Informatica®
10.2.2 Service Pack 1

New Features Guide

Informatica New Features Guide
10.2.2 Service Pack 1
May 2019

Publication Date: 2019-12-18

Table of Contents

Preface

The *Informatica New Features and Enhancements Guide* is written for all Informatica software users. This guide lists the new features and enhancements in Informatica products.

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.

Informatica maintains documentation for many products on the Informatica Knowledge Base in addition to the Documentation Portal. If you cannot find documentation for your product or product version on the Documentation Portal, search the Knowledge Base at https://search.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 (10.2.2 Service Pack 1)

This chapter includes the following topics:

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 hierarchal 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 2

New Features (10.2.2)

This chapter includes the following topics:

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 infamcd 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. - 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 ldm Commands

The following table describes new infacmd ldm 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 ldm 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 ldm 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 21.

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.

--infaownername argument support

You can configure the --infaownername 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 10.2.2, 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
  - APPLYNEWSCHEMA - 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 Salesforce10.2.2 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 Snowflake10.2.2 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*.