objects as dynamic sources and targets in a mapping.
- You can use complex file objects to read data from and write data to a complex file system.
- When you run a mapping in the native environment or on the Spark engine to write data to a complex file data object, you can overwrite target data, the Data Integration Service deletes the target data before writing new data.
- When you create a data object read or write operation, you can read the data present in the FileName port that contains the endpoint name and source path of the file.
- You can now view the data object operations immediately after you create the data object read or write operation.
- You can add new columns or modify the columns, when you create a data object read or write operation.
- You can copy the columns of the source transformations, target transformations, or any other transformations and paste the columns in the data object read or write operation directly when you read or write to an Avro, JSON, ORC, or Parquet file.

For more information, see the *Informatica PowerExchange for HDFS 10.2.2 User Guide*.

PowerExchange for Hive

Effective in version 10.2.2, PowerExchange for Hive includes the following new features:

- You can configure the following target schema strategy options for a Hive target:
 - RETAIN - Retain existing target schema
 - CREATE - Create or replace table at run time
 - APPLYNEWCOLUMNS - Alter table and apply new columns only
 - APPLYNEWSHEMA - Alter table and apply new schema
 - FAIL - fail mapping if target schema is different
 - Assign Parameter
- You can truncate an internal or external partitioned Hive target before loading data. This option is applicable when you run the mapping in the Hadoop environment.
- You can create a read or write transformation for Hive in native mode to read data from Hive source or write data to Hive target.
- When you write data to a Hive target, you can configure the following properties in a Hive connection:
 - Hive Staging Directory on HDFS. Represents the HDFS directory for Hive staging tables. This option is applicable and required when you write data to a Hive target in the native environment.
 - Hive Staging Database Name. Represents the namespace for Hive staging tables. This option is applicable when you run a mapping in the native environment to write data to a Hive target. If you run the mapping on the Blaze or Spark engine, you do not need to configure the Hive staging database name in the Hive connection. The Data Integration Service uses the value that you configure in the Hadoop connection.

For more information, see the *Informatica PowerExchange for Hive 10.2.2 User Guide*.

PowerExchange for MapR-DB

Effective in version 10.2.2, when you create an HBase data object for MapR-DB, you can select an operating system profile to increase security and to isolate the design-time user environment when you import and preview metadata from a Hadoop cluster.

Note: You can choose an operating system profile if the Metadata Access Service is configured to use operating system profiles. The Metadata Access Service imports the metadata with the default operating system profile assigned to the user. You can change the operating system profile from the list of available operating system profiles.

For more information, see the *Informatica PowerExchange for MapR-DB 10.2.2 User Guide*.

PowerExchange for Microsoft Azure Blob Storage

Effective in version 10.2.2, PowerExchange for Microsoft Azure Blob Storage includes the following functionality:

- You can run mappings in the Azure Databricks environment.
- You can configure the US government Microsoft Azure end-points.

- You can compress data in the following formats when you read data from or write data to Microsoft Azure Blob Storage:
 - None
 - Deflate
 - Gzip
 - Bzip2
 - Lzo
 - Snappy
- You can use Microsoft Azure Blob Storage objects as dynamic sources and targets in a mapping.
- You can read the name of the file from which the Data Integration Service reads the data at run-time in the native environment.
- You can configure the relative path in **Blob Container Override** in the advanced source and target properties.

For more information, see the *Informatica PowerExchange for Microsoft Azure Blob Storage 10.2.2 User Guide*.

PowerExchange for Microsoft Azure Cosmos DB SQL API

Effective in version 10.2.2, PowerExchange for Microsoft Azure Cosmos DB SQL API includes the following functionality:

- You can run mappings in the Azure Databricks environment. Databricks support for PowerExchange for Microsoft Azure Cosmos DB SQL API is available for technical preview. Technical preview functionality is supported but is unwarranted and is not production-ready. Informatica recommends that you use these features in non-production environments only.

For more information, see the *Informatica PowerExchange for Microsoft Azure Cosmos DB SQL API 10.2.2 User Guide*.

PowerExchange for Microsoft Azure Data Lake Store

Effective in version 10.2.2, PowerExchange for Microsoft Azure Data Lake Store includes the following functionality:

- You can run mappings in the Azure Databricks environment.
- You can use complex data types, such as array, struct, and map, in mappings that run in the Hadoop environment. With complex data types, the respective engine directly reads, processes, and writes hierarchical data in Avro, JSON, and Parquet complex files. For an intelligent structure source, you can configure only the read operation.
- You can create mappings to read and write Avro and Parquet files that contain only primitive data types in the native environment.
- You can select a directory as a source in a mapping to read multiple files from the directory.
- You can use Microsoft Azure Data Lake Store objects as dynamic sources and targets in a mapping.
- You can create a Microsoft Azure Data Lake Store target using the Create Target option.

For more information, see the *Informatica PowerExchange for Microsoft Azure Data Lake Store 10.2.2 User Guide*.

PowerExchange for Microsoft Azure SQL Data Warehouse

Effective in version 10.2.2, PowerExchange for Microsoft Azure SQL Data Warehouse includes the following functionality:

- You can run mappings in the Azure Databricks environment.
- You can configure the US government Microsoft Azure end-points in mappings that run in the native environment and on the Spark engine.
- You can generate error files in the Microsoft Azure Blob Storage container. The error files contain rejected rows and the cause for the rejected rows.
- You can define the batch size in advance target properties in the native environment.
- You can configure full pushdown optimization to push transformation logic to source databases and target databases. Use pushdown optimization to improve task performance by using the database resources.
- You can use Microsoft Azure SQL Data Warehouse objects as dynamic sources and targets in a mapping.

The full pushdown optimization and the dynamic mappings functionality for PowerExchange for Microsoft Azure SQL Data Warehouse is available for technical preview. Technical preview functionality is supported but is unwarranted and is not production-ready. Informatica recommends that you use these features in non-production environments only.

For more information, see the *Informatica PowerExchange for Microsoft Azure SQL Data Warehouse 10.2.2 User Guide*.

PowerExchange for Salesforce

Effective in version 10.2.2, PowerExchange for Salesforce includes the following new features:

- You can use version 43.0 and 44.0 of Salesforce API to create a Salesforce connection and access Salesforce objects.
- You can configure OAuth for Salesforce connections.
- You can configure the native expression filter for the source data object operation.
- You can parameterize the following read operation properties for a Salesforce data object:
 - SOQL Filter Condition
 - CDC Start Timestamp
 - CDC End Timestamp
 - PK Chunking Size
 - PK Chunking startRow ID

You can parameterize the following write operation properties for a Salesforce data object:

- Set prefix for BULK success and error files
- SFDC Success File Directory
- Set the location of the BULK error files

For more information, see the *Informatica PowerExchange for Salesforce 10.2.2 User Guide*.

PowerExchange for Snowflake

Effective in version 10.2.2, PowerExchange for Snowflake includes the following new features:

- You can configure Okta SSO authentication by specifying the authentication details in the JDBC URL parameters of the Snowflake connection.
- You can configure an SQL override to override the default SQL query used to extract data from the Snowflake source. Specify the SQL override in the Snowflake data object read operation properties.
- You can choose to compress the files before writing to Snowflake tables and optimize the write performance. In the advanced properties. You can set the compression parameter to On or Off in the **Additional Write Runtime Parameters** field in the Snowflake data object write operation advanced properties.
- The Data Integration Service uses the Snowflake Spark Connector APIs to run Snowflake mappings on the Spark engine.
- You can read data from and write data to Snowflake that is enabled for staging data in Azure or Amazon.

For more information, see the *Informatica PowerExchange for Snowflake 10.2.2 User Guide*.

PowerExchange for Teradata Parallel Transporter API

Effective in version 10.2.2, PowerExchange for Teradata Parallel Transporter API includes the following functions in the advanced target properties:

- You can specify a replacement character to use in place of an unsupported Teradata unicode character in the Teradata database while loading data to targets.
- If you specified a character used in place of an unsupported character while loading data to Teradata targets, you can specify version 8.x - 13.x or 14.x and later for the target Teradata database. Use this attribute in conjunction with the Replacement Character attribute. The Data Integration Service ignores this attribute if you did not specify a replacement character while loading data to Teradata targets.
- When you write data to Teradata, you can override the Teradata target table schema and the table name during run time.

For more information, see the *Informatica PowerExchange for Teradata Parallel Transporter API 10.2.2 User Guide*.