Informatica®
10.2.1

Release Guide

Table of Contents

Preface

The *Informatica Release Guide* lists new features and enhancements, behavior changes between versions, and tasks you might need to perform after you upgrade from a previous version. The *Informatica Release Guide* is written for all types of users who are interested in the new features and changed behavior. This guide assumes that you have knowledge of the features for which you are responsible.

Informatica Resources

Informatica Network

Informatica Network hosts Informatica Global Customer Support, the Informatica Knowledge Base, and other product resources. To access Informatica Network, visit https://network.informatica.com.

As a member, you can:

- Access all of your Informatica resources in one place.
- Search the Knowledge Base for product resources, including documentation, FAQs, and best practices.
- View product availability information.
- 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 search Informatica Network for product resources such as documentation, how-to articles, best practices, and PAMs.

To access the Knowledge Base, visit https://kb.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

To get the latest documentation for your product, browse the Informatica Knowledge Base at https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx.

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

Informatica Product Availability Matrixes

Product Availability Matrixes (PAMs) indicate the versions of operating systems, databases, and other types of data sources and targets that a product release supports. If you are an Informatica Network member, you can access 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. Developed from the real-world experience of hundreds of data management projects, Informatica Velocity represents the collective knowledge of our consultants who have worked with organizations from around the world to plan, develop, deploy, and maintain successful data management solutions.

If you are an Informatica Network member, you can access 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 augment, extend, or enhance your Informatica implementations. By leveraging any of the hundreds of solutions from Informatica developers and partners, you can improve your productivity and speed up time to implementation on your projects. You can access Informatica Marketplace at https://marketplace.informatica.com.

Informatica Global Customer Support

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

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link:
http://www.informatica.com/us/services-and-training/support-services/global-support-centers.

If you are an Informatica Network member, you can use Online Support at http://network.informatica.com.

Part I: Version 10.2.1

This part contains the following chapters:

CHAPTER 1

New Features (10.2.1)

This chapter includes the following topics:

Application Services

This section describes new application service features in version 10.2.1.

Content Management Service

Effective in version 10.2.1, you can optionally specify a schema to identify reference tables in the reference data database as a property on the Content Management Service.

To specify the schema, use the Reference Data Location Schema property on the Content Management Service in Informatica Administrator. Or, run the infacmd cms updateServiceOptions command with the DataServiceOptions.RefDataLocationSchema option.

If you do not specify a schema for reference tables on the Content Management Service, the service uses the schema that the database connection specifies. If you do not explicitly set a schema on the database connection, the Content Management Service uses the default database schema.

Note: Establish the database and the schema that the Content Management Service will use for reference data before you create a managed reference table.

For more information, see the "Content Management Service" chapter in the *Informatica 10.2.1 Application Service Guide* and the "infacmd cms Command Reference" chapter in the *Informatica 10.2.1 Command Reference*.

Data Integration Service

Effective in version 10.2.1, the Data Integration Service properties include a new execution option.

JDK Home Directory

> The JDK installation directory on the machine that runs the Data Integration Service. Required to run Sqoop mappings or mass ingestion specifications that use a Sqoop connection on the Spark engine, or to process a Java transformation on the Spark engine. Default is blank.

Mass Ingestion Service

Effective in version 10.2.1, you can create a Mass Ingestion Service. The Mass Ingestion Service is an application service in the Informatica domain that manages mass ingestion specifications. You configure the mass ingestion specifications in the Mass Ingestion tool to ingest large amounts of data from a relational source to a Hive or HDFS target.

To manage mass ingestion specifications, the Mass Ingestion Service performs the following tasks:

- Manages and validates a mass ingestion specification.
- Schedules a mass ingestion job to run on a Data integration Service.
- Monitors the results and statistics of a mass ingestion job.
- Restarts a mass ingestion job.

For more information on the Mass Ingestion Service, see the "Mass Ingestion Service" chapter in the *Informatica 10.2.1 Application Service Guide*.

Metadata Access Service

Effective in version 10.2.1, you can create a Metadata Access Service. The Metadata Access Service is an application service that allows the Developer tool to access Hadoop connection information to import and preview metadata. When you import an object from a Hadoop cluster, the following adapters use Metadata Access Service to extract the object metadata at design time:

- PowerExchange for HBase
- PowerExchange for HDFS
- PowerExchange for Hive
- PowerExchange for MapR-DB

For more information, see the "Metadata Access Service" chapter in the *Informatica 10.2.1 Application Service Guide*.

Model Repository Service

Azure SQL Database as Model Repository

Effective in version 10.2.1, you can use the Azure SQL database as the Model repository.

For more information, see the "Model Repository Service" chapter in the *Informatica 10.2.1 Application Service Guide*.

Git Version Control System

Effective in version 10.2.1, you can integrate the Model repository with the Git version control system. Git is a distributed version control system. When you check out and check in an object, a copy of the version is saved to the local repository and to the Git server. If the Git server goes down, the local repository retains all the versions of the object. To use the Git version control system, enter the URL of the global repository for Git in the **URL** field, login credentials for the global repository in the **Username** and **Password** fields, and the path of the local repository for the Model Repository Service in the **VCS Local Repository Path** field.

For more information, see the "Model Repository Service" chapter in the *Informatica 10.2.1 Application Service Guide*.

Big Data Management

This section describes new Big Data Management features in version 10.2.1.

Blaze Engine Resource Conservation

Effective in version 10.2.1, you can preserve the resources that the Blaze engine infrastructure uses.

Set the infagrid.blaze.service.idle.timeout property to specify the number of minutes that the Blaze engine remains idle before releasing resources. Set the infagrid.orchestrator.svc.sunset.time property to specify the maximum number of hours for the Blaze orchestrator service. You can use the infacmd isp createConnection command, or set the property in the Blaze Advanced properties in the Hadoop connection in the Administrator tool or the Developer tool.

For more information about these properties, see the *Informatica Big Data Management 10.2.1 Administrator Guide*.

Cluster Workflows

You can use new workflow tasks to create a cluster workflow.

A cluster workflow creates a cluster on a cloud platform and runs Mapping and other workflow tasks on the cluster. You can choose to terminate and delete the cluster when workflow tasks are complete to save cluster resources.

Two new workflow tasks enable you to create and delete a Hadoop cluster as part of a cluster workflow:
Create Cluster Task

The Create Cluster task enables you to create, configure and start a Hadoop cluster on the following cloud platforms:

- Amazon Web Services (AWS). You can create an Amazon EMR cluster.

- Microsoft Azure. You can create an HDInsight cluster.

Delete Cluster Task

The optional Delete Cluster task enables you to delete a cluster after Mapping tasks and any other tasks in the workflow are complete. You might want to do this to save costs.

Previously, you could use Command tasks in a workflow to create clusters on a cloud platform. For more information about cluster workflows and workflow tasks, see the *Informatica 10.2.1 Developer Workflow Guide*.

Note: In 10.2.1, the Command task method of creating and deleting clusters now supports Cloudera Altus clusters on AWS. For more information, see the article "How to Create Cloudera Altus Clusters with a Cluster Workflow on Big Data Management" on the Informatica Network.

Mapping Task

Mapping task advanced properties include a new ClusterIdentifier property. The ClusterIdentifier identifies the cluster to use to run the Mapping task.

For more information about cluster workflows, see the *Informatica 10.2.1 Developer Workflow Guide*.

Cloud Provisioning Configuration

A cloud provisioning configuration is an object that contains information about connecting to a Hadoop cluster.

The cloud provisioning configuration includes information about how to integrate the domain with Hadoop account authentication and storage resources. A cluster workflow uses the information in the cloud provisioning configuration to connect to and create a cluster on a cloud platform such as Amazon Web Services or Microsoft Azure.

For more information about cloud provisioning, see the "Cloud Provisioning Configuration" chapter in the *Informatica Big Data Management 10.2.1 Administrator Guide*.

High Availability

Effective in version 10.2.1, you can enable high availability for the following services and security systems in the Hadoop environment on Cloudera CDH, Hortonworks HDP, and MapR Hadoop distributions:

- Apache Ranger
- Apache Ranger KMS
- Apache Sentry
- Cloudera Navigator Encrypt
- HBase
- Hive Metastore
- HiveServer2
- Name node
- Resource Manager

Hive Functionality in the Hadoop Environment

This section describes new features for Hive functionality in the Hadoop environment in version 10.2.1.

Hive Table Truncation

Effective in version 10.2.1, you can truncate external partitioned Hive tables on all run-time engines.

You can truncate tables in the following Hive storage formats:

- Avro
- ORC
- Parquet

- RCFile
- Sequence
- Text

You can truncate tables in the following Hive external table formats:

- Hive on HDFS
- Hive on Amazon S3
- Hive on Azure Blob
- Hive on WASB
- Hive on ADLS

For more information on truncating Hive targets, see the "Mapping Targets in the Hadoop Environment" chapter in the *Informatica Big Data Management 10.2.1 User Guide*.

Pre- and Post-Mapping SQL Commands

Effective in version 10.2.1, you can configure PreSQL and PostSQL commands against Hive sources and targets in mappings that run on the Spark engine.

For more information, see the *Informatica Big Data Management 10.2.1 User Guide*.

Importing from PowerCenter

This section describes new import from PowerCenter features in version 10.2.1.

Import Session Properties from PowerCenter

Effective in version 10.2.1, you can import session properties, such as SQL-based overrides in relational sources and targets and overrides for the Lookup transformation from the PowerCenter repository to the Model repository.

For more information about the import from PowerCenter functionality, see the "Import from PowerCenter" chapter in the *Informatica 10.2.1 Developer Mapping Guide*.

SQL Parameters

Effective in version 10.2.1, you can specify an SQL parameter type to import all SQL-based overrides into the Model repository. The remaining session override properties map to String or a corresponding parameter type.

For more information, see the "Import from PowerCenter" chapter in the *Informatica 10.2.1 Developer Mapping Guide*.

Import a Command Task from PowerCenter

Effective in version 10.2.1, you can import a Command task from PowerCenter into the Model repository.

For more information, see the "Workflows" chapter in the *Informatica 10.2.1 Developer Workflow Guide*.

Intelligent Structure Model

Effective in version 10.2.1, you can use the intelligent structure model in Big Data Management.

Spark Engine Support for Data Objects with Intelligent Structure Model

You can incorporate an intelligent structure model in an Amazon S3, Microsoft Azure Blob, or complex file data object. When you add the data object to a mapping that runs on the Spark engine, 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.

Intelligent structure model in the complex file, Amazon S3, and Microsoft Azure Blob data objects 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 Big Data Management 10.2.1 User Guide*.

Mass Ingestion

Effective in version 10.2.1, you can perform mass ingestion jobs to ingest or replicate large amounts of data for use or storage in a database or a repository. To perform mass ingestion jobs, you use the Mass Ingestion tool to create a mass ingestion specification. You configure the mass ingestion specification to ingest data from a relational database to a Hive or HDFS target. You can also specify parameters to cleanse the data that you ingest.

A mass ingestion specification replaces the need to manually create and run mappings. You can create one mass ingestion specification that ingests all of the data at once.

For more information on mass ingestion, see the *Informatica Big Data Management 10.2.1 Mass Ingestion Guide*.

Monitoring

This section describes the new features related to monitoring in Big Data Management in version 10.2.1.

Hadoop Cluster Monitoring

Effective in version 10.2.1, you can configure the amount of information that appears in the application logs that you monitor for a Hadoop cluster.

The amount of information in the application logs depends on the tracing level that you configure for a mapping in the Developer tool. The following table describes the amount of information that appears in the application logs for each tracing level:

Tracing Level	Messages
None	The log displays FATAL messages. FATAL messages include non-recoverable system failures that cause the service to shut down or become unavailable.
Terse	The log displays FATAL and ERROR code messages. ERROR messages include connection failures, failures to save or retrieve metadata, service errors.
Normal	The log displays FATAL, ERROR, and WARNING messages. WARNING errors include recoverable system failures or warnings.

Tracing Level	Messages
Verbose initialization	The log displays FATAL, ERROR, WARNING, and INFO messages. INFO messages include system and service change messages.
Verbose data	The log displays FATAL, ERROR, WARNING, INFO, and DEBUG messages. DEBUG messages are user request logs.

For more information, see the "Monitoring Mappings in the Hadoop Environment" chapter in the *Informatica Big Data Management 10.2.1 User Guide*.

Spark Monitoring

Effective in version 10.2.1, the Spark executor listens on a port for Spark events as part of Spark monitoring support and it is not required to configure the SparkMonitoringPort.

The Data Integration Service has a range of available ports, and the Spark executor selects a port from the available range. During failure, the port connection remains available and you do not need to restart the Data Integration Service before running the mapping.

The custom property for the monitoring port is retained. If you configure the property, the Data Integration Service uses the specified port to listen to Spark events.

Previously, the Data Integration Service custom property, the Spark monitoring port could configure the Spark listening port. If you did not configure the property, Spark Monitoring was disabled by default.

Tez Monitoring

Effective in 10.2.1, you can view Tez engine monitoring support related properties. You can use the Hive engine to run the mapping on MapReduce or Tez. The Tez engine can process jobs on Hortonworks HDP, Azure HDInsight, and Amazon Elastic MapReduce. To run a Spark mapping on Tez, you can use any of the supported clusters for Tez.

In the Administrator tool, you can also review the Hive query properties for Tez when you monitor the Hive engine. In the Hive session log and in Tez, you can view information related to Tez statistics, such as DAG tracking URL, total vertex count, and DAG progress.

You can monitor any Hive query on the Tez engine. When you enable logging for verbose data or verbose initialization, you can view the Tez engine information in the Administrator tool or in the session log. You can also monitor the status of the mapping on the Tez engine on the Monitoring tab in the Administrator tool.

For more information about Tez monitoring, see the *Informatica Big Data Management 10.2.1 User Guide* and the *Informatica Big Data Management 10.2.1 Hadoop Integration Guide*.

Processing Hierarchical Data on the Spark Engine

Effective in version 10.2.1, the Spark engine includes the following additional functionality to process hierarchical data:

Map data type

You can use map data type to generate and process map data in complex files.

Complex files on Amazon S3

You can use complex data types to read and write hierarchical data in Avro and Parquet files on Amazon S3. You project columns as complex data type in the data object read and write operations.

For more information, see the "Processing Hierarchical Data on the Spark Engine" chapter in the *Informatica Big Data Management 10.2.1 User Guide*.

Rule Specification Support on the Spark Engine

Effective in version 10.2.1, you can run a mapping that contains a rule specification on the Spark engine in addition to the Blaze and Hive engines.

You can also run a mapping that contains a mapplet that you generate from a rule specification on the Spark engine in addition to the Blaze and Hive engines.

For more information about rule specifications, see the *Informatica 10.2.1 Rule Specification Guide*.

Security

This section describes the new features related to security in Big Data Management in version 10.2.1.

Cloudera Navigator Encrypt

Effective in version 10.2.1, you can use Cloudera Navigator Encrypt to secure the data and implement transparent encryption of data at rest.

EMR File System Authorization

Effective in version 10.2.1, you can use EMR File System (EMRFS) authorization to access data in Amazon S3 on Spark engine.

IAM Roles

Effective in version 10.2.1, you can use IAM roles for EMR File System to read and write data from the cluster to Amazon S3 in Amazon EMR cluster version 5.10.

Kerberos Authentication

Effective in version 10.2.1, you can enable Kerberos authentication for the following clusters:

- Amazon EMR
- Azure HDInsight with WASB as storage

LDAP Authentication

Effective in version 10.2.1, you can configure Lightweight Directory Access Protocol (LDAP) authentication for Amazon EMR cluster version 5.10.

Sqoop

Effective in version 10.2.1, you can use the following new Sqoop features:

Support for MapR Connector for Teradata

You can use MapR Connector for Teradata to read data from or write data to Teradata on the Spark engine. MapR Connector for Teradata is a Teradata Connector for Hadoop (TDCH) specialized connector for Sqoop. When you run Sqoop mappings on the Spark engine, the Data Integration Service invokes the connector by default.

For more information, see the *Informatica Big Data Management 10.2.1 User Guide*.

Spark engine optimization for Sqoop pass-through mappings

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

- You read data from a Sqoop source and write data to a Hive target that uses the Text format.
- You read data from a Sqoop source and write data to an HDFS target that uses the Flat, Avro, or Parquet format.

For more information, see the *Informatica Big Data Management 10.2.1 User Guide*.

Spark engine support for high availability and security features

Sqoop honors all the high availability and security features such as Kerberos keytab login and KMS encryption that the Spark engine supports.

For more information, see the "Data Integration Service" chapter in the *Informatica 10.2.1 Application Services Guide* and "infacmd dis Command Reference" chapter in the *Informatica 10.2.1 Command Reference Guide*.

Spark engine support for Teradata data objects

If you use a Teradata data object and you run a mapping on the Spark engine and on a Hortonworks or Cloudera cluster, the Data Integration Service runs the mapping through Sqoop.

If you use a Hortonworks cluster, the Data Integration Service invokes Hortonworks Connector for Teradata at run time. If you use a Cloudera cluster, the Data Integration Service invokes Cloudera Connector Powered by Teradata at run time.

For more information, see the *Informatica PowerExchange for Teradata Parallel Transporter API 10.2.1 User Guide*.

Transformation Support in the Hadoop Environment

This section describes new transformation features in the Hadoop environment in version 10.2.1.

Transformation Support on the Spark Engine

This section describes new transformation features on the Spark engine in version 10.2.1.

Transformation Support

Effective in version 10.2.1, the following transformations are supported on the Spark engine:

- Case Converter
- Classifier
- Comparison
- Key Generator
- Labeler
- Merge
- Parser
- Python
- Standardizer
- Weighted Average

Effective in version 10.2.1, the following transformations are supported with restrictions on the Spark engine:

- Address Validator
- Consolidation
- Decision
- Match
- Sequence Generator

Effective in version 10.2.1, the following transformation has additional support on the Spark engine:

- Java. Supports complex data types such as array, map, and struct to process hierarchical data.

For more information on transformation support, see the "Mapping Transformations in the Hadoop Environment" chapter in the *Informatica Big Data Management 10.2.1 User Guide*.

For more information about transformation operations, see the *Informatica 10.2.1 Developer Transformation Guide*.

Python Transformation

Effective in version 10.2.1, you can create a Python transformation in the Developer tool. Use the Python transformation to execute Python code in a mapping that runs on the Spark engine.

You can use a Python transformation to implement a machine model on the data that you pass through the transformation. For example, use the Python transformation to write Python code that loads a pre-trained model. You can use the pre-trained model to classify input data or create predictions.

Note: The Python transformation is available for technical preview. Technical preview functionality is supported but is not production-ready. Informatica recommends that you use in non-production environments only.

For more information, see the "Python Transformation" chapter in the *Informatica 10.2.1 Developer Transformation Guide*.

Update Strategy Transformation

Effective in version 10.2.1, you can use Hive MERGE statements for mappings that run on the Spark engine to perform update strategy tasks. Using MERGE in queries is usually more efficient and helps increase performance.

Hive MERGE statements are supported for the following Hadoop distributions:

- Amazon EMR 5.10
- Azure HDInsight 3.6
- Hortonworks HDP 2.6

To use Hive MERGE, select the option in the advanced properties of the Update Strategy transformation.

Previously, the Data Integration Service used INSERT, UPDATE and DELETE statements to perform this task using any run-time engine. The Update Strategy transformation still uses these statements in the following scenarios:

- You do not select the Hive MERGE option.
- Mappings run on the Hive or Blaze engine.
- If the Hadoop distribution does not support Hive MERGE.

For more information about using a MERGE statement in Update Strategy transformations, see the chapter on Update Strategy transformation in the *Informatica Big Data Management 10.2.1 User Guide*.

Transformation Support on the Blaze Engine

This section describes new transformation features on the Blaze engine in version 10.2.1.

Aggregator Transformation

Effective in version 10.2.1, the data cache for the Aggregator transformation uses variable length to store binary and string data types on the Blaze engine. Variable length reduces the amount of data that the data cache stores when the Aggregator transformation runs.

When data that passes through the Aggregator transformation is stored in the data cache using variable length, the Aggregator transformation is optimized to use sorted input and a Sorter transformation is inserted before the Aggregator transformation in the run-time mapping.

For more information, see the "Mapping Transformations in the Hadoop Environment" chapter in the *Informatica Big Data Management 10.2.1 User Guide*.

Match Transformation

Effective in version 10.2.1, you can run a mapping that contains a Match transformation that you configure for identity analysis on the Blaze engine.

Configure the Match transformation to write the identity index data to cache files. The mapping fails validation if you configure the Match transformation to write the index data to database tables.

For more information on transformation support, see the "Mapping Transformations in the Hadoop Environment" chapter in the *Informatica Big Data Management 10.2.1 User Guide*.

Rank Transformation

Effective in version 10.2.1, the data cache for the Rank transformation uses variable length to store binary and string data types on the Blaze engine. Variable length reduces the amount of data that the data cache stores when the Rank transformation runs.

When data that passes through the Rank transformation is stored in the data cache using variable length, the Rank transformation is optimized to use sorted input and a Sorter transformation is inserted before the Rank transformation in the run-time mapping.

For more information, see the "Mapping Transformations in the Hadoop Environment" chapter in the *Informatica Big Data Management 10.2.1 User Guide*.

For more information about transformation operations, see the *Informatica 10.2.1 Developer Transformation Guide*.

Big Data Streaming

This section describes new Big Data Streaming features in version 10.2.1.

Sources and Targets

Effective in version 10.2.1, you can read from or write to the following sources and targets in streaming mappings:

- Azure Event Hubs. Create an Azure EventHub data object to read from or write to Event Hub events. You can use an Azure EventHub connection to access Microsoft Azure Event Hubs as source or target. You can create and manage an Azure Eventhub connection in the Developer tool or through infacmd.

- Microsoft Azure Data Lake Store. Create an Azure Data Lake store data object to write to Azure Data Lake Store. You can use an Azure Data Lake Store connection to access Microsoft Azure Data Lake Store tables as targets. You can create and manage a Microsoft Azure Data Lake Store connection in the Developer tool.
- JDBC-compliant database. Create a relational data object with a JDBC connection.

For more information, see the "Sources in a Streaming Mapping" and "Targets in a Streaming Mapping" chapters in the *Informatica Big Data Streaming 10.2.1 User Guide*.

Stateful Computing in Streaming Mappings

Effective in 10.2.1, you can use window functions in an Expression transformation to perform stateful calculations in streaming mappings.

For more information, see the "Streaming Mappings" chapter in the *Informatica Big Data Streaming 10.2.1 User Guide*.

Transformation Support

Effective in version 10.2.1, you can use the following transformations in streaming mappings:

- Data Masking
- Normalizer
- Python

You can perform an uncached lookup on HBase data in streaming mappings with a Lookup transformation.

For more information, see the "Streaming Mappings" chapter in the *Informatica Big Data Streaming 10.2.1 User Guide*.

Truncate Partitioned Hive Target Tables

Effective in version 10.2.1, you can truncate an external or managed Hive table with or without partitions.

For more information about truncating Hive targets, see the "Targets in a Streaming Mapping" chapter in the *Informatica Big Data Streaming 10.2.1 User Guide*.

Command Line Programs

This section describes new commands in version 10.2.1.

infacmd autotune Commands

autotune is a new infacmd plugin that tunes services and connections in the Informatica domain.

The following table describes new infacmd autotune commands:

Command	Description
Autotune	Configures services and connections in the Informatica domain with recommended settings based on the size description.

For more information, see the "infacmd autotune Command Reference" chapter in the *Informatica 10.2.1 Command Reference*.

infacmd ccps Commands

ccps is a new infacmd plugin that performs operations on cloud platform clusters.

The following table describes new infacmd ccps commands:

Command	Description
deleteClusters	Deletes clusters on the cloud platform that a cluster workflow created.
listClusters	Lists clusters on the cloud platform that a cluster workflow created.
updateADLSCertifcate	Updates the Azure Data Lake Service Principal certificate.

For more information, see the "infacmd ccps Command Reference" chapter in the *Informatica 10.2.1 Command Reference*.

infacmd cluster Commands

The following table describes new infacmd cluster commands:

Command	Description
updateConfiguration	Updates the Hadoop distribution version of a cluster configuration. Use the -dv option to change the distribution version of the Hadoop distribution of a cluster configuration.

The following table describes changes to infacmd cluster commands:

Command	Change Description
listConfigurationProperties	Effective in 10.2.1, you can specify the general configuration set when you use the -cs option to return the property values in the general configuration set. Previously, the -cs option accepted only .xml file names.
createConfiguration	Effective in 10.2.1, you can optionally use the -dv option to specify a Hadoop distribution version when you create a cluster configuration. If you do not specify a version, the command creates a cluster configuration with the default version for the specified Hadoop distribution. Previously, the createConfiguration command did not contain the option to specify the Hadoop version.

For more information, see the "infacmd cluster Command Reference" chapter in the *Informatica 10.2.1 Command Reference*.

infacmd cms Commands

The following table describes new Content Management Service options for infacmd cms updateServiceOptions:

Command	Description
DataServiceOptions.RefDataLocationSchema	Identifies the schema that specifies the reference data tables in the reference data database.

For more information, see the "infacmd cms Command Reference" chapter in the *Informatica 10.2.1 Command Reference*.

infacmd dis Commands

The following table describes new infacmd dis commands:

Command	Description
listMappingEngines	Lists the execution engines of the deployed mappings on a Data Integration Service.

For more information, see the "infacmd dis Command Reference" chapter in the *Informatica 10.2.1 Command Reference*.

infacmd ihs Commands

The following table describes new infacmd ihs commands:

Command	Description
ListServiceProcessOptions	Lists process options for the Informatica Cluster Service.
UpdateServiceProcessOptions	Updates service options for the Informatica Cluster Service.

For more information, see the "infacmd ihs Command Reference" chapter in the *Informatica 10.2.1 Command Reference*.

infacmd isp Commands

The following table describes new infacmd isp commands:

Command	Description
PingDomain	Pings a domain, service, domain gateway host, or node.
GetPasswordComplexityConfig	Returns the password complexity configuration for the domain users.
ListWeakPasswordUsers	Lists the users with passwords that do not meet the password policy.

For more information, see the "infacmd isp Command Reference" chapter in the *Informatica 10.2.1 Command Reference*.

infacmd ldm Commands

The following table describes new infacmd ldm commands:

Command	Description
ListServiceProcessOptions	Lists options for the Catalog Administrator process.
UpdateServiceProcessOptions	Updates process options for the Catalog Service.

For more information, see the "infacmd ldm Command Reference" chapter in the *Informatica 10.2.1 Command Reference*.

infacmd mi Commands

mi is a new infacmd plugin that performs mass ingestion operations.

The following table describes new infacmd mi commands:

Command	Description
abortRun	Aborts the ingestion mapping jobs in a run instance of a mass ingestion specification.
createService	Creates a Mass Ingestion Service. Disabled by default. To enable the Mass Ingestion Service, use infacmd isp enableService.
deploySpec	Deploys a mass ingestion specification.
exportSpec	Exports the mass ingestion specification to an application archive file.
extendedRunStats	Gets the extended statistics for a mapping in the deployed mass ingestion specification.
getSpecRunStats	Gets the detailed run statistics for a deployed mass ingestion specification.
listSpecRuns	Lists the run instances of a deployed mass ingestion specification.
listSpecs	Lists the mass ingestion specifications.

Command	Description
restartMapping	Restarts the ingestion mapping jobs in a mass ingestion specification.
runSpec	Runs a mass ingestion specification that is deployed to a Data Integration Service.

For more information, see the "infacmd mi Command Reference" chapter in the *Informatica 10.2.1 Command Reference*.

infacmd mrs Commands

The following table describes new infacmd mrs commands:

Command	Description
listMappingEngines	Lists the execution engines of the mappings that are stored in a Model repository.
listPermissionOnProject	Lists all the permissions on multiple projects for groups and users.
updateStatistics	Updates the statistics for the monitoring Model repository on Microsoft SQL Server.

For more information, see the "infacmd mrs Command Reference" chapter in the *Informatica 10.2.1 Command Reference*.

infacmd wfs Commands

The following table describes new infacmd wfs commands:

Command	Description
pruneOldInstances	Deletes workflow process data from the workflow database.

To delete the process data, you must have the Manage Service privilege on the domain.

For more information, see the "infacmd wfs Command Reference" chapter in the *Informatica 10.2.1 Command Reference*.

infasetup Commands

The following table describes new infasetup commands:

Command	Description
UpdatePasswordComplexityConfig	Enables or disables the password complexity configuration for the domain.

For more information, see the "infasetup Command Reference" chapter in the *Informatica 10.2.1 Command Reference*.

Enterprise Data Catalog

This section describes new Enterprise Data Catalog features in version 10.2.1.

Adding a Business Title to an Asset

Effective in version 10.2.1, you can add a business title to any asset in the catalog except for Business Glossary and Axon glossary assets. You can either associate a business term or provide a display name to add a business title to an asset.

For more information about adding a business title, see the *Informatica 10.2 .1 Enterprise Data Catalog User Guide*.

Cluster Validation Utility in Installer

Effective in version 10.2.1, when you install Enterprise Data Catalog, the installer provides an option to run the cluster-validation utility. The utility helps you validate the prerequisites to install Enterprise Data Catalog in an embedded cluster and existing cluster. The utility also validates the configuration settings for Informatica domain, cluster hosts, and the Hadoop cluster services.

For more information about the utility, see the *Informatica Enterprise Data Catalog 10.2 .1 Installation and Configuration Guide* and the following knowledge base articles:

- HOW TO: Validate Embedded Cluster Prerequisites with Validation Utility in Enterprise Information Catalog
- HOW TO: Validate Informatica Domain, Cluster Hosts, and Cluster Services Configuration

Data Domain Discovery Types

Effective in version 10.2.1, when you configure the data domain discovery profile settings, you can choose one of the following data domain discovery types:

- Run Discovery on Source Data. Scanner runs data domain discovery on source data.
- Run Discovery on Source Metadata. Scanner runs data domain discovery on source metadata.
- Run Discovery on both Source Metadata and Data. Scanner runs data domain discovery on source data and source metadata.
- Run Discovery on Source Data Where Metadata Matches. Scanner runs data domain discovery on the source metadata to identify the columns with inferred data domains. The scanner then runs discovery on the source data for the columns that have inferred data domains.

For more information about data domain discovery types, see the *Informatica 10.2 .1 Catalog Administrator Guide*.

Filter Settings

Effective in version 10.2.1, you can use the filter settings in the Application Configuration page to customize the search filters that you view in the **Filter By** panel of the search results page.

For more information about search filters, see the *Informatica Enterprise Data Catalog 10.2.1 User Guide*.

Missing Links Report

Effective in version 10.2.1, you can now generate a missing links report to identify the connection links that are missing after you assign schemas from a resource to connections.

For more information about the missing links report, see the *Informatica 10.2.1 Catalog Administrator Guide*.

New Resource Types

Effective in version 10.2.1, Informatica Enterprise Data Catalog extracts metadata from several new data sources.

You can create resources in Informatica Catalog Administrator to extract metadata from the following data sources:

Azure Data Lake Store

Online cloud file storage platform.

Database Scripts

Database scripts to extract lineage information. The Database Scripts resource 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.

Microsoft Azure Blob Storage

Cloud-based file storage web service.

QlikView

Business Intelligence tool that allows you to extract metadata from the QlikView source system.

SharePoint

Import metadata from files in SharePoint.

OneDrive

Import metadata from files in OneDrive.

For more information about the new resources, see the *Informatica 10.2 .1 Catalog Administrator Guide*.

REST APIs

Effective in version 10.2.1, you can use Informatica Enterprise Data Catalog REST APIs to load and monitor resources.

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

SAML Authentication for Enterprise Data Catalog Applications

Effective in version 10.2.1, you can enable Single Sign-on using SAML authentication for Enterprise Data Catalog applications. You can either use SAML authentication using OKTA with Active Directory or Active Directory Federation Services with Active Directory.

For more information, see the *Informatica Enterprise Data Catalog 10.2 .1 Installation and Configuration Guide*.

SAP Resource

Effective in version 10.2.1, you can choose the **Enable Streaming for Data Access** option for SAP R/3 resources to extract data by using the HTTP protocol.

For more information about the option, see the *Informatica 10.2 .1 Catalog Administrator Guide.*

Import from ServiceNow

Effective in version 10.2.1, Catalog Administrator now connects to ServiceNow to import connections and extract the configuration metadata into the catalog.

The Import from ServiceNow feature 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 about importing metadata from ServiceNow, see the *Informatica 10.2 .1 Catalog Administrator Guide.*

Similar Columns

Effective in version 10.2.1, you can view the Similar Columns section that displays all the columns that are similar to the column you are viewing. Enterprise Data Catalog discovers similar columns based on column names, column patterns, unique values, and value frequencies.

For more information about column similarity, see the *Informatica 10.2 .1 Enterprise Data Catalog User Guide.*

Specify Load Types for Catalog Service

Effective in version 10.2.1, when you create a Catalog Service, you can choose the option to specify the data size that you want to deploy.

Previously, you had to create the Catalog Service and use the custom properties for the Catalog Service to specify the data size.

For more information, see the *Informatica Enterprise Data Catalog 10.2 .1 Installation and Configuration Guide.*

Supported Resource Types for Data Discovery

Effective in version 10.2.1, you can enable data discovery for the following resources to extract profiling metadata:

- Unstructured file types:
  - Apple Files. Supported extension types include .key, .pages, .numbers, .ibooks, and .ipa.
  - Open Office Files. Supported extension types
    include .odt, .ott, .odm, .ods, .ots, .odp, .odg, .otp, .odg, .otg, and .odf.
- Structured file types:
  - Avro. Supported extension type is .avro.

    This file type is available for HDFS resource and File System resource. For the File System resource, you can choose only the Local File protocol.

- Parquet. Supported extension type is .parquet.

    This file type is available for HDFS resource and File System resource. For the File System resource, you can choose only the Local File protocol.

- Other resources:

    - Azure Data Lake Store

    - File System. Supported protocols include Local File, SFTP, and SMB/CIFS protocol.

    - HDFS. Supported distribution includes MapR FS.

    - Microsoft Azure Blob Storage

    - OneDrive

    - SharePoint

For more information about new resources, see the *Informatica 10.2 .1 Catalog Administrator Guide*.

Enterprise Data Lake

This section describes new Enterprise Data Lake features in version 10.2.1.

Column Data

Effective in version 10.2.1, you can use the following features when you work with columns in worksheets:

- You can categorize or group related values in a column into categories to make analysis easier.

- You can view the source of the data for a selected column in a worksheet. You might want to view the source of the data in a column to help you troubleshoot an issue.

- You can revert types or data domains inferred during sampling on columns to the source type. You might want to revert an inferred type or data domain to the source type if you want to use the column data in a formula.

For more information, see the "Prepare Data" chapter in the *Informatica 10.2.1 Enterprise Data Lake User Guide*.

Manage Data Lake Resources

Effective in version 10.2.1, you can use the Enterprise Data Lake application to add and delete Enterprise Data Catalog resources. Catalog resources represent the external data sources and metadata repositories from which scanners extract metadata that can be used in the data lake.

For more information, see the "Managing the Data Lake" chapter in the *Informatica 10.2.1 Enterprise Data Lake Administrator Guide*.

Data Preparation Operations

Effective in version 10.2.1, you can perform the following operations during data preparation:

Pivot Data

    You can use the pivot operation to reshape the data in selected columns in a worksheet into a summarized format. The pivot operation enables you to group and aggregate data for analysis, such as

summarizing the average price of single family homes sold in each city for the first six months of the year.

Unpivot Data

You can use the unpivot operation to transform columns in a worksheet into rows containing the column data in key value format. The unpivot operation is useful when you want to aggregate data in a worksheet into rows based on keys and corresponding values.

Apply One Hot Encoding

You can use the one hot encoding operation to determine the existence of a string value in a selected column within each row in a worksheet. You might use the one hot encoding operation to convert categorical values in a worksheet to numeric values required by machine learning algorithms.

For more information, see the "Prepare Data" chapter in the *Informatica 10.2.1 Enterprise Data Lake User Guide*.

Prepare JSON Files

Effective in version 10.2.1, you can sample the hierarchal data in JavaScript Object Notation Lines (JSONL) files you add to your project as the first step in data preparation. Enterprise Data Lake converts the JSON file structure into a flat structure, and presents the data in a worksheet that you use to sample the data.

For more information, see the "Prepare Data" chapter in the *Informatica 10.2.1 Enterprise Data Lake User Guide*.

Recipe Steps

Effective in version 10.2.1, you can use the following features when you work with recipes in worksheets:

- You can reuse recipe steps created in a worksheet, including steps that contain complex formulas or rule definitions. You can reuse recipe steps within the same worksheet or in a different worksheet, including a worksheet in another project. You can copy and reuse selected steps from a recipe, or you can reuse the entire recipe.
- You can insert a step at any position in a recipe.
- You can add a filter or modify a filter applied to a recipe step.

For more information, see the "Prepare Data" chapter in the *Informatica 10.2.1 Enterprise Data Lake User Guide*.

Schedule Export, Import, and Publish Activities

Effective in version 10.2.1, you can schedule the exporting, importing, and publishing of data assets. Scheduling an activity enables you to import, export or publish updated data assets on a recurring basis.

When you schedule an activity, you can create a new schedule, or you can select an existing schedule. You can use schedules created by other users, and other users can use schedules that you create.

For more information, see the "Scheduling Export, Import, and Publish Activities" chapter in the *Informatica 10.2.1 Enterprise Data Lake User Guide*.

Security Assertion Markup Language Authentication

Effective in version 10.2.1, the Enterprise Data Lake application supports Security Assertion Markup Language (SAML) authentication.

For more information on configuring SAML authentication, see the *Informatica 10.2.1 Security Guide*.

View Project Flows and Project History

Effective in version 10.2.1, you can view project flow diagrams and review the activities performed within a project.

You can view a flow diagram that shows you how worksheets in a project are related and how they are derived. The diagram is especially useful when you work on a complex project that contains numerous worksheets and includes numerous assets.

You can also review the complete history of the activities performed within a project, including activities performed on worksheets within the project. Viewing the project history might help you determine the root cause of issues within the project.

For more information, see the "Create and Manage Projects" chapter in the *Informatica 10.2.1 Enterprise Data Lake User Guide*.

Informatica Developer

This section describes new Developer tool features in version 10.2.1.

Default Layout

Effective in version 10.2.1, the following additional views appear by default in the Developer tool workbench:

- Connection Explorer view
- Progress view

The following image shows the default Developer tool workbench in version 10.2.1:
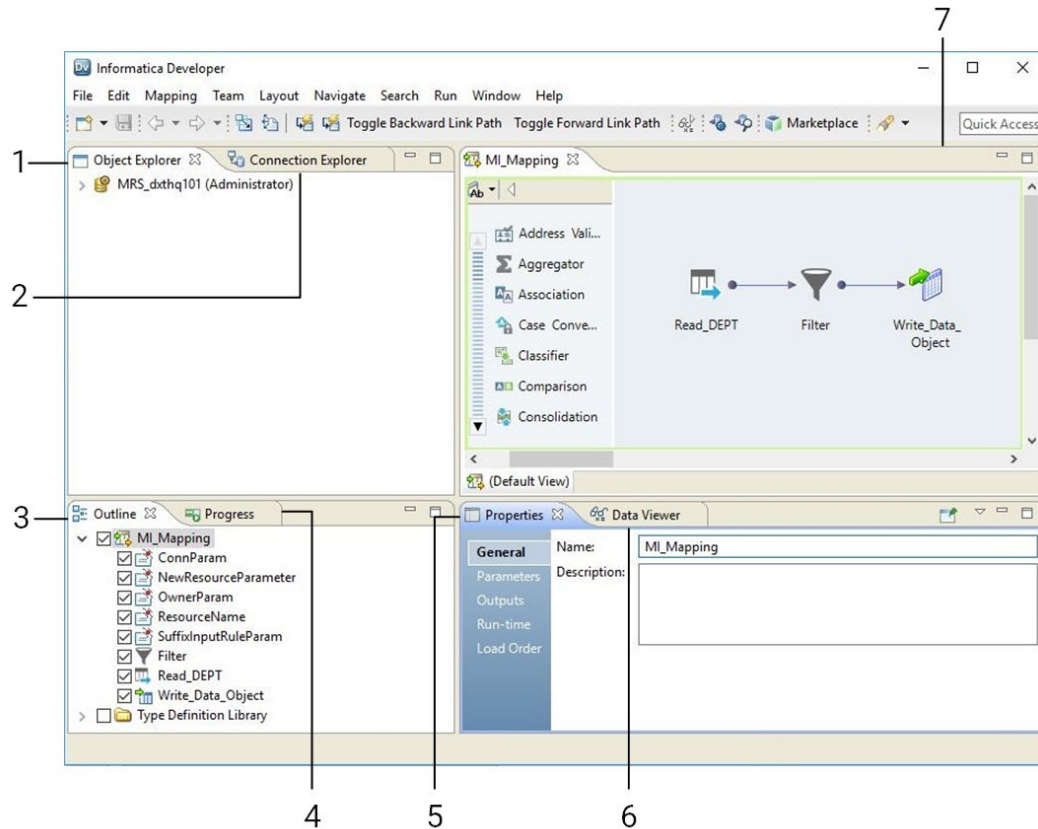
1. Object Explorer view
2. Connection Explorer view
3. Outline view
4. Progress view
5. Properties view
6. Data Viewer view
7. Editor

For more information, see the "Informatica Developer" chapter in the *Informatica 10.2.1 Developer Tool Guide*.

Editor Search

Effective in version 10.2.1, you can search for a complex data type definition in mappings and mapplets in the Editor view. You can also show link paths using a complex data type definition.

For more information, see the "Searches in Informatica Developer" chapter in the *Informatica 10.2.1 Developer Tool Guide*.

Import Session Properties from PowerCenter

Effective in version 10.2.1, you can import session properties, such as SQL-based overrides in relational sources and targets and overrides for the Lookup transformation from the PowerCenter repository to the Model repository.

For more information about the import from PowerCenter functionality, see the "Import from PowerCenter" chapter in the *Informatica 10.2.1 Developer Mapping Guide*.

Views

Effective in version 10.2.1, you can expand complex data types to view the complex data type definition in the following views:

- Editor view
- Outline view
- Properties view

For more information, see the "Informatica Developer" chapter in the *Informatica 10.2.1 Developer Tool Guide*.

Informatica Mappings

This section describes new Informatica mapping features in version 10.2.1.

Dynamic Mappings

This section describes new dynamic mapping features in version 10.2.1.

Input Rules

Effective in version 10.2.1, you can perform the following tasks when you create an input rule:

- Create an input rule by complex data type definition.
- Restore source port names when you rename generated ports.
- Select ports by source name when you create an input rule by column name or a pattern.
- View source names and complex data type definitions in the port preview.

For more information, see the "Dynamic Mappings" chapter in the *Informatica 10.2.1 Developer Mapping Guide*.

Port Selectors

Effective in version 10.2.1, you can configure a port selector to select ports by complex data type definition.

For more information, see the "Dynamic Mappings" chapter in the *Informatica 10.2.1 Developer Mapping Guide*.

Validate Dynamic Sources and Targets

Effective in version 10.2.1, you can validate dynamic sources and targets. To validate dynamic sources and targets, resolve the mapping parameters to view a run-time instance of the mapping. Validate the run-time instance of the mapping.

For more information, see the "Dynamic Mappings" chapter in the *Informatica 10.2.1 Developer Mapping Guide*.

Mapping Parameters

This section describes new mapping parameters features in version 10.2.1

Assign Parameters

Effective in version 10.2.1, you can assign parameters to the following mapping objects and object fields:

Object	Field
Customized data object read operation	Custom query Filter condition Join condition PreSQL PostSQL
Customized data object write operation	PreSQL PostSQL Update override
Flat file data object	Compression codec Compression format
Lookup transformation	Custom query. Relational only.
Read transformation	Custom query. Relational only. Filter condition. Relational only. Join condition. Relational only. PreSQL. Relational only. PostSQL. Relational only.
Write transformation	PreSQL. Relational only. PostSQL. Relational only. Update override. Relational only.

For more information, see the "Mapping Parameters" chapter in the *Informatica 10.2.1 Developer Mapping Guide*.

Resolve Mapping Parameters

Effective in version 10.2.1, you can resolve mapping parameters in the Developer tool. When you resolve mapping parameters, the Developer tool generates a run-time instance of the mapping that shows how the Data Integration Service resolves the parameters at run time. You can run the instance of the mapping where the parameters are resolved to run the mapping with the selected parameters.

The following table describes the options that you can use to resolve mapping parameters:

Mapping Parameters	Description
Apply the default values in the mapping	Resolves the mapping parameters based on the default values configured for the parameters in the mapping. If parameters are not configured for the mapping, no parameters are resolved in the mapping.
Apply a parameter set	Resolves the mapping parameters based on the parameter values defined in the specified parameter set.
Apply a parameter file	Resolves the mapping parameters based on the parameter values defined in the specified parameter file.

To quickly resolve mapping parameters based on a parameter set. Drag the parameter set from the Object Explorer view to the mapping editor to view the resolved parameters in the run-time instance of the mapping.

For more information, see the "Mapping Parameters" chapter in the *Informatica 10.2.1 Developer Mapping Guide*.

Validate Mapping Parameters

Effective in version 10.2.1, you can validate mapping parameters in the Developer tool. To validate mapping parameters, first resolve the mapping parameters. When you resolve mapping parameters, the Developer tool generates a run-time instance of the mapping that shows the resolved parameters. Validate the run-time instance of the mapping to validate the mapping parameters.

For more information, see the "Mapping Parameters" chapter in the *Informatica 10.2.1 Developer Mapping Guide*.

Running Mappings

This section describes new run mapping features in version 10.2.1.

Run a Mapping from the Object Explorer View

Effective in version 10.2.1, you can run a mapping from the Object Explorer view. You do not have to open the mapping in the mapping editor. Right-click the mapping in the Object Explorer view and click Run.

For more information, see the *Informatica 10.2.1 Developer Tool Guide*.

Run a Mapping Using Advanced Options

Effective in version 10.2.1, you can run a mapping in the Developer tool using advanced options. In the advanced options, you can specify a mapping configuration and mapping parameters. Specify the mapping configuration and mapping parameters each time that you run the mapping.

The following table describes the options that you can use to specify a mapping configuration:

Option	Description
Select a mapping configuration	Select a mapping configuration from the drop-down menu. To create a new mapping configuration, select New Configuration.
Specify a custom mapping configuration	Create a custom mapping configuration that persists for the current mapping run.

The following table describes the options that you can use to specify mapping parameters:

Mapping Parameters	Description
Apply the default values in the mapping	Resolves the mapping parameters based on the default values configured for the parameters in the mapping. If parameters are not configured for the mapping, no parameters are resolved in the mapping.
Apply a parameter set	Resolves the mapping parameters based on the parameter values defined in the specified parameter set.
Apply a parameter file	Resolves the mapping parameters based on the parameter values defined in the specified parameter file.

For more information, see the *Informatica 10.2.1 Developer Mapping Guide*.

Truncate Partitioned Hive Target Tables

Effective in version 10.2.1, you can truncate an external or managed Hive table with or without partitions.

Previously, you could design a mapping to truncate a Hive target table, but not an external, partitioned Hive target table.

For more information on truncating Hive targets, see the "Mapping Targets in the Hadoop Environment" chapter in the *Informatica Big Data Management 10.2.1 User Guide*.

Informatica Transformation Language

This section describes Informatica Transformation Language new features in 10.2.1.

Complex Functions for Map Data Type

Effective in version 10.2.1, the transformation language introduces complex functions for map data type. Use complex functions for map data type to generate or process map data on the Spark engine.

The transformation language includes the following complex functions for map data type:

- COLLECT_MAP
- MAP
- MAP_FROM_ARRAYS
- MAP_KEYS
- MAP_VALUES

Effective in version 10.2.1, you can use the SIZE function to determine the size of map data.

For more information about complex functions, see the "Functions" chapter in the *Informatica 10.2.1 Developer Transformation Language Reference*.

Complex Operator for Map Data Type

Effective in version 10.2.1, you can use a complex operator in mappings that run on the Spark engine to access elements in a map data type.

Map data type contains an unordered collection of key-value pair elements. Use the subscript operator [ ] to access the value corresponding to a given key in the map data type.

For more information about complex operators, see the "Operators" chapter in the *Informatica 10.2.1 Developer Transformation Language Reference.*

Informatica Transformations

This section describes new Informatica transformation features in version 10.2.1.

Address Validator Transformation

This section describes the new Address Validator transformation features.

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

Argentina

Effective in version 10.2.1, you can configure Informatica to return valid suggestions for an Argentina address that you enter on a single line.

Enter an Argentina address in the following format:

    [Street] [House Number] [Dependent Locality] [Post Code] [Locality]

To verify single-line addresses, enter the addresses in the `Complete Address` port.

Brazil

Effective in version 10.2.1, you can configure Informatica to return valid suggestions for a Brazil address that you enter on a single line.

Enter a Brazil address in the following format:

    [Street] [House Number] [Locality] [State Code] [Post Code]

To verify single-line addresses, enter the addresses in the `Complete Address` port.

Colombia

Effective in version 10.2.1, Informatica validates an address in Colombia to house number level.

Hong Kong

Effective in version 10.2.1, Informatica supports rooftop geocoding for Hong Kong addresses. Informatica can return rooftop geocoordinates for a Hong Kong address that you submit in the Chinese language or the English language.

Informatica can consider all three levels of building information when it generates the geocoordinates. It delivers rooftop geocoordinates to the lowest level available in the verified address.

To retrieve rooftop geocoordinates for Hong Kong addresses, install the `HKG5GCRT.MD` database.

India

Effective in version 10.2.1, Informatica validates an address in India to house number level.

Mexico

Effective in version 10.2.1, you can configure Informatica to return valid suggestions for a Mexico address that you enter on a single line.

Enter a Mexico address in the following format:

```
[Street] [House Number] [Sub-locality] [Post Code] [Locality] [Province]
```

To verify single-line addresses, enter the addresses in the `Complete Address` port.

South Africa

Effective in version 10.2.1, Informatica improves the parsing and verification of delivery service descriptors in South Africa addresses.

Informatica improves the parsing and verification of the delivery service descriptors in the following ways:

- Address Verification recognizes Private Bag, Cluster Box, Post Office Box, and Postnet Suite as different types of delivery service. Address Verification does not standardize one delivery service descriptor to another. For example, Address Verification does not standardize Postnet Suite to Post Office Box.

- Address Verification parses Postnet Box as a non-standard delivery service descriptor and corrects Postnet Box to the valid descriptor Postnet Suite.

- Address Verification does not standardize the sub-building descriptor Flat to Fl.

South Korea

Effective in version 10.2.1, Informatica introduces the following features and enhancements for South Korea:

- The South Korea address reference data includes building information. Informatica can read, verify, and correct building information in a South Korea address.

- Informatica returns all of the current addresses at a property that an older address represents. The older address might represent a single current address or it might represent multiple addresses, for example if multiple residences occupy the site of the property.

  To return the current addresses, first find the address ID for the older property. When you submit the address ID with the final character A in address code lookup mode, Informatica returns all current addresses that match the address ID.

  Note: The Address Validator transformation uses the `Max Result Count` property to determine the maximum number of addresses to return for the address ID that you enter. The `Count Overflow` property indicates whether the database contains additional addresses for the address ID.

Thailand

Effective in version 10.2.1, Informatica introduces the following features and enhancements for Thailand:

Improvements to Thailand Addresses

Informatica improves the parsing and validation of Thailand addresses in a Latin script.

Additionally, Informatica validates an address to house number level.

Native Support for Thailand Addresses

Informatica can read and write Thailand addresses in native Thai and Latin scripts. Informatica updates the reference data for Thailand and adds reference data in the native Thai script.

Informatica provides separate reference databases for Thailand addresses in each script. To verify addresses in the native Thai script, install the native Thai databases. To verify addresses in a Latin script, install the Latin databases.

Note: If you verify Thailand addresses, do not install both database types. Accept the default option for the `Preferred Script` property.

United Arab Emirates

Effective in version 10.2.1, Informatica verifies street names in United Arab Emirates addresses. To verify street names in United Arab Emirates, install the current reference address databases for the United Arab Emirates.

United Kingdom

Effective in version 10.2.1, Informatica can return a United Kingdom territory name.

Informatica returns the territory name in the Country_2 element. Informatica returns the country name in the Country_1 element. You can configure an output address with both elements, or you can omit the Country_1 element if you post mail within the United Kingdom. The territory name appears above the postcode in a United Kingdom address on an envelope or label.

To return the territory name, install the current United Kingdom reference data.

United States

Effective in version 10.2.1, Informatica can recognize up to three sub-building levels in a United States address.

In compliance with the United States Postal Service requirements, Informatica matches the information in a single sub-building element with the reference data. If the Sub-building_1 information does not match, Informatica compares the Sub-building_2 information. If the Sub-building_2 information does not match, Address Verification compares the Sub-building_3 information. Address Verification copies the unmatched sub-building information from the input address to the output address.

Austria, Germany, and Switzerland

Effective in version 10.2.1, Informatica supports the uppercase character ẞ in Austria, Germany, and Switzerland addresses.

Informatica supports the character ẞ in the following ways:

- If you set the `Casing` property to UPPER, Informatica returns the German character ß as ẞ. If you set the `Casing` property to LOWER, Informatica returns the German character ẞ as ß.

- Informatica treats ẞ and ß as equally valid characters in an address. In reference data matches, Informatica can identify a perfect match when the same values contain either ẞ or ß.

- Informatica treats ß and `ss` as equally valid characters in an address. In reference data matches, Informatica can identify a standardized match when the same values contain either ß or `ss`.

- If you set the `Preferred Script` property to ASCII_SIMPLIFIED, Informatica returns the character ß as `S`.

- If you set the `Preferred Script` property to ASCII_EXTENDED, Informatica returns the character ß as `SS`.

For comprehensive information about the features and operations of the address verification software engine version that Informatica embeds in version 10.2.1, see the *Informatica Address Verification 5.12.0 Developer Guide*.

Informatica Workflows

This section describes new Informatica workflow features in version 10.2.1.

Import a Command Task from PowerCenter

Effective in version 10.2.1, you can import a Command task from PowerCenter into the Model repository.

For more information, see the "Workflows" chapter in the *Informatica 10.2.1 Developer Workflow Guide*.

PowerExchange Adapters for Informatica

This section describes new Informatica adapter features in version 10.2.1.

PowerExchange for Amazon Redshift

Effective in version 10.2.1, PowerExchange for Amazon Redshift includes the following features:

- You can configure a cached lookup operation to cache the lookup table on the Spark engine and an uncached lookup operation in the native environment.
- For a server-side encryption, you can configure the customer master key ID generated by AWS Key Management Service in the connection in the native environment and Spark engine.

For more information, see the *Informatica PowerExchange for Amazon Redshift 10.2.1 User Guide*.

PowerExchange for Amazon S3

Effective in version 10.2.1, PowerExchange for Amazon S3 includes the following features:

- For a client-side encryption, you can configure the customer master key ID generated by AWS Key Management Service in the connection in the native environment. For a server-side encryption, you can configure the customer master key ID generated by AWS Key Management Service in the connection in the native environment and Spark engine.
- For a server-side encryption, you can configure the Amazon S3-managed encryption key or AWS KMS-managed customer master key to encrypt the data while uploading the files to the buckets.
- You can create an Amazon S3 file data object from the following data source formats in Amazon S3:
  - Intelligent Structure Model
    The intelligent structure model feature for PowerExchange for Amazon S3 is available for technical preview. Technical preview functionality is supported but is not production-ready. Informatica recommends that you use in non-production environments only.
  - JSON
  - ORC
- You can compress an ORC data in the Zlib compression format when you write data to Amazon S3 in the native environment and Spark engine.
- You can create an Amazon S3 target using the **Create Target** option in the target session properties.
- You can use complex data types on the Spark engine to read and write hierarchical data in the Avro and Parquet file formats.
- You can use Amazon S3 sources as dynamic sources in a mapping. Dynamic mapping support for PowerExchange for Amazon S3 sources 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 Amazon S3 10.2.1 User Guide*.

PowerExchange for Cassandra

Effective in version 10.2.1, the Informatica Cassandra ODBC driver supports asynchronous write.

To enable asynchronous write on a Linux operating system, you must add the **EnableAsynchronousWrites** key name in the `odbc.ini` file and set the value to 1.

To enable asynchronous write on a Windows operating system, you must add the **EnableAsynchronousWrites** property in the Windows registry for the Cassandra ODBC data source name and set the value as 1.

For more information, see the *Informatica PowerExchange for Cassandra 10.2.1 User Guide*.

PowerExchange for HBase

Effective in version 10.2.1, you can use an HBase data object read operation to look up data in an HBase resource. You can enable lookup caching and also parameterize the lookup condition.

For more information, see the *Informatica PowerExchange for HBase 10.2.1 User Guide*.

PowerExchange for HDFS

Effective in version 10.2.1, you can use the following new PowerExchange for HDFS features:

Intelligent structure model support for complex file data objects

You can incorporate an intelligent structure model in a complex file data object. When you add the data object to a mapping that runs on the Spark engine, you can process any input type that the model can parse.

The intelligent structure model feature for PowerExchange for HDFS is available for technical preview. Technical preview functionality is supported but is not production-ready. Informatica recommends that you use in non-production environments only.

For more information, see the *Informatica PowerExchange for HDFS 10.2.1 User Guide*.

Dynamic mapping support for complex file sources

You can use complex file sources as dynamic sources in a mapping.

Dynamic mapping support for complex file sources 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 about dynamic mappings, see the *Informatica Developer Mapping Guide*.

PowerExchange for Hive

Effective in version 10.2.1, PowerExchange for Hive supports mappings that run PreSQL and PostSQL queries against Hive sources and targets on the Spark engine.

For more information, see the *Informatica PowerExchange for Hive 10.2.1 User Guide*.

PowerExchange for Microsoft Azure Blob Storage

Effective in version 10.2.1, PowerExchange for Microsoft Azure Blob Storage includes the following functionality:

- You can run mappings on the Spark engine.

- You can read and write .csv, Avro, and Parquet files when you run a mapping on the Spark engine and in the native environment.
- You can read and write JSON and intelligent structure files when you run a mapping on the Spark engine.
- You can read a directory when you run a mapping on the Spark engine.
- You can generate or skip header rows when you run a mapping in the native environment. On the Spark engine, the header row is created by default.
- You can append an existing blob. The append operation is applicable to only to the append blob and in the native environment.
- You can override the blob or container name. In the Blob Container Override field, specify the container name or sub-folders in the root container with the absolute path.
- You can read and write .csv files compressed in the gzip format.

All new functionality for PowerExchange for Microsoft Azure Blob Storage is available for technical preview. Technical preview functionality is supported but is not production-ready. Informatica recommends that you use in non-production environments only.

For more information, see the *Informatica PowerExchange for Microsoft Azure Blob Storage 10.2.1 User Guide*.

PowerExchange for Microsoft Azure SQL Data Warehouse

Effective in version 10.2.1, PowerExchange for Microsoft Azure SQL Data Warehouse includes the following features:

- You can run mappings on the Spark engine.
- You can configure key range partitioning when you read data from Microsoft Azure SQL Data Warehouse objects.
- You can override the SQL query and define constraints when you read data from a Microsoft Azure SQL Data Warehouse object.
- You can configure pre-SQL and post-SQL queries for source and target objects in a mapping.
- You can configure the native expression filter for the source data object operation.
- You can perform update, upsert, and delete operations against Microsoft Azure SQL Data Warehouse tables.
- You can configure a cached lookup operation to cache the lookup table on the Spark engine and an uncached lookup operation in the native environment.

For more information, see the *Informatica PowerExchange for Microsoft Azure SQL Data Warehouse 10.2.1 User Guide*.

PowerExchange for Salesforce

Effective in version 10.2.1, you can use version 41 of Salesforce API to create a Salesforce connection and access Salesforce objects. You can use big objects with source and target transformations.

For more information, see the *Informatica PowerExchange for Salesforce 10.2.1 User Guide*.

PowerExchange for SAP NetWeaver

Effective in version 10.2.1, you can run mappings on the Spark engine to read data from SAP tables.

For more information, see the *Informatica PowerExchange for SAP NetWeaver 10.2.1 User Guide*.

PowerExchange for Snowflake

Effective in version 10.2.1, PowerExchange for Snowflake includes the following features:

- You can configure a lookup operation on a Snowflake table. You can also enable lookup caching for a lookup operation to increase the lookup performance. The Data Integration Service caches the lookup source and runs the query on the rows in the cache.

- You can parameterize the Snowflake connection, and data object read and write operation properties.

- You can configure key range partitioning for Snowflake data objects in a read or write operation. The Data Integration Service distributes the data based on the port or set of ports that you define as the partition key.

- You can specify a table name in the advanced target properties to override the table name in the Snowflake connection properties.

For more information, see the *Informatica PowerExchange for Snowflake 10.2.1 User Guide*.

Security

This section describes new security features in version 10.2.1.

Password Complexity

Effective in version 10.2.1, you can enable password complexity to validate the password strength. By default this option is disabled.

For more information, see the "Security Management in Informatica Administrator" chapter in the *Informatica 10.2.1 Security Guide*.

CHAPTER 2

Changes (10.2.1)

This chapter includes the following topics:

Support Changes

This section describes the support changes in 10.2.1.

Upgrade Support Changes

In version 10.2.1, Informatica supports upgrade for Informatica big data products only, such as Big Data Management and Big Data Quality. When you upgrade the domain, functionality for traditional products such as PowerCenter and Informatica Data Quality will not be available.

If you run traditional and big data products in the same domain, you must split the domain before you upgrade. When you split the domain, you create a copy of the domain so that you can run big data products and traditional products in separate domains. You duplicate the nodes on each machine in the domain. You also duplicate the services that are common to both traditional and big data products. After you split the domain, you can upgrade the domain that runs big data products.

Note: Although Informatica traditional products are not supported in version 10.2.1, the documentation does contain some references to PowerCenter and Metadata Manager services.

Big Data Hadoop Distribution Support

Informatica big data products support a variety of Hadoop distributions. In each release, Informatica adds, defers, and drops support for Hadoop distribution versions. Informatica might reinstate support for deferred versions in a future release.

The following table lists the supported Hadoop distribution versions for Informatica 10.2.1 big data products:

Product	EMR	HDI	CDH	HDP	MapR
Big Data Management	5.10, 5.14[3]	3.6.x	5.11[1], 5.12[1], 5.13, 5.14, 5.15	2.5, 2.6	6.x MEP 5.0.x[2]
Big Data Streaming	5.10, 5.14[3]	3.6.x	5.11[1], 5.12[1], 5.13, 5.14, 5.15	2.5, 2.6	6.x MEP 4.0.x
Enterprise Data Catalog	N/A	3.6.x	5.13	2.6.x	N/A
Enterprise Data Lake	5.10	3.6.x	5.13	2.6.x	N/A

[1] Big Data Management and Big Data Streaming support for CDH 5.11 and 5.12 requires EBF-11719. See KB article 533310.

[2] Big Data Management support for MapR 6.x with MEP 5.0.x requires EBF-12085. See KB article 553273.

[3] Big Data Management and Big Data Streaming support for Amazon EMR 5.14 requires EBF-12444. See KB article 560632.

Note: Informatica dropped support for IBM BigInsights.

To see a list of the latest supported versions, see the Product Availability Matrix on the Informatica Customer Portal: https://network.informatica.com/community/informatica-network/product-availability-matrices.

Big Data Management Hadoop Distributions

The following table lists the supported Hadoop distribution versions and changes in Big Data Management 10.2.1:

Hadoop Distribution	Supported Distribution Versions	10.2.1 Changes
Amazon EMR	5.10, 5.14	Added support for version 5.10 and 5.14. Dropped support for version 5.8.
Azure HDInsight	3.6.x	Added support for version 3.6.x. Dropped support for 3.5x.
Cloudera CDH	5.11, 5.12, 5.13, 5.14, 5.15	Added support for versions 5.13, 5.14, 5.15.

Hadoop Distribution	Supported Distribution Versions	10.2.1 Changes
Hortonworks HDP	2.5.x, 2.6.x	Added support for version 2.6.x. Dropped support for version 2.4.x.
MapR	6.x MEP 5.0.x	Added support for versions 6.x MEP 5.0.x. Dropped support for versions 5.2 MEP 2.0.x, 5.2.MEP 3.0.x.

Note: Informatica dropped support for IBM BigInsights.

Informatica big data products support a variety of Hadoop distributions. In each release, Informatica adds, defers, and drops support for Hadoop distribution versions. Informatica might reinstate support for deferred versions in a future release.

To see a list of the latest supported versions, see the Product Availability Matrix on the Informatica network: https://network.informatica.com/community/informatica-network/product-availability-matrices

Big Data Streaming Hadoop Distributions

The following table lists the supported Hadoop distribution versions and changes in Big Data Streaming 10.2.1:

Hadoop Distribution	Supported Distribution Versions	10.2.1 Changes
Amazon EMR	5.10, 5.14	Added support for 5.10, 5.14. Dropped support for version 5.4.
Azure HDInsight	3.6.x	Added support for version 3.6.x.
Cloudera CDH	5.11, 5.12, 5.13, 5.14, 5.15	Added support for versions 5.13, 5.14, 5.15.
Hortonworks HDP	2.5.x, 2.6.x	Added support for version 2.6.x. Dropped support for version 2.4.x.
MapR	6.x MEP 4.0.x	Added support for versions 6.x MEP 4.0. Dropped support for versions 5.2 MEP 2.0.x, 5.2.MEP 3.0.x.

Informatica big data products support a variety of Hadoop distributions. In each release, Informatica adds, defers, and drops support for Hadoop distribution versions. Informatica might reinstate support for deferred versions in a future release.

To see a list of the latest supported versions, see the Product Availability Matrix on the Informatica network: https://network.informatica.com/community/informatica-network/product-availability-matrices.

Hive Run-Time Engine

Effective in version 10.2.1, the MapReduce mode of the Hive run-time engine is deprecated, and Informatica will drop support for it in a future release. The Tez mode remains supported.

Mapping

When you choose to run a mapping in the Hadoop environment, the Blaze and Spark run-time engines are selected by default.

Previously, the Hive run-time engine was also selected.

If you select Hive to run a mapping, the Data Integration Service will use Tez. You can use the Tez engine only on the following Hadoop distributions:

- Amazon EMR
- Azure HDInsight
- Hortonworks HDP

In a future release, when Informatica drops support for MapReduce, the Data Integration Service will ignore the Hive engine selection and run the mapping on Blaze or Spark.

Profiles

Effective in version 10.2.1, the Hive run-time engine is deprecated, and Informatica will drop support for it in a future release.

The Hive option appears as Hive (deprecated) in Informatica Analyst, Informatica Developer, and Catalog Administrator. You can still choose to run the profiles on the Hive engine. Informatica recommends that you choose the Hadoop option to run the profiles on the Blaze engine.

Installer Changes

Effective in version 10.2.1, the installer includes new functionality and is updated to include the installation and upgrade of all big data products. Enterprise Data Catalog and Enterprise Data Lake installation is combined with the Informatica platform installer.

Install Options

When you run installer, you can choose the install options that fit your requirements.

The following image illustrates the install options and different installer tasks for version 10.2.1:

Choose Informatica Domain Services

Install
- Domain services

Choose Enterprise Data Catalog

Install
- Domain services
- Enterprise Data Catalog

Install
- Enterprise Data Catalog

Choose Enterprise Data Lake

Install
- Domain services
- Enterprise Data Catalog
- Enterprise Data Lake

Install
- Enterprise Data Catalog
- Enterprise Data Lake

Install
- Enterprise Data Lake
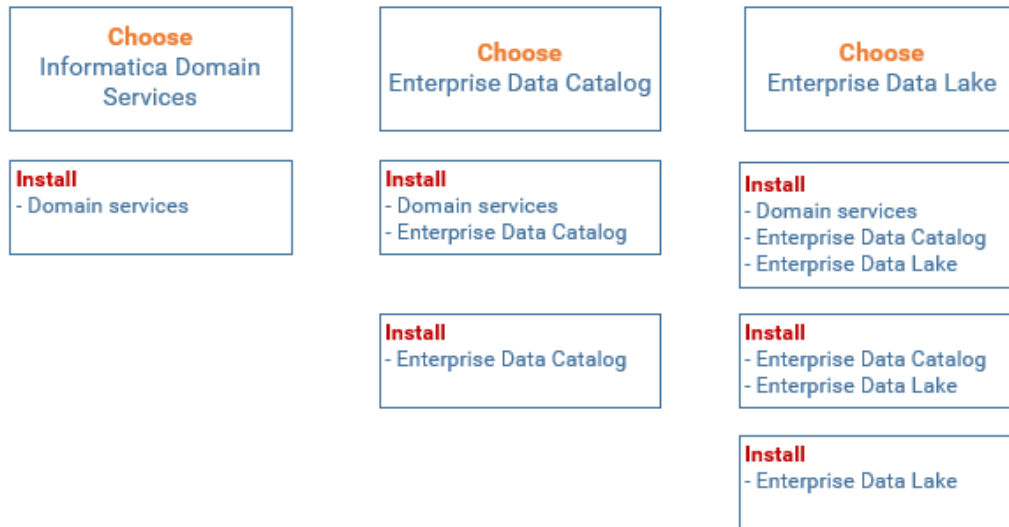
Note: When you install the domain services, the installer also installs application service binaries to support Big Data Management, Bit Data Quality, and Big Data Streaming.

Upgrade Options

When you run the installer, you can choose the upgrade options and actions based on your current installation. When you choose a product to upgrade, the installer upgrades parent products as needed, and either installs or upgrades the product that you choose.

For example, if you choose Enterprise Data Catalog, the installer will upgrade the domain if it is running a previous version. If Enterprise Data Catalog is installed, the installer will upgrade it. If Enterprise Data Catalog is not installed, the installer will install it.

The following image illustrates the upgrade options and the different installer tasks for version 10.2.1:

Choose Informatica Domain Services

- Upgrade domain services

Choose Enterprise Data Catalog

- Upgrade domain services
- Upgrade Enterprise Data Catalog

- Upgrade domain services
- Install Enterprise Data Catalog

Choose Enterprise Data Lake

- Upgrade domain services
- Upgrade Enterprise Data Catalog
- Upgrade Enterprise Data Lake

- Upgrade domain services
- Upgrade Enterprise Data Catalog
- Install Enterprise Data Lake

- Upgrade domain services
- Install Enterprise Data Catalog
- Install Enterprise Data Lake

Note: After the installer performs an upgrade, you need to complete the upgrade of some application services within the Administrator tool.

Installer Task Enhancements

The unified installer is enhanced to perform the following tasks:

- Create a separate monitoring Model Repository Service when you install Informatica domain services.

- Tune the Data Integration Service and the Model Repository Service based on the Big Data Management deployment size.
- Create a cluster configuration and associated connections required by the Enterprise Data Lake.
- Enables the Data Preparation Service for Enterprise Data Lake.

Installer Restricts Traditional Products

The installer includes big data products only. It does not include traditional products such as PowerCenter and Informatica Data Quality. The traditional products and big data products are on separate release trains. If you are upgrading, and the domain includes traditional and big data products, you must split the domain before you upgrade.

Product Name Changes

This section describes changes to product names in version 10.2.1.

The following product names are changed:

- The product Intelligent Data Lake is renamed to Enterprise Data Lake.
- The product Intelligent Streaming is renamed to Big Data Streaming.
- The product Enterprise Information Catalog is renamed to Enterprise Data Catalog.

Application Services

This section describes changes to Application Services in version 10.2.1.

Model Repository Service

Monitoring Model Repository Service

Effective in version 10.2.1, configure a Model Repository Service as a monitoring Model Repository Service to monitor the statistics for ad hoc jobs, applications, logical data objects, SQL data services, web services, and workflows. Use separate database user accounts when you configure monitoring Model repository and Model repository.

Previously, you could use a Model Repository Service to store design-time and run-time objects in the Model repository.

For more information, see the "Model Repository Service" chapter in the *Informatica 10.2.1 Application Service Guide*.

Big Data Management

This section describes the changes to Big Data Management in version 10.2.1.

Azure Storage Access

Effective in version 10.2.1, you must override the properties in the cluster configuration core-site.xml before you run a mapping on the Azure HDInsight cluster.

WASB

If you use a cluster with WASB as storage, you can get the storage account key associated with the HDInsight cluster from the administrator or you can decrypt the encrypted storage account key, and then override the decrypted value in the cluster configuration core-site.xml.

ADLS

If you use a cluster with ADLS as storage, you must copy the client credentials from the web application, and then override the values in the cluster configuration core-site.xml.

Previously, you copied the files from the Hadoop cluster to the machine that runs the Data Integration Service.

Configuring the Hadoop Distribution

This section describes changes to Hadoop distribution configuration.

Hadoop Distribution Configuration

Effective in version 10.2.1, you configure the Hadoop distribution in cluster configuration properties.

The Distribution Name and Distribution Version properties are populated when you import a cluster configuration from the cluster. You can edit the distribution version after you finish the import process.

Previously, the Hadoop distribution was identified by the path to the distribution directory on the machine that hosts the Data Integration Service.

Effective in version 10.2.1, the following property is removed from the Data Integration Service properties:

- Data Integration Service Hadoop Distribution Directory

For more information about the Distribution Name and Distribution Version properties, see the *Big Data Management 10.2.1 Administration Guide*.

MapR Configuration

Effective in version 10.2.1, it is no longer necessary to configure Data Integration Service process properties for the domain when you use Big Data Management with MapR. Big Data Management supports Kerberos authentication with no user action necessary.

Previously, you configured JVM Option properties in the Data Integration Service custom properties, as well as environment variables, to enable support for Kerberos authentication.

For more information about integrating the domain with a MapR cluster, see the *Big Data Management 10.2.1 Hadoop Integration Guide*.

Developer Tool Configuration

Effective in version 10.2.1, you can create a Metadata Access Service. The Metadata Access Service is an application service that allows the Developer tool to access Hadoop connection information to import and preview metadata. When you import an object from a Hadoop cluster, the following adapters use Metadata Access Service to extract the object metadata at design time:

- PowerExchange for HBase
- PowerExchange for HDFS

- PowerExchange for Hive
- PowerExchange for MapR-DB

Previously, you performed the following steps manually on each Developer tool to establish communication between the Developer tool machine and Hadoop cluster at design time:

- Extracted cluster configuration files.
- Ran `krb5.ini` file to import metadata from Hive, HBase, and complex file sources from a kerberos-enabled Hadoop cluster.

The Metadata Access Service eliminates the need to configure each Developer tool machine for design-time connectivity to Hadoop cluster.

For more information, see the "Metadata Access Service" chapter in the *Informatica 10.2.1 Application Service Guide*.

Hadoop Connection Changes

Effective in version 10.2.1, the Hadoop connection contains new and different properties and functionality. These include several properties that you previously configured in other connections or configuration files, and other changes.

This section lists changes to the Hadoop connection in version 10.2.1.

Properties Moved from hadoopEnv.properties to the Hadoop Connection

Effective in version 10.2.1, the properties that you previously configured in the hadoopEnv.properties file are now configurable in advanced properties for the Hadoop connection.

For information about Hive and Hadoop connections, see the *Informatica Big Data Management 10.2.1 User Guide*. For more information about configuring Big Data Management, see the *Informatica Big Data Management 10.2.1 Hadoop Integration Guide*.

Properties Moved from the Hive Connection to the Hadoop Connection

The following Hive connection properties to enable mappings to run on a Hadoop cluster are now in the Hadoop connection:

- Database Name. Namespace for tables. Use the name `default` for tables that do not have a specified database name.
- Advanced Hive/Hadoop Properties. Configures or overrides Hive or Hadoop cluster properties in the hive-site.xml configuration set on the machine on which the Data Integration Service runs. You can specify multiple properties.
- Temporary Table Compression Codec. Hadoop compression library for a compression codec class name.
- Codec Class Name. Codec class name that enables data compression and improves performance on temporary staging tables.

Previously, you configured these properties in the Hive connection.

For information about Hive and Hadoop connections, see the *Informatica Big Data Management 10.2.1 Administrator Guide*.

Advanced Properties for Hadoop Run-time Engines

Effective in version 10.2.1, configure advanced properties for the Blaze, Spark and Hive run-time engines in Hadoop connection properties.

Informatica standardized the property names for run-time engine-related properties. The following table shows the old and new names:

Pre-10.2.1 Property Name	10.2.1 Hadoop Connection Properties Section	10.2.1 Property Name
Blaze Service Custom Properties	Blaze Configuration	Advanced Properties
Spark Execution Parameters	Spark Configuration	Advanced Properties
Hive Custom Properties	Hive Pushdown Configuration	Advanced Properties

Previously, you configured advanced properties for run-time engines in the `hadoopRes.properties` or `hadoopEnv.properties` files, or in the Hadoop Engine Custom Properties field under **Common Properties** in the Administrator tool.

Additional Properties for the Blaze Engine

Effective in version 10.2.1, you can configure an additional property in the Blaze Configuration Properties section of the Hadoop connection properties.

The following table describes the property:

Property	Description
Blaze YARN Node Label	Node label that determines the node on the Hadoop cluster where the Blaze engine runs. If you do not specify a node label, the Blaze engine runs on the nodes in the default partition. If the Hadoop cluster supports logical operators for node labels, you can specify a list of node labels. To list the node labels, use the operators `&&` (AND), `||` (OR), and `!` (NOT).

For more information on using node labels on the Blaze engine, see the "Mappings in the Hadoop Environment" chapter in the *Informatica Big Data Management 10.2.1 User Guide*.

Hive Connection Properties

Effective in version 10.2.1, properties for the Hive connection have changed.

The following Hive connection properties have been removed:

- Access Hive as a source or target
- Use Hive to run mappings in a Hadoop cluster

Previously, these properties were deprecated. Effective in version 10.2.1, they are obsolete.

Configure the following Hive connection properties in the Hadoop connection:

- Database Name
- Advanced Hive/Hadoop Properties
- Temporary Table Compression Codec

- Codec Class Name

Previously, you configured these properties in the Hive connection.

For information about Hive and Hadoop connections, see the *Informatica Big Data Management 10.2.1 User Guide.*

Monitoring

This section describes the changes to monitoring in Big Data Management in version 10.2.1.

Spark Monitoring

Effective in version 10.2.1, changes in Spark monitoring relate to the following areas:

- Event changes
- Updates in the Summary Statistics view

Event Changes

Effective in version 10.2.1, only monitoring information is checked in the Spark events in the session log.

Previously, all the Spark events were relayed as is from the Spark application to the Spark executor. When the events relayed took a long time, performance issues occurred.

For more information, see the *Informatica Big Data Management 10.2.1 User Guide*.

Summary Statistics View

Effective in version 10.2.1, you can view the statistics for Spark execution based on the run stages. For instance, Spark Run Stages shows the statistics of spark application run stages. Stage_0 shows the statistics related to run stage with ID=0 in the spark application. Rows and Average Rows/Sec show the number of rows written out of the stage and the corresponding throughput. Bytes and Average Bytes/Sec show the bytes and throughput broadcasted in the stage.

Previously, you could only view the Source and Target rows and average rows for each second processed for the Spark run.

For more information, see the *Informatica Big Data Management 10.2.1 User Guide*.

Precision and Scale on the Hive Engine

Effective in version 10.2.1, the output of user-defined functions that perform multiplication on the Hive engine can have a maximum scale of 6 if the following conditions are true:

- The difference between the precision and scale is greater than or equal to 32.
- The resultant precision is greater than 38.

Previously, the scale could be as low as 0.

For more information, see the "Mappings in the Hadoop Environment" chapter in the *Informatica Big Data Management 10.2.1 User Guide*.

Sqoop

Effective in version 10.2.1, the following changes apply to Sqoop:

- When you run Sqoop mappings on the Spark engine, the Data Integration Service prints the Sqoop log events in the mapping log. Previously, the Data Integration Service printed the Sqoop log events in the Hadoop cluster log.

  For more information, see the *Informatica Big Data Management 10.2.1 User Guide*.

- If you add or delete a Type 4 JDBC driver .jar file required for Sqoop connectivity from the `externaljdbcjars` directory, changes take effect after you restart the Data Integration Service. If you run the mapping on the Blaze engine, changes take effect after you restart the Data Integration Service and Blaze Grid Manager.

  Note: When you run the mapping for the first time, you do not need to restart the Data Integration Service and Blaze Grid Manager. You need to restart the Data Integration Service and Blaze Grid Manager only for the subsequent mapping runs.

  Previously, you did not have to restart the Data Integration Service and Blaze Grid Manager after you added or deleted a Sqoop .jar file.

  For more information, see the *Informatica Big Data Management 10.2.1 Hadoop Integration Guide*.

Transformation Support on the Hive Engine

Effective in version 10.2.1, a Labeler or Parser transformation that performs probabilistic analysis requires the Java 8 Development Kit on any node on which it runs.

Previously, the transformations required the Java 7 Development Kit.

If you run a mapping that contains a Labeler or Parser transformation that you configured for probabilistic analysis, verify the Java version on the Hive nodes.

Note: On a Blaze or Spark node, the Data Integration Service uses the Java Development Kit that installs with the Informatica engine. Informatica 10.2.1 installs with version 8 of the Java Development Kit.

For more information, see the *Informatica 10.2.1 Installation Guide* or the *Informatica 10.2.1 Upgrade Guide* that applies to the Informatica version that you upgrade.

Big Data Streaming

This section describes changes to Big Data Streaming in version 10.2.1.

Configuring the Hadoop Distribution

Effective in version 10.2.1, you configure the Hadoop distribution in cluster configuration properties.

The Distribution Name and Distribution Version properties are populated when you import a cluster configuration from the cluster. You can edit the distribution version after you finish the import process.

Previously, the Hadoop distribution was identified by the path to the distribution directory on the machine that hosts the Data Integration Service.

For more information about the Distribution Name and Distribution Version properties, see the *Informatica Big Data Management 10.2.1 Administration Guide*.

Developer Tool Configuration

Effective in version 10.2.1, you can create a Metadata Access Service. The Metadata Access Service is an application service that allows the Developer tool to access Hadoop connection information to import and preview metadata.

The following sources and targets use Metadata Access Service at design time to extract the metadata:

- HBase
- HDFS
- Hive
- MapR-DB
- MapRStreams

Previously, you performed the following steps manually on each Developer tool client machine to establish communication between the Developer tool machine and Hadoop cluster at design time:

- Extracted cluster configuration files.
- Ran `krb5.ini` file to import metadata from Hive, HBase, and complex file sources from a kerberos-enabled Hadoop cluster.

The Metadata Access Service eliminates the need to configure each Developer tool machine for design-time connectivity to Hadoop cluster.

For more information, see the "Metadata Access Service" chapter in the *Informatica 10.2.1 Application Service Guide*.

Kafka Connection Properties

Effective in version 10.2.1, properties for the Kafka connection have changed.

You can now configure the Kafka broker version in the connection properties.

Previously, you configured this property in the hadoopEnv.properties file and the hadoopRes.properties file.

For more information about the Kafka connection, see the "Connections" chapter in the *Informatica Big Data Streaming 10.2.1 User Guide*.

Command Line Programs

This section describes changes to commands in 10.2.1

infacmd ihs Commands

Changed Commands

The following table describes changed infacmd ihs commands:

Command	Description
createservice	Effective in 10.2.1, the **-kc** option is added to the createservice command.
createservice	Effective in 10.2.1, the **-bn** option is added to the createservice command.

infacmd ldm Commands

Changed Commands

The following table describes changed infacmd ldm commands:

Command	Change Description
CreateService	Effective in 10.2.1, the **-lt** option is added to the CreateService command.
CreateService	Effective in 10.2.1, the **-dis** option is removed from the CreateService command.
CreateService	Effective in 10.2.1, the **-cms** option is removed from the CreateService command.

For more information, see the *Informatica 10.2 .1 Command Reference*.

Content Installer

Effective in version 10.2.1. Informatica no longer provides a Content Installer utility for accelerator files and reference data files. To add accelerator files or reference data files to an Informatica installation, extract and copy the files to the appropriate directories in the installation.

Previously, you used the Content Installer to extract and copy the files to the Informatica directories.

For more information, see the *Informatica 10.2.1 Content Guide*.

Enterprise Data Catalog

This section describes the changes to Informatica Enterprise Data Catalog in version 10.2.1.

Additional Properties Section in the General Tab

Effective in version 10.2.1, when you create a resource, you can assign custom attribute values to a resource in the **Additional Properties** section of the General tab. Custom attribute values that you can assign include Department, Data Owner, Data Steward, and Subject Matter Experts.

For more information about assigning custom attributes, see the *Informatica 10.2 .1 Catalog Administrator* and *Informatica 10.2 .1 Enterprise Data Catalog User Guide*.

Connection Assignment

Effective in version 10.2.1, you can assign a database to a connection for a PowerCenter resource.

For more information about connection assignment, see the *Informatica 10.2 .1 Catalog Administrator Guide*.

Column Similarity

Effective in version 10.2.1, you can discover similar columns based on column names, column patterns, unique values, and value frequencies in a resource.

Previously, the Similarity Discovery system resource identified similar columns in the source data.

For more information about column similarity, see the *Informatica 10.2 .1 Catalog Administrator Guide*.

Create a Catalog Service

Effective in version 10.2.1, when you create a Catalog Service, you do not have to provide the details of the Data Integration Service and the Content Management Service that you want to associate with the Catalog Service.

For more information, see the *Informatica Enterprise Data Catalog 10.2 .1 Installation and Configuration Guide*.

HDFS Resource Type Enhancemets

Effective in version 10.2.1, you can now use one of the following Hadoop distribution types for an HDFS resource:

- Hortonworks
- IBM BigInsights
- Azure HDInsight
- Amazon EMR
- MapR FS

Hive Resources

Effective in version 10.2.1, when you create a Hive resource, and choose Hive as the **Run On** option, you need to select a Hadoop connection to run the profiling scanner on the Hive engine.

Previously, a Hadoop connection was not required to run the profiling scanner on Hive resources.

For more information about Hive resources, see the *Informatica 10.2 .1 Catalog Administrator Guide*.

Informatica Platform Scanner

Effective in version 10.2.1, you can use the parameter file and parameter set options to extract detailed lineage using the Informatica Platform scanner.

Overview Tab

Effective in version 10.2.1, the Asset Details view is titled Overview in Enterprise Data Catalog.

You can now view the details of an asset in the Overview tab. The Overview tab displays the different sections, such as the source description, description, people, business terms, business classifications, system properties, and other properties. The sections that the Overview tab displays depends on the type of the asset.

For more information about the overview of assets, see the "View Assets" chapter in the *Informatica Enterprise Data Catalog 10.2.1 User Guide*.

Product Name Changes

Effective in version 10.2.1, Enterprise Data Catalog includes the following name changes:

- The product name is changed to Informatica Enterprise Data Catalog. Previously, the product name was Enterprise Information Catalog.
- The installer name is changed to Enterprise Data Catalog. Previously, the installer name was Enterprise Information Catalog.

Proximity Data Domains

Effective in version 10.2.1, you can add one or more data domains as proximity data domains when you create or edit a data domain that has data rules or column rules. The profiling scanner scans the data source for the data domain and the proximity data domains in the resource and displays a match score in Enterprise Data Catalog. The match score is the ratio of the number of proximal data domains discovered in the data source to the number of configured proximal data domains for an inferred data domain.

Previously, you could add proximity rules to a data domain that had a data rule. If the data domains were not found in the source tables, the data conformance percentage for the data domain was reduced in the source tables by the specified value.

For more information about proximity data domains, see the *Informatica 10.2 .1 Catalog Administrator Guide*.

Search Results

Effective in version 10.2.1, the search results page includes the following changes:

- You can now sort the search results based on the asset name and relevance. Previously, you could sort the search results based on the asset name, relevance, system attributes, and custom attributes.
- You can now add a business title to an asset from the search results. Previously, you could associate only a business term.
- The search results page now displays the asset details of assets, such as the resource name, source description, description, path to asset, and asset type. Previously, you could view the details, such as the asset type, resource type, the date the asset was last updated, and size of the asset.

For more information about search results, see the *Informatica Enterprise Data Catalog 10.2.1 User Guide*.

Universal Connectivity Framework

Effective in version 10.2.1, all the resources that you create using the Universal Connectivity Framework require the Catalog Agent to be up and running.

Previously, only resources that run on Microsoft Windows required the Catalog Agent to be up and running..

For more information, see the *Informatica 10.2 .1 Catalog Administrator Guide*.

Informatica Analyst

This section describes changes to the Analyst tool in version 10.2.1.

Scorecards

This section describes the changes to scorecard behavior in version 10.2.1.

Edit Existing Metrics in a Scorecard

Effective in version 10.2.1, you cannot edit existing metrics or metric groups when you add columns to an existing scorecard. To modify the existing metrics or metric groups in the scorecard, navigate to the Scorecard workspace, edit the scorecard, and modify the metrics.

Previously, you could view and edit the existing metrics or metric groups when you add the columns to an existing scorecard.

For more information about scorecards, see the *Informatica 10.2 .1 Data Discovery Guide*.

Configure Threshold for a Metric

Effective in version 10.2.1, you can configure a decimal number up to two decimal places as the threshold for a metric in a scorecard.

Previously, you could configure only integer values as the threshold value for a metric.

For more information about scorecards, see the *Informatica 10.2 .1 Data Discovery Guide*.

Informatica Developer

This section describes changes to the Developer tool in version 10.2.1.

Importing and Exporting Objects from and to PowerCenter

Effective in version 10.2.1, the Developer tool does not include options to import objects from and export objects to PowerCenter.

Informatica Transformations

This section describes changes to the Informatica transformations in version 10.2.1.

Address Validator Transformation

This section describes changes to the Address Validator transformation in version 10.2.1.

The Address Validator transformation contains the following updates to address functionality:

All Countries

Effective in version 10.2.1, the Address Validator transformation uses version 5.12.0 of the Informatica Address Verification software engine. The engine enables the features that Informatica adds to the Address Validator transformation in version 10.2.1.

Previously, the transformation used version 5.11.0 of the Informatica Address Verification software engine.

United Kingdom

Effective from November 2017, Informatica ceases the delivery of reference data files that contain the names and addresses of businesses in the United Kingdom. Informatica ceases to support the verification of the business names and addresses.

For comprehensive information about the features and operations of the address verification software engine version that Informatica embeds in version 10.2.1, see the *Informatica Address Verification 5.12.0 Developer Guide*.

Data Transformation

This section describes changes to the Data Processor transformation in version 10.2.1.

Effective in 10.2.1, Data Processor transformation performs strict validation for hierarchical input. When strict validation applies, the hierarchical input file must conform strictly to its schema. This option can be applied when the Data Processor mode is set to Output Mapping, which creates output ports for relational output.

This option does not apply to mappings with JSON input from versions previous to version 10.2.1.

For more information, see the *Data Transformation 10.2.1 User Guide*.

Sequence Generator Transformation

This section describes changes to the Sequence Generator transformation in version 10.2.1.

Maintain Row Order

Effective in version 10.2.1, the Maintain Row Order property for the Sequence Generator transformation is set to False by default.

Previously, the default value was True.

If you upgrade from an earlier version, the Maintain Row Order property on any Sequence Generator transformation in the repository does not change.

For more information, see the "Sequence Generator Transformation" chapter in the *Informatica 10.2.1 Developer Transformation Guide*.

Sorter Transformation

This section describes changes to the Sorter transformation in version 10.2.1.

Sorter Caches

Effective in version 10.2.1, the sorter cache for the Sorter transformation uses variable length to store data up to 8 MB in the native environment and on the Blaze engine in the Hadoop environment.

Previously, the sorter cache used variable length to store data up to 64 KB. For data that exceeded 64 KB, the sorter cache stored the data using fixed length.

For more information, see the "Sorter Transformation" chapter in the *Informatica 10.2.1 Developer Transformation Guide*.

Sorter Performance

Effective in version 10.2.1, the Sorter transformation is optimized to perform faster sort key comparisons for data up to 8 MB.

The sort key comparison rate is not optimized in the following situations:

- Binary sort order is not selected.
- The sort key is a timestamp with time zone data type.
- You perform case-sensitive string comparison and any of the sort key columns is a string data type.

Previously, the Sorter transformation performed faster sort key comparisons for data up to 64 KB. For data that exceeded 64 KB, the sort key comparison rate was not optimized.

For more information, see the "Sorter Transformation" chapter in the *Informatica 10.2.1 Developer Transformation Guide*.

PowerExchange Adapters for Informatica

This section describes changes to Informatica adapters in version 10.2.1.

PowerExchange for Amazon Redshift

Effective in version 10.2.1, after you connect to PowerExchange for Amazon Redshift, the following prerequisite tasks are completed automatically:

- The required Amazon Redshift JDBC `.jar` file are downloaded.
- The `.jar` file are copied on the node that runs on the Data Integration Service and on the client machine.

Previously, you had to perform the prerequisite tasks manually and restart the Data Integration Service before you can use PowerExchange for Amazon Redshift.

PowerExchange for Cassandra

Effective in version 10.2.1, PowerExchange for Cassandra has the following changes:

- The name and directory of the Informatica PowerExchange for Cassandra ODBC driver file has changed.

The following table lists the Cassandra ODBC driver file name and file directory based on Linux and Windows operating systems:

Operating System	Cassandra ODBC Driver File Name	File Directory
Linux	`libcassandraodbc_sb64.so`	`<Informatica installation directory>\tools` `\cassandra\lib\libcassandraodbc_sb64.so`
Windows	`CassandraODBC_sb64.dll`	`<Informatica installation directory>\tools` `\cassandra\lib\CassandraODBC_sb64.dll`

On Linux operating systems, you must update the value of the `Driver` property to `<Informatica installation directory>\tools\cassandra\lib\libcassandraodbc_sb64.so` for the existing Cassandra data sources in the `odbc.ini` file.

On Windows, you must update the following properties in the Windows registry for the existing Cassandra data source name:

```
Driver=<Informatica installation directory>\tools\cassandra\lib\CassandraODBC_sb64.dll
Setup=<Informatica installation directory>\tools\cassandra\lib\CassandraODBC_sb64.dll
```

- The new key name for Load Balancing Policy option is LoadBalancingPolicy.
  Previously, the key name for Load Balancing Policy was COLoadBalancingPolicy

- The default values of the following Cassandra ODBC driver properties has changed:

Driver Property Name	Key Name	New Default Value
Concurrent Requests	NumConcurrentRequests	100
Insert Query Threads	NumInsertQueryThreads	2
Iterations Per Insert Thread	NumIterationsPerInsertThread	50

For more information, see the *Informatica PowerExchange for Cassandra 10.2.1 User Guide*.

PowerExchange for Snowflake

Effective in version 10.2.1, PowerExchange for Snowflake installs with Informatica 10.2.1.

Previously, PowerExchange for Snowflake had a separate installer.

For more information, see the *Informatica PowerExchange for Snowflake 10.2.1 User Guide* .

CHAPTER 3

Release Tasks (10.2.1)

This chapter includes the following topic:

PowerExchange Adapters for Informatica

This section describes release tasks for Informatica adapters in version 10.2.1.

PowerExchange Adapters for Amazon S3

Effective in version 10.2.1, to successfully preview data from the Avro and Parquet files or run a mapping in the native environment with the Avro and Parquet files, you must configure the INFA_PARSER_HOME property for the Data Integration Service in Informatica Administrator. Perform the following steps to configure the INFA_PARSER_HOME property:

- Log in to Informatica Administrator.
- Click the Data Integration Service and then click the **Processes** tab on the right pane.
- Click **Edit** in the **Environment Variables** section.
- Click **New** to add an environment variable.
- Enter the name of the environment variable as **INFA_PARSER_HOME**.
- Set the value of the environment variable to the absolute path of the Hadoop distribution directory on the machine that runs the Data Integration Service. Verify that the version of the Hadoop distribution directory that you define in the INFA_PARSER_HOME property is the same as the version you defined in the cluster configuration.

For more information, see the *Informatica PowerExchange for Amazon S3 10.2.1 User Guide*.

Part II: 10.2

This part contains the following chapters:

CHAPTER 4

New Products (10.2)

This chapter includes the following topic:

-

PowerExchange Adapters

PowerExchange Adapters for Informatica

This section describes new Informatica adapters in 10.2.

PowerExchange for Microsoft Azure Data Lake Store

Effective in version 10.2, you can create a Microsoft Azure Data Lake Store connection to specify the location of Microsoft Azure Data Lake Store sources and targets you want to include in a data object. You can use the Microsoft Azure Data Lake Store connection in data object read and write operations. You can validate and run mappings in the native environment or on the Blaze engine in the Hadoop environment.

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

CHAPTER 5

New Features (10.2)

This chapter includes the following topics:

Application Services

This section describes new application service features in 10.2.

Model Repository Service

This section describes new Model Repository Service features in 10.2.

Import Objects from Previous Versions

Effective in version 10.2, you can use infacmd to upgrade objects exported from an Informatica 10.1 or 10.1.1 Model repository to the current metadata format, and then import the upgraded objects into the current Informatica release.

For more information, see the "Object Import and Export" chapter in the *Informatica 10.2 Developer Tool Guide*, or the "infacmd mrs Command Reference" chapter in the *Informatica 10.2 Command Reference*.

Big Data

This section describes new big data features in 10.2.

Big Data Management Installation

Effective in version 10.2, the Data Integration Service automatically installs the Big Data Management binaries on the cluster.

When you run a mapping , the Data Integration Service checks for the binary files on the cluster. If they do not exist or if they are not synchronized, the Data Integration Service prepares the files for transfer. It transfers the files to the distributed cache through the Informatica Hadoop staging directory on HDFS. By default, the staging directory is /tmp. This process replaces the requirement to install distribution packages on the Hadoop cluster.

For more information, see the *Informatica Big Data Management 10.2 Hadoop Integration Guide*.

Cluster Configuration

A cluster configuration is an object in the domain that contains configuration information about the Hadoop cluster. The cluster configuration enables the Data Integration Service to push mapping logic to the Hadoop environment.

When you create the cluster configuration, you import cluster configuration properties that are contained in configuration site files. You can import these properties directly from a cluster or from a cluster configuration archive file. You can also create connections to associate with the cluster configuration.

Previously, you ran the Hadoop Configuration Manager utility to configure connections and other information to enable the Informatica domain to communicate with the cluster.

For more information about cluster configuration, see the "Cluster Configuration" chapter in the *Informatica Big Data Management 10.2 Administrator Guide*.

Processing Hierarchical Data

Effective in version 10.2, you can use complex data types, such as array, struct, and map, in mappings that run on the Spark engine. With complex data types, the Spark engine directly reads, processes, and writes hierarchical data in Avro, JSON, and Parquet complex files.

Develop mappings with complex ports, operators, and functions to perform the following tasks:

- Generate and modify hierarchical data.
- Transform relational data to hierarchical data.
- Transform hierarchical data to relational data.

- Convert data from one complex file format to another.

When you process hierarchical data, you can use hierarchical conversion wizards to simplify the mapping development tasks. Use these wizards in the following scenarios:

- To generate hierarchical data of type struct from one or more ports.
- To generate hierarchical data of a nested struct type from ports in two transformations.
- To extract elements from hierarchical data in a complex port.
- To flatten hierarchical data in a complex port.

For more information, see the "Processing Hierarchical Data on the Spark Engine" chapter in the *Informatica Big Data Management 10.2 User Guide*.

Stateful Computing on the Spark Engine

Effective in version 10.2, you can use window functions in an Expression transformation to perform stateful calculations on the Spark engine. Window functions operate on a group of rows and calculate a single return value for every input row. You can use window functions to perform the following tasks:

- Retrieve data from previous or subsequent rows.
- Calculate a cumulative sum based on a group of rows.
- Calculate a cumulative average based on a group of rows.

For more information, see the "Stateful Computing on the Spark Engine" chapter of the *Big Data Management 10.2 User Guide*.

Data Integration Service Queuing

Effective in version 10.2, if you deploy multiple mapping jobs or workflow mapping tasks at the same time, the Data Integration Service queues the jobs in a persisted queue and runs the jobs when resources are available. You can view the current status of mapping jobs on the Monitor tab of the Administrator tool.

All queues are persisted by default. If the Data Integration Service node shuts down unexpectedly, the queue does not fail over when the Data Integration Service fails over. The queue remains on the Data Integration Service machine, and the Data Integration Service resumes processing the queue when you restart it.

By default, each queue can hold 10,000 jobs at a time. When the queue is full, the Data Integration Service rejects job requests and marks them as failed. When the Data Integration Service starts running jobs in the queue, you can deploy additional jobs.

For more information, see the "Queuing" chapter in the *Informatica Big Data Management 10.2 Administrator Guide*.

Blaze Job Monitor

Effective in version 10.2, you can configure the host and port number to start the Blaze Job Monitor application in the Hadoop connection properties. The default value is <hostname>:9080. If you do not configure the host name, the Blaze engine uses the first alphabetical node in the cluster.

For more information, see the "Connections" chapter in the *Big Data Management 10.2 User Guide*.

Data Integration Service Properties for Hadoop Integration

Effective in version 10.2, the Data Integration Service added properties required to integrate the domain with the Hadoop environment.

The following table describes the new properties:

Property	Description
Hadoop Staging Directory	The HDFS directory where the Data Integration Services pushes Informatica Hadoop binaries and stores temporary files during processing. Default is /tmp.
Hadoop Staging User	Required if the Data Integration Service user is empty. The HDFS user that performs operations on the Hadoop staging directory. The user needs write permissions on Hadoop staging directory. Default is the Data Integration Service user.
Custom Hadoop OS Path	The local path to the Informatica Hadoop binaries compatible with the Hadoop operating system. Required when the Hadoop cluster and the Data Integration Service are on different supported operating systems.
	Download and extract the Informatica binaries for the Hadoop cluster on the machine that hosts the Data Integration Service. The Data Integration Service uses the binaries in this directory to integrate the domain with the Hadoop cluster.
	The Data Integration Service can synchronize the following operating systems: - SUSE 11 and Redhat 6.5
	Changes take effect after you recycle the Data Integration Service.

As a result of the changes in cluster integration, the following properties are removed from the Data Integration Service:

- Informatica Home Directory on Hadoop
- Hadoop Distribution Directory

For more information, see the *Informatica 10.2 Hadoop Integration Guide*.

Sqoop

Effective in version 10.2, if you use Sqoop data objects, you can use the following specialized Sqoop connectors to run mappings on the Spark engine:

- Cloudera Connector Powered by Teradata
- Hortonworks Connector for Teradata

These specialized connectors use native protocols to connect to the Teradata database.

For more information, see the *Informatica Big Data Management 10.2 User Guide*.

Autoscaling in an Amazon EMR Cluster

Effective in version 10.2, Big Data Management adds support for Spark mappings to take advantage of autoscaling in an Amazon EMR cluster.

Autoscaling enables the EMR cluster administrator to establish threshold-based rules for adding and subtracting cluster task and core nodes. Big Data Management certifies support for Spark mappings that run on an autoscaling-enabled EMR cluster.

Transformation Support on the Blaze Engine

Effective in version 10.2, the following transformations have additional support on the Blaze engine

- Update Strategy. Supports targets that are ORC bucketed on all columns.

For more information, see the "Mapping Objects in a Hadoop Environment" chapter in the *Informatica Big Data Management 10.2 User Guide.*

Hive Functionality for the Blaze Engine

Effective in version 10.2, mappings that run on the Blaze engine can read and write to bucketed and sorted targets.

For information about how to configure mappings for the Blaze engine, see the "Mappings in a Hadoop Environment" chapter in the *Informatica Big Data Management 10.2 User Guide.*

Transformation Support on the Spark Engine

Effective in version 10.2, the following transformations are supported with restrictions on the Spark engine:

- Normalizer
- Rank
- Update Strategy

Effective in version 10.2, the following transformations have additional support on the Spark engine:

- Lookup. Supports unconnected lookup from the Filter, Aggregator, Router, Expression, and Update Strategy transformation.

For more information, see the "Mapping Objects in a Hadoop Environment" chapter in the *Informatica Big Data Management 10.2 User Guide.*

Hive Functionality for the Spark Engine

Effective in version 10.2, the following functionality is supported for mappings that run on the Spark engine:

- Reading and writing to Hive resources in Amazon S3 buckets
- Reading and writing to transactional Hive tables
- Reading and writing to Hive table columns that are secured with fine-grained SQL authorization

For information about how to configure mappings for the Spark engine, see the "Mappings in a Hadoop Environment" chapter in the *Informatica Big Data Management 10.2 User Guide.*

Command Line Programs

This section describes new commands in 10.2.

infacmd cluster Commands

cluster is a new infacmd plugin that performs operations on cluster configurations.

The following table describes new infacmd cluster commands:

Command	Description
clearConfigurationProperties	Clears overridden property values in the cluster configuration set.
createConfiguration	Creates a new cluster configuration either from XML files or remote cluster manager.
deleteConfiguration	Deletes a cluster configuration from the domain.
exportConfiguration	Exports a cluster configuration to a compressed file or a combined XML file.
listAssociatedConnections	Lists connections by type that are associated with the specified cluster configuration.
listConfigurationGroupPermissions	Lists the permissions that a group has for a cluster configuration.
listConfigurationSets	Lists configuration sets in the cluster configuration.
listConfigurationProperties	Lists configuration properties in the cluster configuration set.
listConfigurations	Lists cluster configuration names.
listConfigurationUserPermissions	Lists the permissions that a user has for a cluster configuration.
refreshConfiguration	Refreshes a cluster configuration either from XML files or remote cluster manager.
setConfigurationPermissions	Sets permissions on cluster configuration to a user or a group after removing previous permissions.
setConfigurationProperties	Sets overridden property values in the cluster configuration set.

For more information, see the "infacmd cluster Command Reference" chapter in the *Informatica 10.2 Command Reference.*

infacmd dis Options

The following table describes new Data Integration Service options for infacmd UpdateServiceOptions:

Command	Description
ExecutionOptions.MaxHadoopBatchExecutionPoolSize	The maximum number of deployed Hadoop jobs that can run concurrently.
ExecutionOptions.MaxNativeBatchExecutionPoolSize	The maximum number of deployed native jobs that each Data Integration Service process can run concurrently.

Command	Description
ExecutionOptions.MaxOnDemandExecutionPoolSize	The maximum number of on-demand jobs that can run concurrently. Jobs include data previews, profiling jobs, REST and SQL queries, web service requests, and mappings run from the Developer tool.
WorkflowOrchestrationServiceOptions.MaxWorkerThreads	The maximum number of threads that the Data Integration Service can use to run parallel tasks between a pair of inclusive gateways in a workflow. The default value is 10. If the number of tasks between the inclusive gateways is greater than the maximum value, the Data Integration Service runs the tasks in batches that the value specifies.

For more information, see the "infacmd dis Command Reference" chapter in the *Informatica 10.2 Command Reference*.

infacmd ipc Commands

The following table describes a new option for an infacmd ipc command:

Command	Description
genReuseReportFromPC	Contains the following new option: -BlockSize: Optional. The number of mappings that you want to run the infacmd ipc genReuseReportFromPC command against.

For more information, see the "infacmd ipc Command Reference" chapter in the *Informatica 10.2 Command Reference*.

infacmd isp Commands

The following table describes changes to infacmd isp commands:

Command	Description
createConnection	Defines a connection and the connection options. Added, changed, and removed Hadoop connection options. See infacmd isp createConnection.
getDomainSamlConfig	Renamed from getSamlConfig. Returns the value of the cst option set for Secure Assertion Markup Language (SAML) authentication. Specifies the allowed time difference between the Active Directory Federation Services (AD FS) host system clock and the system clock on the master gateway node.

Command	Description
getUserActivityLog	Returns user activity log data, which now includes successful and unsuccessful user login attempts from Informatica clients. The user activity data includes the following properties for each login attempt from an Informatica client: - Application name - Application version - Host name or IP address of the application host If the client sets custom properties on login requests, the data includes the custom properties.
listConnections	Lists connection names by type. You can list by all connection types or filter the results by one connection type. The -ct option is now available for the command. Use the -ct option to filter connection types.
purgeLog	Purges log events and database records for license usage. The -lu option is now obsolete.
SwitchToGatewayNode	The following options are added for configuring SAML authentication: - asca. The alias name specified when importing the identity provider assertion signing certificate into the truststore file used for SAML authentication. - saml. Enabled or disabled SAML authentication in the Informatica domain. - std. The directory containing the custom truststore file required to use SAML authentication on gateway nodes within the domain. - stp. The custom truststore password used for SAML authentication.

For more information, see the "infacmd isp Command Reference" chapter in the *Informatica 10.2 Command Reference*.

infacmd isp createConnection

This section lists new, changed, and removed Hadoop connection options for the property infacmd isp createConnection in 10.2.

Hadoop Connection Options

The following tables describes new Hadoop connection options available in 10.2:

Option	Description
clusterConfigId	The cluster configuration ID associated with the Hadoop cluster.
blazeJobMonitorURL	The host name and port number for the Blaze Job Monitor.
rejDirOnHadoop	Enables hadoopRejDir. Used to specify a location to move reject files when you run mappings.
hadoopRejDir	The remote directory where the Data Integration Service moves reject files when you run mappings. Enable the reject directory using rejDirOnHadoop.

Option	Description
sparkEventLogDir	An optional HDFS file path of the directory that the Spark engine uses to log events.
sparkYarnQueueName	The YARN scheduler queue name used by the Spark engine that specifies available resources on a cluster.

The following table describes Hadoop connection options that are renamed in 10.2:

Current Name	Previous Name	Description
blazeYarnQueueName	cadiAppYarnQueueName	The YARN scheduler queue name used by the Blaze engine that specifies available resources on a cluster. The name is case sensitive.
blazeExecutionParameterList	cadiExecutionParameterList	Custom properties that are unique to the Blaze engine.
blazeMaxPort	cadiMaxPort	The maximum value for the port number range for the Blaze engine.
blazeMinPort	cadiMinPort	The minimum value for the port number range for the Blaze engine.
blazeUserName	cadiUserName	The owner of the Blaze service and Blaze service logs.
blazeStagingDirectory	cadiWorkingDirectory	The HDFS file path of the directory that the Blaze engine uses to store temporary files.
hiveStagingDatabaseName	databaseName	Namespace for Hive staging tables.
impersonationUserName	hiveUserName	Hadoop impersonation user. The user name that the Data Integration Service impersonates to run mappings in the Hadoop environment.
sparkStagingDirectory	SparkHDFSStagingDir	The HDFS file path of the directory that the Spark engine uses to store temporary files for running jobs.

The following table describes Hadoop connection options that are removed from the UI and imported into the cluster configuration:

Option	Description
RMAddress	The service within Hadoop that submits requests for resources or spawns YARN applications. Imported into the cluster configuration as the property `yarn.resourcemanager.address`.
defaultFSURI	The URI to access the default Hadoop Distributed File System. Imported into the cluster configuration as the property `fs.defaultFS` or `fs.default.name`.

The following table describes Hadoop connection options that are deprecated in 10.2 and are no longer available in the UI:

Option	Description
metastoreDatabaseDriver*	Driver class name for the JDBC data store.
metastoreDatabasePassword*	The password for the metastore user name.
metastoreDatabaseURI*	The JDBC connection URI used to access the data store in a local metastore setup.
metastoreDatabaseUserName*	The metastore database user name.
metastoreMode*	Controls whether to connect to a remote metastore or a local metastore.
remoteMetastoreURI*	The metastore URI used to access metadata in a remote metastore setup. This property is imported into the cluster configuration as the property `hive.metastore.uris`.
jobMonitoringURL	The URL for the MapReduce JobHistory server.
These properties are deprecated in 10.2. When you upgrade to 10.2, the property values you set in a previous release are saved in the repository, but they do not appear in the connection properties.	

The following properties are dropped. If they appear in connection strings, they will have no effect:

- hadoopClusterInfoExecutionParametersList
- passThroughSecurityEnabled
- hiverserver2Enabled
- hiveInfoExecutionParametersList
- cadiPassword
- sparkMaster
- sparkDeployMode

HBase Connection

The following table describes HBase connection options that are removed from the connection and imported into the cluster configuration:

Property	Description
ZOOKEEPERHOSTS	Name of the machine that hosts the ZooKeeper server.
ZOOKEEPERPORT	Port number of the machine that hosts the ZooKeeper server.
ISKERBEROSENABLED	Enables the Informatica domain to communicate with the HBase master server or region server that uses Kerberos authentication.
hbaseMasterPrincipal	Service Principal Name (SPN) of the HBase master server.
hbaseRegionServerPrincipal	Service Principal Name (SPN) of the HBase region server.

Hive Connection

The following table describes Hive connection options that are removed from the connection and imported into the cluster configuration:

Property	Description
defaultFSURI	The URI to access the default Hadoop Distributed File System.
jobTrackerURI	The service within Hadoop that submits the MapReduce tasks to specific nodes in the cluster.
hiveWarehouseDirectoryOnHDFS	The absolute HDFS file path of the default database for the warehouse that is local to the cluster.
metastoreExecutionMode	Controls whether to connect to a remote metastore or a local metastore.
metastoreDatabaseURI	The JDBC connection URI used to access the data store in a local metastore setup.
metastoreDatabaseDriver	Driver class name for the JDBC data store.
metastoreDatabaseUserName	The metastore database user name.
metastoreDatabasePassword	The password for the metastore user name.
remoteMetastoreURI	The metastore URI used to access metadata in a remote metastore setup. This property is imported into the cluster configuration as the property `hive.metastore.uris`.

HBase Connection Options for MapR-DB

The ISKERBEROSENABLED connection option is obsolete and imported into the cluster configuration.

infacmd mrs Commands

The following table describes new infacmd mrs commands:

Command	Description
manageGroupPermissionOnProject	Manages permissions on multiple projects for a group.
manageUserPermissionOnProject	Manages permissions on multiple projects for a user.
upgradeExportedObjects	Upgrades objects exported to an .xml file from a previous Informatica release to the current metadata format. The command generates an .xml file that contains the upgraded objects.

For more information, see the "infacmd mrs Command Reference" chapter in the *Informatica 10.2 Command Reference*.

infacmd ms Commands

The following table describes new infacmd ms commands:

Command	Description
GetMappingStatus	Gets the current status of a mapping job by job ID.

For more information, see the "infacmd ms Command Reference" chapter in the *Informatica 10.2 Command Reference*.

infacmd wfs Commands

The following table describes new infacmd wfs commands:

Command	Description
completeTask	Completes a Human task instance that you specify.
delegateTask	Assigns ownership of a Human task instance to a user or group.
listTasks	Lists the Human task instances that meet the filter criteria that you specify.
releaseTask	Releases a Human task instance from the current owner, and returns ownership of the task instance to the business administrator that the workflow configuration identifies.
startTask	Changes the status of a Human task instance to IN_PROGRESS.

For more information, see the "infacmd wfs Command Reference" chapter in the *Informatica 10.2 Command Reference*.

infasetup Commands

The following table describes changes to infasetup commands:

Command	Description
DefineDomain	The following options are added for configuring Secure Assertion Markup Language (SAML) authentication: - asca. The alias name specified when importing the identity provider assertion signing certificate into the truststore file used for SAML authentication. - cst. The allowed time difference between the Active Directory Federation Services (AD FS) host system clock and the system clock on the master gateway node. - std. The directory containing the custom truststore file required to use SAML authentication on gateway nodes within the domain. - stp. The custom truststore password used for SAML authentication.
DefineGatewayNode	The following options are added for configuring SAML authentication: - asca. The alias name specified when importing the identity provider assertion signing certificate into the truststore file used for SAML authentication. - saml. Enables or disables SAML authentication in the Informatica domain. - std. The directory containing the custom truststore file required to use SAML authentication on gateway nodes within the domain. - stp. The custom truststore password used for SAML authentication.

Command	Description
UpdateDomainSaml Config	Renamed from UpdateSamlConfig. The following option is added for configuring SAML authentication: - cst. The allowed time difference between the AD FS host system clock and the system clock on the master gateway node.
UpdateGatewayNod e	The following options are added for configuring SAML authentication. - asca. The alias name specified when importing the identity provider assertion signing certificate into the truststore file used for SAML authentication. - saml. Enables or disables SAML authentication in the Informatica domain. - std. The directory containing the custom truststore file required to use SAML authentication on gateway nodes within the domain. - stp. The custom truststore password used for SAML authentication.

For more information, see the "infasetup Command Reference" chapter in the *Informatica 10.2 Command Reference*.

pmrep Commands

The following table describes new pmrep commands:

Command	Description
CreateQuery	Creates a query in the repository.
DeleteQuery	Deletes a query from the repository.

The following table describes updates to pmrep commands:

Command	Description
CreateConnection	Contains the following updated option: -w. Enables you to use a parameter in the password option.
ListObjectDependencies	Contains the following updated option: -o. The object type list includes query and deploymentgroup.
UpdateConnection	Contains the following updated options: -w. Enables you to use a parameter in the password option. -x. Disables the use of password parameters if you use the parameter in password.

For more information, see the "pmrep Command Reference" chapter in the *Informatica 10.2 Command Reference*.

Data Types

This section describes new data type features in 10.2.

Informatica Data Types

This section describes new data types in the Developer tool.

Complex Data Types

Effective in version 10.2, some transformations support complex data types in mappings that run on the Spark engine.

The following table describes the complex data types you can use in transformations:

Complex Data Type	Description
array	Contains an ordered collection of elements. All elements in the array must be of the same data type. The elements can be of primitive or complex data type.
map	Contains an unordered collection of key-value pairs. The key part must be of primitive data type. The value part can be of primitive or complex data type.
struct	Contains a collection of elements of different data types. The elements can be of primitive or complex data types.

For more information, see the "Data Type Reference" appendix in the *Informatica Big Data Management 10.2 User Guide*.

Documentation

This section describes new or updated guides in 10.2.

The Informatica documentation contains the following changes:

Informatica Big Data Management Security Guide

Effective in version 10.2, the *Informatica Big Data Management Security Guide* is renamed to *Informatica Big Data Management Administrator Guide*. It contains the security information and additional administrator tasks for Big Data Management.

For more information see the *Informatica Big Data Management 10.2 Administrator Guide*.

Informatica Big Data Management Installation and Upgrade Guide

Effective in version 10.2, the *Informatica Big Data Management Installation and Upgrade Guide* is renamed to *Informatica Big Data Management Hadoop Integration Guide*. Effective in version 10.2, the Data Integration Service can automatically install the Big Data Management binaries to the Hadoop cluster to integrate the domain with the cluster. The integration tasks in the guide do not include installation of the distribution package.

For more information see the *Informatica Big Data Management 10.2 Hadoop Integration Guide*.

Informatica Catalog Administrator Guide

Effective in version 10.2, the *Informatica Live Data Map Administrator Guide* is renamed to *Informatica Catalog Administrator Guide*.

For more information, see the *Informatica Catalog Administrator Guide 10.2*.

Informatica Administrator Reference for Enterprise Information Catalog

Effective in version 10.2, the *Informatica Administrator Reference for Live Data Map* is renamed to *Informatica Administrator Reference for Enterprise Information Catalog*.

For more information, see the *Informatica Administrator Reference for Enterprise Information Catalog 10.2*.

Informatica Enterprise Information Catalog Custom Metadata Integration Guide

Effective in version 10.2, you can ingest custom metadata into the catalog using Enterprise Information Catalog. You can see the new guide *Informatica Enterprise Information Catalog 10.2 Custom Metadata Integration Guide* for more information.

Informatica Enterprise Information Catalog Installation and Configuration Guide

Effective in version 10.2, the *Informatica Live Data Map Installation and Configuration Guide* is renamed to *Informatica Enterprise Information Catalog Installation and Configuration Guide*.

For more information, see the *Informatica Enterprise Information Catalog 10.2 Installation and Configuration Guide*.

Informatica Enterprise Information Catalog REST API Reference

Effective in version 10.2, you can use REST APIs exposed by Enterprise Information Catalog. You can see the new guide *Informatica Enterprise Information Catalog 10.2 REST API Reference* for more information.

Informatica Enterprise Information Catalog Upgrade Guide

Effective in version 10.2, the *Informatica Live Data Map Upgrading from version <x>* is renamed to *Informatica Enterprise Information Catalog Upgrading from versions 10.1, 10.1.1, 10.1.1 HF1, and 10.1.1 Update 2*.

For more information, see the *Informatica Enterprise Information Catalog Upgrading from versions 10.1, 10.1.1, 10.1.1 HF1, and 10.1.1 Update 2 guide.*.

Enterprise Information Catalog

This section describes new Enterprise Information Catalog features in 10.2.

New Data Sources

Effective in version 10.2, Informatica Enterprise Information Catalog allows you to extract metadata from new data sources.

You can create resources in Informatica Catalog Administrator to extract metadata from the following data sources:

Apache Atlas

Metadata framework for Hadoop.

Azure Microsoft SQL Data Warehouse

Cloud-based relational database to process a large volume of data.

Azure Microsoft SQL Server

Managed cloud database.

Azure WASB File Systems

Windows Azure Storage Blobs interface to load data to Azure blobs.

Erwin

Data modeling tool.

Informatica Axon

Enterprise data governance solution.

For more information about new resources, see the *Informatica Catalog Administrator Guide 10.2*.

Custom Scanner Framework

Effective in version 10.2, you can ingest custom metadata into the catalog.

Custom metadata is metadata that you define. You can define a custom model, create a custom resource type, and create a custom resource to ingest custom metadata from a custom data source. You can use custom metadata integration to extract and ingest metadata from custom data sources for which Enterprise Information Catalog does not provide a model.

For more information about custom metadata integration, see the *Informatica Enterprise Information Catalog 10.2 Custom Metadata Integration Guide*.

REST APIs

Effective in version 10.2, you can use Informatica Enterprise Information Catalog REST APIs to access and configure features related to the objects and models associated with a data source.

The REST APIs allow you to retrieve information related to objects and models associated with a data source. In addition, you can create, update, or delete entities related to models and objects such as attributes, associations, and classes.

For more information about unstructured file sources, see the *Informatica Enterprise Information Catalog 10.2 REST API Reference*.

Composite Data Domains

Effective in version 10.2, you can create composite data domains. A composite data domain is a collection of data domains or other composite data domains that you can link using rules. You can use a composite data domain to search for the required details of an entity across multiple schemas in a data source.

You can view composite data domains for tabular assets in the Asset Details view after you create and enable composite data domain discovery for resources in the Catalog Administrator. You can also search for composite data domains and view details of the composite data domains in the Asset Details view.

For more information about composite data domains, see the "View Assets" chapter in the *Informatica Enterprise Information Catalog 10.2 User Guide* and see the "Catalog Administrator Concepts" and "Managing Composite Data Domains" chapters in the *Informatica Catalog Administrator Guide 10.2*.

Data Domains

This section describes new features related to data domains in Enterprise Information Catalog.

Define Data Domains

Effective in version 10.2, you can configure the following additional options when you create a data domain:

- Use reference tables, rules, and regular expressions to create a data rule or column rule.
- Use minimum conformance percentage or minimum conforming rows for data domain match.

- Use the auto-accept option to accept a data domain automatically in Enterprise Information Catalog when the data domain match exceeds the configured auto-accept percentage.

For more information about data domains in Catalog Administrator, see the "Managing Data Domains" chapter in the *Informatica Catalog Administrator Guide 10.2* .

Configure Data Domains

Effective in version 10.2, you can use predefined values or enter a conformance value for data domain match when you create or edit a resource.

For more information about data domains and resources, see the "Managing Resources" chapter in the *Informatica Catalog Administrator Guide 10.2*.

Data Domain Privileges

Effective in version 10.2, configure the **Domain Management: Admin - View Domain** and **Domaingroup and Domain Management: Admin - Edit Domain and Domaingroup** privileges in Informatica Administrator to view, create, edit, or delete data domains or data domain groups in the Catalog Administrator.

For more information about privileges see the "Privileges and Roles" chapter in the *Informatica Administrator Reference for Enterprise Information Catalog 10.2*.

Data Domain Curation

Effective in version 10.2, Enterprise Information Catalog accepts a data domain automatically if the data domain match percentage exceeds the configured auto-accept percentage in Catalog Administrator.

For more information about data domain curation, see the "View Assets" chapter in the *Informatica Enterprise Information Catalog 10.2 User Guide.*

Export and Import of Custom Attributes

Effective in version 10.2, you can export the custom attributes configured in a resource to a CSV file and import the CSV file back into Enterprise Information Catalog. You can use the exported CSV file to assign custom attribute values to multiple assets at the same time.

For more information about export and import of custom attributes, see the "View Assets" chapter in the *Informatica Enterprise Information Catalog 10.2 User Guide.*

Rich Text as Custom Attribute Value

Effective in version 10.2, you can edit a custom attribute to assign multiple rich text strings as the attribute value.

For more information about assigning custom attribute values to an asset, see the "View Assets" chapter in the *Informatica Enterprise Information Catalog 10.2 User Guide.*

Transformation Logic

Effective in version 10.2, you can view transformation logic for assets in the Lineage and Impact view. The Lineage and Impact view displays transformation logic for assets that contain transformations. The transformation view displays transformation logic for data structures, such as tables and columns. The view also displays various types of transformations, such as filter, joiner, lookup, expression, sorter, union, and aggregate.

For more information about transformation logic, see the "View Lineage and Impact" chapter in the *Informatica Enterprise Information Catalog 10.2 User Guide.*

Unstructured File Types

Effective in version 10.2, you can run the **Data Domain Discovery** profile or **Column Profile and Data Domain Discovery** profile on unstructured file types and extended unstructured formats for all the rows in the data source. The unstructured file types include compressed files, email formats, webpage files, Microsoft Excel, Microsoft PowerPoint, Microsoft Word, and PDF. The extended unstructured formats include mp3, mp4, bmp, and jpg.

For more information about unstructured file types, see the "Managing Resources" chapter in the *Informatica Catalog Administrator Guide 10.2*.

Value Frequency

Configure and View Value Frequency

Effective in version 10.2, you can enable value frequency along with column data similarity in the Catalog Administrator to compute the frequency of values in a data source. You can view the value frequency for view column, table column, CSV field, XML file field, and JSON file data assets in the **Asset Details** view after you run the value frequency on a data source in the Catalog Administrator.

For more information about configuring value frequency, see the "Catalog Administrator Concepts" chapter in the *Informatica Catalog Administrator Guide 10.2* . To view value frequency for a data asset, see the "View Assets" chapter in the *Informatica Enterprise Information Catalog 10.2 User Guide*.

Privileges to View Value Frequency in Enterprise Information Catalog

Effective in version 10.2, you need the following permission and privileges to view the value frequency for a data asset:

- Read permission for the data asset.
- **Data Privileges: View Data** privilege.
- **Data Privileges: View Sensitive Data** privilege.

For more information about permissions and privileges, see the "Permissions Overview" and "Privileges and Roles Overview" chapter in the *Informatica Administrator Reference for Enterprise Information Catalog 10.2* .

Deployment Support for Azure HDInsight

Effective in version 10.2, you can deploy Enterprise Information Catalog on Azure HDInsight Hadoop distribution.

For more information, see the "Create the Application Services" chapter in the *Informatica Enterprise Information Catalog 10.2 Installation and Configuration Guide*.

Informatica Analyst

This section describes new Analyst tool features in 10.2.

Profiles

This section describes new features for profiles and scorecards.

Rule Specification

Effective in version 10.2, you can configure a rule specification in the Analyst tool and use the rule specification in the column profile.

For more information about using rule specifications in the column profiles, see the "Rules in Informatica Analyst" chapter in the *Informatica 10.2 Data Discovery Guide*.

Intelligent Data Lake

This section describes new Intelligent Data Lake features in 10.2.

Validate and Assess Data Using Visualization with Apache Zeppelin

Effective in version 10.2, after you publish data, you can validate your data visually to make sure that the data is appropriate for your analysis from content and quality perspectives. You can then choose to fix the recipe thus supporting an iterative Prepare-Publish-Validate process.

Intelligent Data Lake uses Apache Zeppelin to view the worksheets in the form of a visualization Notebook that contains graphs and charts. For more details about Apache Zeppelin, see Apache Zeppelin documentation. When you visualize data using Zeppelin's capabilities, you can view relationships between different columns and create multiple charts and graphs.

When you open the visualization Notebook for the first time after a data asset is published, Intelligent Data Lake uses CLAIRE engine to create Smart Visualization suggestions in the form of histograms of the numeric columns created by the user.

For more information about the visualization notebook, see the "Validate and Assess Data Using Visualization with Apache Zeppelin" chapter in the *Informatica Intelligent Data Lake 10.2 User Guide*.

Assess Data Using Filters During Data Preview

Effective in version 10.2, you can filter the data during data preview for better assessment of data assets. You can add filters for multiple fields and apply combinations of such filters. Filter conditions depend on the data types. If available, you can view column value frequencies found during profiling for string values.

For more information, see the "Discover Data" chapter in the *Informatica Intelligent Data Lake 10.2 User Guide*.

Enhanced Layout of Recipe Panel

Effective in version 10.2, you can see a dedicated panel for Recipe steps during data preparation. The recipe steps are clearer and concise with color codes to indicate function name, columns involved, and input sources. You can edit the steps or delete them. You can also go back-in-time to a specific step in the recipe and see the state of data. You can refresh the recipe from the source. You can also see a separate Ingredients panel which shows the sources used for this sheet.

For more information, see the "Prepare Data" chapter in the *Informatica Intelligent Data Lake 10.2 User Guide*.

Apply Data Quality Rules

Effective in version 10.2, while preparing data, you can use pre-built rules that are available during interactive data preparation. These rules are created using Informatica Developer or Informatica Analyst tool. If you have a Big Data Quality license, thousands of pre-built rules are available that can be used by Intelligent Data Lake users as well. Using pre-built rules promotes effective collaboration within Business and IT with reusability of rules and knowledge, consistency of usage and extensibility.

For more information, see the "Prepare Data" chapter in the *Informatica Intelligent Data Lake 10.2 User Guide*.

View Business Terms for Data Assets in Data Preview and Worksheet View

Effective in version 10.2, you can view business terms associated with columns of data assets in data preview as well as during data preparation.

For more information, see the "Discover Data" chapter in the *Informatica Intelligent Data Lake 10.2 User Guide*.

Prepare Data for Delimited Files

Effective in version 10.2, as a data analyst, you can cleanse, transform, combine, aggregate, and perform other operations on delimited HDFS files that are already in the lake. You can preview these files before adding them to a project. You can then configure the sampling settings of these assets and perform data preparation operations on them.

For more information, see the "Prepare Data" chapter in the *Informatica Intelligent Data Lake 10.2 User Guide*.

Edit Joins in a Joined Worksheet

Effective in version 10.2, you can edit the joinconditions for an existing joined worksheet such as join keys, join types (such as inner and outer joins).

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

Edit Sampling Settings for Data Preparation

Effective in version 10.2, you can edit the sampling settings while preparing your data asset. You can change the columns selected for sampling, edit the filters selected, and change the sampling criteria.

For more information, see the "Prepare Data" chapter in the *Informatica Intelligent Data Lake 10.2 User Guide*.

Support for Multiple Enterprise Information Catalog Resources in the Data Lake

Effective in version 10.2, you can configure multiple Enterprise Information Catalog resources so that the users can work with all types of assets and all applicable Hive schemas in the lake.

Use Oracle for the Data Preparation Service Repository

Effective in version 10.2, you can now use Oracle 11gR2 and 12c for the Data Preparation Service repository.

Improved Scalability for the Data Preparation Service

Effective in version 10.2, you can ensure horizontal scalability by using grid for the Data Preparation Service with multiple Data Preparation Service nodes. Improved scalability supports high performance, interactive data preparation during increased data volumes and increased number of users.

Informatica Developer

This section describes new Developer tool features in 10.2.

Nonrelational Data Objects

Effective in version 10.2, you can import multiple nonrelational data objects at a time.

For more information, see the "Physical Data Objects" chapter in the *Informatica 10.2 Developer Tool Guide*.

Profiles

This section describes new features for profiles and scorecards.

Rule Specification

Effective in version 10.2, you can use rule specifications when you create a column profile in the Developer tool. To use the rule specification, generate a mapplet from the rule specification and validate the mapplet as a rule.

For more information about using rule specifications in the column profiles, see the "Rules in Informatica Developer" chapter in the *Informatica 10.2 Data Discovery Guide*.

Informatica Installation

This section describes new installation features in 10.2.

Informatica Upgrade Advisor

Effective in version 10.2, you can run the Informatica Upgrade Advisor to validate the services and check for obsolete services, supported databases, and supported operating systems in the domain before you perform an upgrade.

For more information about the upgrade advisor, see the *Informatica Upgrade Guides*.

Intelligent Streaming

This section describes new Intelligent Streaming features in 10.2.

CSV Format

Effective in version 10.2, Streaming mappings can read and write data in CSV format.

For more information about the CSV format, see the "Sources and Targets in a Streaming Mapping" chapter in the *Informatica Intelligent Streaming 10.2 User Guide*.

Data Types

Effective in version 10.2, Streaming mappings can read, process, and write hierarchical data. You can use array, struct, and map complex data types to process the hierarchical data.

For more information, see the "Sources and Targets in a Streaming Mapping" chapter in the *Informatica Intelligent Streaming 10.2 User Guide*.

Connections

Effective in version 10.2, you can use the following new messaging connections in Streaming mappings:

- AmazonKinesis. Access Amazon Kinesis Stream as source or Amazon Kinesis Firehose as target. You can create and manage an AmazonKinesis connection in the Developer tool or through infacmd.
- MapRStreams. Access MapRStreams as targets. You can create and manage a MapRStreams connection in the Developer tool or through infacmd.

For more information, see the "Connections" chapter in the *Informatica Intelligent Streaming 10.2 User Guide*.

Pass-Through Mappings

Effective in version 10.2, you can pass any payload format directly from source to target in Streaming mappings.

You can project columns in binary format to pass a payload from source to target in its original form or to pass a payload format that is not supported.

For more information, see the "Sources and Targets in a Streaming Mapping" chapter in the *Informatica Intelligent Streaming 10.2 User Guide*.

Sources and Targets

Effective in version 10.2, you can create the following new physical data objects:

- AmazonKinesis. Represents data in a Amazon Kinesis Stream or Amazon Kinesis Firehose Delivery Stream.
- MapRStreams. Represents data in a MapR Stream.

For more information, see the "Sources and Targets in a Streaming Mapping" chapter in the *Informatica Intelligent Streaming 10.2 User Guide*.

Transformation Support

Effective in version 10.2, you can use the Rank transformation with restrictions in Streaming mappings.

For more information, see the "Intelligent Streaming Mappings" chapter in the *Informatica Intelligent Streaming 10.2 User Guide*.

Metadata Manager

This section describes new Metadata Manager features in 10.2.

Cloudera Navigator

Effective in version 10.2, you can provide the truststore file information to enable a secure connection to a Cloudera Navigator resource. When you create or edit a Cloudera Navigator resource, enter the path and file name of the truststore file for the Cloudera Navigator SSL instance and the password of the truststore file.

For more information about creating a Cloudera Navigator Resource, see the "Database Management Resources" chapter in the *Informatica Metadata Manager 10.2 Administrator Guide*.

PowerCenter

This section describes new PowerCenter features in 10.2.

Audit Logs

Effective in version 10.2, you can generate audit logs when you import an .xml file into the PowerCenter repository. When you import one or more repository objects, you can generate audit logs. You can enable Security Audit Trail configuration option in the PowerCenter Repository Service properties in the Administrator tool to generate audit logs when you import an .xml file into the PowerCenter repository. The user activity logs captures all the audit messages.

The audit logs contain the following information about the file, such as the file name and size, the number of objects imported, and the time of the import operation.

For more information, see the "pmrep Command Reference" chapter in the *Informatica 10.2 Command Reference*, the *Informatica 10.2 Application Service Guide*, and the *Informatica 10.2 Administrator Guide*.

Bulk Upsert for SAP HANA Targets

Effective in version 10.2, when you upsert data into SAP HANA targets, you can configure the EnableArrayUpsert custom property to upsert data in bulk and improve the session performance. You can configure the EnableArrayUpsert custom property at the session level or at the PowerCenter Integration Service level, and set its value to yes.

For more information, see the "Working with Targets" chapter in the *Informatica 10.2 PowerCenter Designer Guide*.

Object Queries

Effective in version 10.2, you can create and delete object queries with the *pmrep* commands.

For more information, see the "pmrep Command Reference" chapter in the *Informatica 10.2 Command Reference*.

Use Parameter in a Password

Effective in version 10.2, you can create or update a connection with a parameter in password with the *pmrep* commands.

You can also update a connection with or without a parameter in password with the *pmrep* command.

For more information, see the "pmrep Command Reference" chapter in the *Informatica 10.2 Command Reference*.

PowerExchange Adapters

This section describes new PowerExchange adapter features in 10.2.

PowerExchange Adapters for Informatica

This section describes new Informatica adapter features in 10.2.

PowerExchange for Amazon Redshift

Effective in version 10.2, PowerExchange for Amazon Redshift includes the following new features:

- You can read data from or write data to the Amazon S3 buckets in the following regions:
  - Asia Pacific (Mumbai)
  - Asia Pacific (Seoul)
  - Canada (Central)
  - China(Beijing)
  - EU (London)
  - US East (Ohio)
- You can run Amazon Redshift mappings on the Spark engine. When you run the mapping, the Data Integration Service pushes the mapping to a Hadoop cluster and processes the mapping on the Spark engine, which significantly increases the performance.
- You can use AWS Identity and Access Management (IAM) authentication to securely control access to Amazon S3 resources.
- You can connect to Amazon Redshift Clusters available in Virtual Private Cloud (VPC) through VPC endpoints.
- You can use AWS Identity and Access Management (IAM) authentication to run a session on the EMR cluster.

For more information, see the *Informatica PowerExchange for Amazon Redshift 10.2 User Guide*.

PowerExchange for Amazon S3

Effective in version 10.2, PowerExchange for Amazon S3 includes the following new features:

- You can read data from or write data to the Amazon S3 buckets in the following regions:
  - Asia Pacific (Mumbai)
  - Asia Pacific (Seoul)
  - Canada (Central)
  - China (Beijing)
  - EU (London)
  - US East (Ohio)

- You can compress data in the following formats when you read data from or write data to Amazon S3 in the native environment and Spark engine:

Compression format	Read	Write
Bzip2	Yes	Yes
Deflate	No	Yes
Gzip	Yes	Yes
Lzo	Yes	Yes
None	Yes	Yes
Snappy	No	Yes

- You can select the type of source from which you want to read data in the **Source Type** option under the advanced properties for an Amazon S3 data object read operation. You can select **Directory** or **File** source types.
- You can select the type of the data sources in the **Resource Format** option under the Amazon S3 data objects properties. You can read data from the following source formats:
  - Binary
  - Flat
  - Avro
  - Parquet
- You can connect to Amazon S3 buckets available in Virtual Private Cloud (VPC) through VPC endpoints.
- You can run Amazon S3 mappings on the Spark engine. When you run the mapping, the Data Integration Service pushes the mapping to a Hadoop cluster and processes the mapping on the Spark engine.
- You can choose to overwrite the existing files. You can select the **Overwrite File(s) If Exists** option in the Amazon S3 data object write operation properties to overwrite the existing files.
- You can use AWS Identity and Access Management (IAM) authentication to securely control access to Amazon S3 resources.
- You can filter the metadata to optimize the search performance in the **Object Explorer** view.
- You can use AWS Identity and Access Management (IAM) authentication to run a session on the EMR cluster.

For more information, see the *Informatica PowerExchange for Amazon S3 10.2 User Guide*.

PowerExchange for HBase

Effective in version 10.2, PowerExchange for HBase contains the following new features:

- You can use PowerExchange for HBase to read from sources and write to targets stored in the WASB file system on Azure HDInsight.
- You can associate a cluster configuration with an HBase connection. A cluster configuration is an object in the domain that contains configuration information about the Hadoop cluster. The cluster configuration enables the Data Integration Service to push mapping logic to the Hadoop environment.

For more information, see the *Informatica PowerExchange for HBase 10.2 User Guide*.

PowerExchange for HDFS

Effective in version 10.2, you can associate a cluster configuration with an HDFS connection. A cluster configuration is an object in the domain that contains configuration information about the Hadoop cluster. The cluster configuration enables the Data Integration Service to push mapping logic to the Hadoop environment.

For more information, see the *Informatica PowerExchange for HDFS 10.2 User Guide*.

PowerExchange for Hive

Effective in version 10.2, you can associate a cluster configuration with an Hive connection. A cluster configuration is an object in the domain that contains configuration information about the Hadoop cluster. The cluster configuration enables the Data Integration Service to push mapping logic to the Hadoop environment.

For more information, see the *Informatica PowerExchange for Hive 10.2 User Guide*.

PowerExchange for MapR-DB

Effective in version 10.2, PowerExchange for MapR-DB contains the following new features:

- You can run MapR-DB mappings on the Spark engine. When you run the mapping, the Data Integration Service pushes the mapping to a Hadoop cluster and processes the mapping on the Spark engine, which significantly increases the performance.
- You can configure dynamic partitioning for MapR-DB mappings that you run on the Spark engine.
- You can associate a cluster configuration with an HBase connection for MapR-DB. A cluster configuration is an object in the domain that contains configuration information about the Hadoop cluster. The cluster configuration enables the Data Integration Service to push mapping logic to the Hadoop environment.

For more information, see the *Informatica PowerExchange for MapR-DB 10.2 User Guide*.

PowerExchange for Microsoft Azure Blob Storage

Effective in version 10.2, you can read data from or write data to a subdirectory in Microsoft Azure Blob Storage. You can use the **Blob Container Override** and **Blob Name Override** fields to read data from or write data to a subdirectory in Microsoft Azure Blob Storage.

For more information, see the *Informatica PowerExchange for Microsoft Azure Blob Storage 10.2 User Guide*.

PowerExchange for Microsoft Azure SQL Data Warehouse

Effective in version 10.2, you can run Microsoft Azure SQL Data Warehouse mappings in a Hadoop environment on Kerberos enabled clusters.

For more information, see the *Informatica PowerExchange for Microsoft Azure SQL Data Warehouse 10.2 User Guide*.

PowerExchange for Salesforce

Effective in version 10.2, you can use version 39 of Salesforce API to create a Salesforce connection and access Salesforce objects.

For more information, see the *Informatica PowerExchange for Salesforce 10.2 User Guide*.

PowerExchange Adapters for PowerCenter

This section describes new PowerCenter adapter features in version 10.2.

PowerExchange for Amazon Redshift

Effective in version 10.2, PowerExchange for Amazon Redshift includes the following new features:

- You can read data from or write data to the China (Beijing) region.

- When you import objects from **AmazonRSCloudAdapter** in the PowerCenter Designer, the PowerCenter Integration Service lists the table names alphabetically.

- In addition to the existing recovery options in the vacuum table, you can select the **Reindex** option to analyze the distribution of the values in an interleaved sort key column.

- You can configure the multipart upload option to upload a single object as a set of independent parts. TransferManager API uploads the multiple parts of a single object to Amazon S3. After uploading, Amazon S3 assembles the parts and creates the whole object. TransferManager API uses the multipart uploads option to achieve performance and increase throughput when the content size of the data is large and the bandwidth is high.
  You can configure the **Part Size** and **TransferManager Thread Pool Size** options in the target session properties.

- PowerExchange for Amazon Redshift uses the `commons-beanutils.jar` file to address potential security issues when accessing properties. The following is the location of the `commons-beanutils.jar` file:
  `<Informatica installation directory>server/bin/javalib/505100/commons-beanutils-1.9.3.jar`

For more information, see the *Informatica PowerExchange for Amazon Redshift 10.2 User Guide for PowerCenter*.

PowerExchange for Amazon S3

Effective in version 10.2, PowerExchange for Amazon S3 includes the following new features:

- You can read data from or write data to the China (Beijing) region.

- You can read multiple files from Amazon S3 and write data to a target.

- You can write multiple files to Amazon S3 target from a single source. You can configure the **Distribution Column** options in the target session properties.

- When you create a mapping task to write data to Amazon S3 targets, you can configure partitions to improve performance. You can configure the **Merge Partition Files** option in the target session properties.

- You can specify a directory path that is available on the PowerCenter Integration Service in the **Staging File Location** property.

- You can configure the multipart upload option to upload a single object as a set of independent parts. TransferManager API uploads the multiple parts of a single object to Amazon S3. After uploading, Amazon S3 assembles the parts and creates the whole object. TransferManager API uses the multipart uploads option to achieve performance and increase throughput when the content size of the data is large and the bandwidth is high.
  You can configure the **Part Size** and **TransferManager Thread Pool Size** options in the target session properties.

For more information, see the *Informatica PowerExchange for Amazon S3 version 10.2 User Guide for PowerCenter*.

PowerExchange for Microsoft Dynamics CRM

Effective in version 10.2, you can use the following target session properties with PowerExchange for Microsoft Dynamics CRM:

- Add row reject reason. Select to include the reason for rejection of rows to the reject file.

- Alternate Key Name. Indicates whether the column is an alternate key for an entity. Specify the name of the alternate key. You can use alternate key in update and upsert operations.
- You can configure PowerExchange for Microsoft Dynamics CRM to run on AIX platform.

For more information, see the *Informatica PowerExchange for Microsoft Dynamics CRM 10.2 User Guide for PowerCenter*.

PowerExchange for SAP NetWeaver

Effective in version 10.2, PowerExchange for SAP NetWeaver includes the following new features:

- When you run ABAP mappings to read data from SAP tables, you can use the STRING, SSTRING, and RAWSTRING data types. The SSTRING data type is represented as SSTR in PowerCenter.
- When you read or write data through IDocs, you can use the SSTRING data type.
- When you run ABAP mappings to read data from SAP tables, you can configure HTTP streaming.

For more information, see the *Informatica PowerExchange for SAP NetWeaver 10.2 User Guide for PowerCenter*.

Rule Specifications

Effective in version 10.2, you can select a rule specification from the Model repository in Informatica Developer and add the rule specification to a mapping. You can also deploy a rule specification as a web service.

A rule specification is a read-only object in the Developer tool. Add a rule specification to a mapping in the same way that you add a mapplet to a mapping. You can continue to select a mapplet that you generated from a rule specification and add the mapplet to a mapping.

Add a rule specification to a mapping when you want the mapping to apply the logic that the current rule specification represents. Add the corresponding mapplet to a mapping when you want to use or update the mapplet logic independently of the rule specification.

When you add a rule specification to a mapping, you can specify the type of outputs on the rule specification. By default, a rule specification has a single output port that contains the final result of the rule specification analysis for each input data row. You can configure the rule specification to create an output port for every rule set in the rule specification.

For more information, see the "Mapplets" chapter in the *Informatica 10.2 Developer Mapping Guide*.

Security

This section describes new security features in 10.2.

User Activity Logs

Effective in version 10.2, you can view login attempts from Informatica client applications in user activity logs.

The user activity data includes the following properties for each login attempt from an Informatica client:

- Application name
- Application version
- Host name or IP address of the application host

If the client set custom properties on login requests, the data includes the custom properties.

For more information, see the "Users and Groups" chapter in the *Informatica 10.2 Security Guide*.

Transformation Language

This section describes new transformation language features in 10.2.

Informatica Transformation Language

This section describes Informatica Transformation Language new features in 10.2.

Complex Functions

Effective in version 10.2, the transformation language introduces complex functions for complex data types. Use complex functions to process hierarchical data on the Spark engine.

The transformation language includes the following complex functions:

- ARRAY
- CAST
- COLLECT_LIST
- CONCAT_ARRAY
- RESPEC
- SIZE
- STRUCT
- STRUCT_AS

For more information about complex functions, see the "Functions" chapter in the *Informatica 10.2 Developer Transformation Language Reference.*

Complex Operators

Effective in version 10.2, the transformation language introduces complex operators for complex data types. In mappings that run on the Spark engine, use complex operators to access elements of hierarchical data.

The transformation language includes the following complex operators:

- Subscript operator [ ]
- Dot operator .

For more information about complex operators, see the "Operators" chapter in the *Informatica 10.2 Developer Transformation Language Reference*.

Window Functions

Effective in version 10.2, the transformation language introduces window functions. Use window functions to process a small subset of a larger set of data on the Spark engine.

The transformation language includes the following window functions:

- LEAD. Provides access to a row at a given physical offset that comes after the current row.
- LAG. Provides access to a row at a given physical offset that comes before the current row.

For more information, see the "Functions" chapter in the *Informatica 10.2 Transformation Language Reference*.

Transformations

This section describes new transformation features in version 10.2.

Informatica Transformations

This section describes new features in Informatica transformations in 10.2.

Address Validator Transformation

This section describes the new Address Validator transformation features.

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

Austria

Effective in version 10.2, you can configure the Address Validator transformation to return a postal address code identifier for a mailbox that has two valid street addresses. For example, a building at an intersection of two streets might have an address on both streets. The building might prefer to receive mail at one of the addresses. The other address remains a valid address, but the postal carrier does not use it to deliver mail.

Austria Post assigns a postal address code to both addresses. Austria Post additionally assigns a postal address code identifier to the address that does not receive mail. The postal address code identifier is identical to the postal address code of the preferred address. You can use the postal address code identifier to look up the preferred address with the Address Validator transformation.

To find the postal address code identifier for an address in Austria, select the Postal Address Code Identifier AT output port. Find the port in the AT Supplementary port group.

To find the address that a postal address identifier represents, select the Postal Address Code Identifier AT input port. Find the port in the Discrete port group.

Czech Republic

Effective in version 10.2, you can configure the Address Validator transformation to add RUIAN ID values to a valid Czech Republic address.

You can find the following RUIAN ID values:

- RUIANAM_ID. Uniquely identifies the address delivery point.
  To find the RUIAN ID value that uniquely identifies the address delivery point, select the RUIAN Delivery Point Identifier output port.

- RUIANSO_ID. Identifies the address to building level.
  To find the RUIAN ID value that identifies the address to building level, select the RUIAN Building Identifier output port.

- RUIANTEA_ID. Identifies the building entrance.
  To find the RUIAN ID value that identifies the entrance to building, select the RUIAN Building Entrance Identifier output port.

Find the ports in the CZ Supplementary port group.

Hong Kong

The Address Validator transformation includes the following features for Hong Kong:

Multilanguage support for Hong Kong addresses

Effective in version 10.2, the Address Validator transformation can read and write Hong Kong addresses in Chinese or in English.

Use the Preferred Language property to select the preferred language for the addresses that the transformation returns. The default language is Chinese. To return Hong Kong addresses in English, update the property to ENGLISH.

Use the Preferred Script property to select the preferred character set for the address data. The default character set is Hanzi. To return Hong Kong addresses in Latin characters, update the property to a Latin or ASCII option. When you select a Latin script, address validation transliterates the address data into Pinyin.

Single-line address validation in suggestion list mode

Effective in version 10.2, you can configure the Address Validator transformation to return valid suggestions for a Hong Kong address that you enter on a single line. To return the suggestions, configure the transformation to run in suggestion list mode.

Submit the address in the native Chinese language and in the Hanzi script. The Address Validator transformation reads the address in the Hanzi script and returns the address suggestions in the Hanzi script.

Submit a Hong Kong address in the following format:

```
[Province] [Locality] [Street] [House Number] [Building 1] [Building 2] [Sub-
building]
```

When you submit a partial address, the transformation returns one or more address suggestions for the address that you enter. When you enter a complete or almost complete address, the transformation returns a single suggestion for the address that you enter.

To verify single-line addresses, use the Complete Address port.

Macau

The Address Validator transformation includes the following features for Macau:

Multilanguage support for Macau addresses

Effective in version 10.2, the Address Validator transformation can read and write Macau addresses in Chinese or in Portuguese.

Use the Preferred Language property to select the preferred language for the addresses that the transformation returns. The default language is Chinese. To return Macau addresses in Portuguese, update the property to ALTERNATIVE_2.

Use the Preferred Script property to select the preferred character set for the address data. The default character set is Hanzi. To return Macau addresses in Latin characters, update the property to a Latin or ASCII option.

Note: When you select a Latin script with the default preferred language option, address validation transliterates the Chinese address data into Cantonese or Mandarin. When you select a Latin script with the ALTERNATIVE_2 preferred language option, address validation returns the address in Portuguese.

Single-line address verification for native Macau addresses in suggestion list mode

Effective in version 10.2, you can configure the Address Validator transformation to return valid suggestions for a Macau address that you enter on a single line in suggestion list mode. When you enter a partial address in suggestion list mode, the transformation returns one or more address suggestions for the address that you enter. Submit the address in the Chinese language and in the Hanzi script. The transformation returns address suggestions in the Chinese language and in the Hanzi script. Enter a Macau address in the following format:

```
[Locality] [Street] [House Number] [Building]
```

Use the Preferred Language property to select the preferred language for the addresses. The default preferred language is Chinese. Use the Preferred Script property to select the preferred character set for the address data. The default preferred script is Hanzi. To verify single-line addresses, enter the addresses in the Complete Address port.

Taiwan

Effective in version 10.2, you can configure the Address Validator transformation to return a Taiwan address in the Chinese language or the English language.

Use the Preferred Language property to select the preferred language for the addresses that the transformation returns. The default language is traditional Chinese. To return Taiwan addresses in English, update the property to ENGLISH.

Use the Preferred Script property to select the preferred character set for the address data. The default character set is Hanzi. To return Taiwan addresses in Latin characters, update the property to a Latin or ASCII option.

Note: The Taiwan address structure in the native script lists all address elements in a single line. You can submit the address as a single string in a Formatted Address Line port.

When you format an input address, enter the elements in the address in the following order:

```
Postal Code, Locality, Dependent Locality, Street, Dependent Street, House or Building
Number, Building Name, Sub-Building Name
```

United States

The Address Validator transformation includes the following features for the United States:

Support for the Secure Hash Algorithm-compliant versions of CASS data files

Effective in version 10.2, the Address Validator transformation reads CASS certification data files that comply with the SHA-256 standard.

The current CASS certification files are numbered USA5C101.MD through USA5C126.MD. To verify United States addresses in certified mode, you must use the current files.

Note: The SHA-256-compliant files are not compatible with older versions of Informatica.

Support for Door Not Accessible addresses in certified mode

Effective in version 10.2, you can configure the Address Validator transformation to identify United States addresses that do not provide a door or entry point for a mail carrier. The mail carrier might be unable to deliver a large item to the address.

The United States Postal Service maintains a list of addresses for which a mailbox is accessible but for which a physical entrance is inaccessible. For example, a residence might locate a mailbox outside a locked gate or on a rural route. The address reference data includes the list of inaccessible addresses that the USPS recognizes. Address validation can return the accessible status of an address when you verify the address in certified mode.

To identify DNA addresses, select the Delivery Point Validation Door not Accessible port. Find the port in the US Specific port group.

Support for No Secure Location address in certified mode

Effective in version 10.2, you can configure the Address Validator transformation to identify United States addresses that do not provide a secure mailbox or reception point for mail. The mail carrier might be unable to deliver a large item to the address.

The United States Postal Service maintains a list of addresses at which the mailbox is not secure. For example, a retail store is not a secure location if the mail carrier can enter the store but cannot find a mailbox or an employee to receive the mail. The address reference data includes the list of non-secure addresses that the USPS recognizes. Address validation can return the non-secure status of an address when you verify the address in certified mode.

To identify DNA addresses, select the Delivery Point Validation No Secure Location port. Find the port in the US Specific port group.

Support for Post Office Box Only Delivery Zones

Effective in version 10.2, you can configure the Address Validator transformation to identify ZIP Codes that contain post office box addresses and no other addresses. When all of the addresses in a ZIP Code are post office box addresses, the ZIP Code represents a Post Office Box Only Delivery Zone.

The Address Validator transformation adds the value Y to an address to indicate that it contains a ZIP Code in a Post Office Box Only Delivery Zone. The value enables the postal carrier to sort mail more easily. For example, the mailboxes in a Post Office Box Only Delivery Zone might reside in a single post office building. The postal carrier can deliver all mail to the Post Office Box Only Delivery Zone in a single trip.

To identify Post Office Box Only Delivery Zones, select the Post Office Box Delivery Zone Indicator port. Find the port in the US Specific port group.

For more information, see the *Informatica 10.2 Developer Transformation Guide* and the *Informatica 10.2 Address Validator Port Reference*.

Data Processor Transformation

This section describes new Data Processor transformation features.

JsonStreamer

Use the JsonStreamer object in a Data Processor transformation to process large JSON files. The transformation splits very large JSON files into complete JSON messages. The transformation can then call other Data Processor transformation components, or a Hierarchical to Relational transformation, to complete the processing.

For more information, see the "Streamers" chapter in the *Informatica Data Transformation 10.2 User Guide*.

RunPCWebService

Use the RunPCWebService action to call a PowerCenter mapplet from within a Data Processor transformation.

For more information, see the "Actions" chapter in the *Informatica Data Transformation 10.2 User Guide*.

PowerCenter Transformations

Evaluate Expression

Effective in version 10.2, you can evaluate expressions that you configure in the Expression Editor of an Expression transformation. When you test an expression, you can enter sample data and then evaluate the expression.

For more information about evaluating an expression, see the "Working with Transformations" chapter and the "Expression Transformation" chapter in the *Informatica PowerCenter 10.2 Transformation Guide*.

Workflows

This section describes new workflow features in version 10.2.

Informatica Workflows

This section describes new features in Informatica workflows in 10.2.

Human Task Distribution Properties

Effective in version 10.2, you can store a list of the users or groups who can work on Human task instances in an external database table. You select the table when you configure the Human task to define task instances based on the values in a column of source data.

The table identifies the users or groups who can work on the task instances and specifies the column values to associate with each user or group. You can update the table independently of the workflow configuration, for example as users join or leave the project. When the workflow runs, the Data Integration Service uses the current information in the table to assign task instances to users or groups.

You can also specify a range of numeric values or date values when you associate users or groups with the values in a source data column. When one or more records contain a value in a range that you specify, the Data Integration Service assigns the task instance to a user or group that you specify.

For more information, see the "Human Task" chapter in the *Informatica 10.2 Developer Workflow Guide*.

Human Task Notification Properties

Effective in version 10.2, you can edit the subject line of an email notification that you configure in a Human task. You can also add a workflow variable to the subject line of the notification.

A Human task can send email notifications when the Human task completes in the workflow and when a task instance that the Human task defines changes status. To configure notifications for a Human task, update the Notifications properties on the Human task in the workflow. To configure notifications for a task

instance, update the Notification properties on the step within the Human task that defines the task instances.

When you configure notifications for a Human task instance, you can select an option to notify the task instance owner in addition to any recipient that you specify. The option applies when a single user owns the task instance. When you select the option to notify the task instance owner, you can optionally leave the Recipients field empty

For more information, see the "Human Task" chapter in the *Informatica 10.2 Developer Workflow Guide*.

Import from PowerCenter

Effective in version 10.2, you can import mappings with multiple pipelines, sessions, workflows, and worklets from PowerCenter into the Model repository. Sessions within a workflow are imported as Mapping tasks in the Model repository. Workflows are imported as workflows within the Model repository. Worklets within a workflow are expanded and objects are imported into the Model repository.

Multiple pipelines within a mapping are imported as separate mappings into the Model repository based on the target load order. If a workflow contains a session that runs a mapping with multiple pipelines, the import process creates a separate Model repository mapping and mapping task for each pipeline in the PowerCenter mapping to preserve the target load order.

For more information about importing from PowerCenter, see the "Import from PowerCenter" chapter in the *Informatica 10.2 Developer Mapping Guide* and the "Workflows" chapter in the *Informatica 10.2 Developer Workflow Guide*.

CHAPTER 6

Changes (10.2)

This chapter includes the following topics:

Support Changes

This section describes the support changes in 10.2.

Big Data Hadoop Distribution Support

Informatica big data products support a variety of Hadoop distributions. In each release, Informatica adds, defers, and drops support for Hadoop distribution versions. Informatica might reinstate support for deferred versions in a future release.

The following table lists the supported Hadoop distribution versions for Informatica 10.2 big data products:

Product	Amazon EMR	Azure HDInsight	Cloudera CDH	Hortonworks HDP	IBM BigInsights	MapR
Big Data Management	5.4, 5.8	3.5, 3.6	5.9, 5.10, 5.11, 5.12, 5.13	2.4, 2.5, 2.6	4.2	5.2 MEP 2.0 5.2 MEP 3.0
Informatica Intelligent Streaming	5.8	NA	5.11, 5.12, 5.13	2.6	NA	5.2 MEP 2.0
Enterprise Information Catalog	NA	3.6	5.8, 5.9, 5.10, 5.11	2.5, 2.6	4.2.x	3.1
Intelligent Data Lake	5.4	3.6	5.11, 5.12	2.6	4.2	5.2 MEP 2.0

To see a list of the latest supported versions, see the Product Availability Matrix on the Informatica Customer Portal: https://network.informatica.com/community/informatica-network/product-availability-matrices.

Big Data Management Hadoop Distributions

The following table lists the supported Hadoop distribution versions and changes in Big Data Management 10.2:

Hadoop Distribution	Supported Distribution Versions	10.2 Changes
Amazon EMR	5.4 5.8	Dropped support for version 5.0. Added support for version 5.8. **Note:** To use Amazon EMR 5.8 with Big Data Management 10.2, you must apply Emergency Bug Fix 10571. See Knowledge Base article KB 525399.
Azure HDInsight	3.5.x 3.6.x	Added support for version 3.6.
Cloudera CDH	5.9.x 5.10.x 5.11.x 5.12.x 5.13.x	Added support for versions 5.12, 5.13 Dropped support for version 5.8.

Hadoop Distribution	Supported Distribution Versions	10.2 Changes
Hortonworks HDP	2.4x 2.5x 2.6x	Dropped support for version 2.3. **Note:** To use Hortonworks 2.4 or 2.5 with Big Data Management 10.2, you must apply Emergency Bug Fix patches. See the following Knowledge Base articles: - Hotwonworks 2.4 support: KB 521845. - Hortonworks 2.5 support: KB 521847.
IBM BigInsights	4.2.x	No change.
MapR	5.2 MEP 2.0.x 5.2 MEP 3.0.x	Added support for versions 5.2 MEP 2.0 and 5.2 MEP 3.0. Dropped support for version 5.2 MEP 1.0.

Informatica big data products support a variety of Hadoop distributions. In each release, Informatica adds, defers, and drops support for Hadoop distribution versions. Informatica might reinstate support for deferred versions in a future release.

To see a list of the latest supported versions, see the Product Availability Matrix on the Informatica network: `https://network.informatica.com/community/informatica-network/product-availability-matrices`.

Enterprise Information Catalog Hadoop Distributions

The following table lists the supported Hadoop distribution versions and changes in Enterprise Information Catalog 10.2:

Hadoop Distribution	Supported Distribution Versions	Changes since 10.1.1 HotFix1
Azure HDInsight	3.6	Added support for Azure HDInsight.
Cloudera CDH	5.8, 5.9, 5.10, 5.11	No changes.
Hortonworks HDP	2.5.x (Kerberos version), 2.6.x (Non Kerberos version)	Added support for 2.6 non-Kerberos version.
IBM BigInsights	4.2	No change.

Intelligent Data Lake Hadoop Distributions

The following table lists the supported Hadoop distribution versions and changes in Intelligent Data Lake 10.2:

Hadoop Distribution	Supported Distribution Versions	Changes since 10.1.1 HotFix1
Amazon EMR	5.4	Added support for version 5.4. Dropped support for version 5.0.
Azure HDInsight	3.6	Added support for version 3.6. Dropped support for version 3.5.

Hadoop Distribution	Supported Distribution Versions	Changes since 10.1.1 HotFix1
Cloudera CDH	5.10 5.11 5.12	Added support for version 5.10 and 5.12. Dropped support for version 5.8. Deferred support for version 5.9.
Hortonworks HDP	2.6	Dropped support for version 2.3. Deferred support for versions 2.4 and 2.5.
IBM BigInsights	4.2	No change.
MapR	5.2 MEP 2.0	Added support for MapR.

Intelligent Streaming Hadoop Distributions

The following table lists the supported Hadoop distribution versions and changes in Intelligent Streaming 10.2:

Distribution	Supported Versions	Changes Since 10.1.1 HotFix1
Amazon EMR	5.4 5.8	Added support for 5..8.
Cloudera CDH	5.10.x 5.11.x 5.12.x 5.13.x	Added support for 5.13. Dropped support for versions 5.8. Deferred support for versions 5.9.
Hortonworks HDP	2.5.x 2.6.x	Dropped support for versions 2.3. Deferred support for versions 2.4.
MapR	5.2 MEP 2.0	Added support for version 5.2 MEP 2.0.

To see a list of the latest supported versions, see the Product Availability Matrix on the Informatica network: `https://network.informatica.com/community/informatica-network/product-availability-matrices`.

Metadata Manager

Custom Metadata Configurator (Deprecated)

Effective in version 10.2, Informatica deprecated the Custom Metadata Configurator in Metadata Manager.

You can use the load template to load metadata from metadata source files into a custom resource. Create a load template for the models that use Custom Metadata Configurator templates.

For more information about using load templates, see the "Custom XConnect Created with a Load Template" in the *Informatica Metadata Manager 10.2 Custom Metadata Integration Guide*.

Application Services

This section describes changes to Application Services in 10.2.

Content Management Service

Effective in version 10.2, you do not need to update the search index on the Model repository before you run the infacmd cms purge command. The infacmd cms purge command updates the search index before it purges unused tables from the reference data warehouse.

Previously, you updated the search index before you ran the command so that the Model repository held an up-to-date list of reference tables. The Content Management Service used the list of objects in the index to select the tables to delete.

For more information, see the "Content Management Service" chapter in the *Informatica 10.2 Application Service Guide*.

Data Integration Service

This section describes changes to the Data Integration Service in 10.2.

Execution Options

Effective in version 10.2, you configure the following execution options on the Properties view for the Data Integration Service:

- Maximum On-Demand Execution Pool Size. Controls the number of on-demand jobs that can run concurrently. Jobs include data previews, profiling jobs, REST and SQL queries, web service requests, and mappings run from the Developer tool.

- Maximum Native Batch Execution Pool Size. Controls the number of deployed native jobs that each Data Integration Service process can run concurrently.

- Maximum Hadoop Batch Execution Pool Size. Controls the number of deployed Hadoop jobs that can run concurrently.

Previously, you configured the **Maximum Execution Pool Size** property to control the maximum number of jobs the Data Integration Service process could run concurrently.

When you upgrade to 10.2, the value of the maximum execution pool size upgrades to the following properties:

- Maximum On-Demand Batch Execution Pool Size. Inherits the value of the Maximum Execution Pool Size property.

- Maximum Native Batch Execution Pool Size. Inherits the value of the Maximum Execution Pool Size property.

- Maximum Hadoop Batch Execution Pool Size. Inherits the value of the Maximum Execution Pool size property if the original value has been changed from 10. If the value is 10, the Hadoop batch pool retains the default size of 100.

For more information, see the "Data Integration Service" chapter in the *Informatica 10.2 Application Service Guide*.

Big Data

This section describes the changes to big data in 10.2.

Hadoop Connection

Effective in version 10.2, the following changes affect Hadoop connection properties.

You can use the following properties to configure your Hadoop connection:

Property	Description
Cluster Configuration	The name of the cluster configuration associated with the Hadoop environment. Appears in General Properties.
Write Reject Files to Hadoop	Select the property to move the reject files to the HDFS location listed in the property Reject File Directory when you run mappings. Appears in Reject Directory Properties.
Reject File Directory	The directory for Hadoop mapping files on HDFS when you run mappings. Appears in Reject Directory Properties
Blaze Job Monitor Address	The host name and port number for the Blaze Job Monitor. Appears in Blaze Configuration.
YARN Queue Name	The YARN scheduler queue name used by the Spark engine that specifies available resources on a cluster. Appears in Blaze Configuration.

Effective in version 10.2, the following properties are renamed:

Current Name	Previous Name	Description
ImpersonationUserName	HiveUserName	Hadoop impersonation user. The user name that the Data Integration Service impersonates to run mappings in the Hadoop environment.
Hive Staging Database Name	Database Name	Namespace for Hive staging tables. Appears in Common Properties. Previously appeared in Hive Properties.
HiveWarehouseDirectory	HiveWarehouseDirectoryOnHDFS	The absolute HDFS file path of the default database for the warehouse that is local to the cluster.
Blaze Staging Directory	Temporary Working Directory on HDFS CadiWorkingDirectory	The HDFS file path of the directory that the Blaze engine uses to store temporary files. Appears in Blaze Configuration.

Current Name	Previous Name	Description
Blaze User Name	Blaze Service User Name CadiUserName	The owner of the Blaze service and Blaze service logs. Appears in Blaze Configuration.
YARN Queue Name	Yarn Queue Name CadiAppYarnQueueName	The YARN scheduler queue name used by the Blaze engine that specifies available resources on a cluster. Appears in Blaze Configuration.
BlazeMaxPort	CadiMaxPort	The maximum value for the port number range for the Blaze engine.
BlazeMinPort	CadiMinPort	The minimum value for the port number range for the Blaze engine.
BlazeExecutionParameterList	CadiExecutionParameterList	An optional list of configuration parameters to apply to the Blaze engine.
SparkYarnQueueName	YarnQueueName	The YARN scheduler queue name used by the Spark engine that specifies available resources on a cluster.
Spark Staging Directory	Spark HDFS Staging Directory	The HDFS file path of the directory that the Spark engine uses to store temporary files for running jobs.

Effective in version 10.2, the following properties are removed from the connection and imported into the cluster configuration:

Property	Description
Resource Manager Address	The service within Hadoop that submits requests for resources or spawns YARN applications. Imported into the cluster configuration as the property `yarn.resourcemanager.address`. Previously appeared in Hadoop Cluster Properties.
Default File System URI	The URI to access the default Hadoop Distributed File System. Imported into the cluster configuration as the property `fs.defaultFS` or `fs.default.name`. Previously appeared in Hadoop Cluster Properties.

Effective in version 10.2, the following properties are deprecated and are removed from the connection:

Property	Description
Type	The connection type. Previously appeared in General Properties.
Metastore Execution Mode*	Controls whether to connect to a remote metastore or a local metastore. Previously appeared in Hive Configuration.
Metastore Database URI*	The JDBC connection URI used to access the data store in a local metastore setup. Previously appeared in Hive Configuration.
Metastore Database Driver*	Driver class name for the JDBC data store. Previously appeared in Hive Configuration.
Metastore Database User Name*	The metastore database user name. Previously appeared in Hive Configuration.
Metastore Database Password*	The password for the metastore user name. Previously appeared in Hive Configuration.
Remote Metastore URI*	The metastore URI used to access metadata in a remote metastore setup. This property is imported into the cluster configuration as the property `hive.metastore.uris`. Previously appeared in Hive Configuration.
Job Monitoring URL	The URL for the MapReduce JobHistory server. Previously appeared in Hive Configuration.
These properties are deprecated in 10.2. When you upgrade to 10.2, the property values that you set in a previous release are saved in the repository, but they do not appear in the connection properties.	

HBase Connection Properties

Effective in version 10.2, the following properties are removed from the connection and imported into the cluster configuration:

Property	Description
ZooKeeper Host(s)	Name of the machine that hosts the ZooKeeper server.
ZooKeeper Port	Port number of the machine that hosts the ZooKeeper server.
Enable Kerberos Connection	Enables the Informatica domain to communicate with the HBase master server or region server that uses Kerberos authentication.
HBase Master Principal	Service Principal Name (SPN) of the HBase master server.
HBase Region Server Principal	Service Principal Name (SPN) of the HBase region server.

Hive Connection Properties

Effective in version 10.2, PowerExchange for Hive has the following changes:

- You cannot use a PowerExchange for Hive connection if you want the Hive driver to run mappings in the Hadoop cluster. To use the Hive driver to run mappings in the Hadoop cluster, use a Hadoop connection.

- The following properties are removed from the connection and imported into the cluster configuration:

Property	Description
Default FS URI	The URI to access the default Hadoop Distributed File System.
JobTracker/Yarn Resource Manager URI	The service within Hadoop that submits the MapReduce tasks to specific nodes in the cluster.
Hive Warehouse Directory on HDFS	The absolute HDFS file path of the default database for the warehouse that is local to the cluster.
Metastore Execution Mode	Controls whether to connect to a remote metastore or a local metastore.
Metastore Database URI	The JDBC connection URI used to access the data store in a local metastore setup.
Metastore Database Driver	Driver class name for the JDBC data store.
Metastore Database User Name	The metastore database user name.
Metastore Database Password	The password for the metastore user name.
Remote Metastore URI	The metastore URI used to access metadata in a remote metastore setup. This property is imported into the cluster configuration as the property `hive.metastore.uris`.

HBase Connection Properties for MapR-DB

Effective in version 10.2, the **Enable Kerberos Connection** property is removed from the HBase connection for MapR-DB and imported into the cluster configuration.

Mapping Run-time Properties

This section lists changes to mapping-run time properties.

Execution Environment

Effective in version 10.2, you can configure the Reject File Directory as a new property in the Hadoop Execution Environment.

Name	Value
Reject File Directory	The directory for Hadoop mapping files on HDFS when you run mappings in the Hadoop environment.
	The Blaze engine can write reject files to the Hadoop environment for flat file, HDFS, and Hive targets. The Spark and Hive engines can write reject files to the Hadoop environment for flat file and HDFS targets.
	Choose one of the following options:
	- On the Data Integration Service machine. The Data Integration Service stores the reject files based on the RejectDir system parameter.
	- On the Hadoop Cluster. The reject files are moved to the reject directory configured in the Hadoop connection. If the directory is not configured, the mapping will fail.
	- Defer to the Hadoop Connection. The reject files are moved based on whether the reject directory is enabled in the Hadoop connection properties. If the reject directory is enabled, the reject files are moved to the reject directory configured in the Hadoop connection. Otherwise, the Data Integration Service stores the reject files based on the RejectDir system parameter.

Monitoring

Effective in version 10.2, the AllHiveSourceTables row in the Summary Statistics view in the Administrator tool includes records read from the following sources:

- Original Hive sources in the mapping.
- Staging Hive tables defined by the Hive engine.
- Staging data between two linked MapReduce jobs in each query.

If the LDTM session includes one MapReduce job, the AllHiveSourceTables statistic only includes original Hive sources in the mapping.

For more information, see the "Monitoring Mappings in the Hadoop Environment" chapter of the *Big Data Management 10.2 User Guide*.

S3 Access and Secret Key Properties

Effective in version 10.2, the following properties are included in the list of sensitive properties of a cluster configuration:

- fs.s3a.access.key
- fs.s3a.secret.key
- fs.s3n.awsAccessKeyId
- fs.s3n.awsSecretAccessKey
- fs.s3.awsAccessKeyId
- fs.s3.awsSecretAccessKey

Sensitive properties are included but masked when you generate a cluster configuration archive file to deploy on the machine that runs the Developer tool.

Previously, you configured these properties in .xml configuration files on the machines that run the Data Integration Service and the Developer tool.

For more information about sensitive properties, see the *Informatica Big Data Management 10.2 Administrator Guide.*

Sqoop

Effective in version 10.2, if you create a password file to access a database, Sqoop ignores the password file. Sqoop uses the value that you configure in the **Password** field of the JDBC connection.

Previously, you could create a password file to access a database.

For more information, see the "Mapping Objects in the Hadoop Environment" chapter in the *Informatica Big Data Management 10.2 User Guide*.

Command Line Programs

This section describes changes to commands in 10.2.

infacmd ihs Commands

Obsolete Commands

The following table describes obsolete infacmd ihs commands:

Command	Description
BackupData	Backs up HDFS data in the internal Hadoop cluster to a zip file. When you back up the data, the Informatica Cluster Service saves all the data created by Enterprise Information Catalog, such as HBase data, scanner data, and ingestion data.
removesnapshot	Removes existing HDFS snapshots so that you can run the infacmd ihs BackupData command successfully to back up HDFS data.

infacmd ldm Commands

Changed Commands

The following table describes changed infacmd ldm commands:

Command	Change Description
BackupData	Effective in 10.2, the name of the command is changed to BackupContents.
LocalDestination	Effective in 10.2, the **-of** option is added to the BackupContents command.
restoreData	Effective in 10.2, the name of the command is changed to restoreContents.

For more information, see the "infacmd ldm Command Reference" chapter in the *Informatica 10.2 Command Reference*.

Enterprise Information Catalog

This section describes the changes to Informatica Enterprise Information Catalog in 10.2.

Product Name Changes

Effective in version 10.2, Enterprise Information Catalog includes the following name changes:

- The product Informatica Live Data Map is renamed to Informatica Enterprise Information Catalog.
- The Informatica Live Data Map Administrator tool is renamed to Informatica Catalog Administrator.
- The installer is renamed from Live Data Map to Enterprise Information Catalog.

Informatica Analyst

This section describes changes to the Analyst tool in 10.2.

Parameters

This section describes changes to Analyst tool parameters.

System Parameters

Effective in version 10.2, the Analyst tool displays the file path of system parameters in the following format: `$$[Parameter Name]/[Path]`.

Previously, the Analyst tool displayed the local file path of the data object and did not resolve the system parameter.

For more information about viewing data objects, see the *Informatica 10.2 Analyst Tool Guide*.

Intelligent Streaming

This section describes the changes to Informatica Intelligent Streaming in 10.2.

Kafka Data Object Changes

Effective in version 10.2, when you configure the data operation read properties, you can specify the time from which the Kafka source starts reading Kafka messages from a Kafka topic. You can read from or write to a Kafka cluster that is configured for Kerberos authentication.

For more information, see the "Sources and Targets in a Streaming Mapping" chapter in the *Informatica Intelligent Streaming 10.2 User Guide*.

PowerExchange Adapters

This section describes changes to PowerExchange adapters in version 10.2.

PowerExchange Adapters for Informatica

This section describes changes to Informatica adapters in 10.2.

PowerExchange for Amazon S3

Effective in version 10.2, PowerExchange for Amazon S3 has the following changes:

- You can provide the folder path without specifying the bucket name in the advanced properties for read and write operation in the following format: `/<folder_name>`. The Data Integration Service appends this folder path with the folder path that you specify in the connection properties.
  Previously, you specified the bucket name along with the folder path in the advanced properties for read and write operation in the following format: `<bucket_name>/<folder_name>`.

- You can view the bucket name directory following sub directory list in the left panel and selected list of files in the right panel of metadata import browser.
  Previously, PowerExchange for Amazon S3 displayed the list of bucket names in the left panel and folder path along with file names in right panel of metadata import browser.

- PowerExchange for Amazon S3 creates the data object read operation and data object write operation for the Amazon S3 data object automatically.
  Previously, you had to create the data object read operation and data object write operation for the Amazon S3 data object manually.

For more information, see the *Informatica PowerExchange for Amazon S3 10.2 User Guide*

PowerExchange Adapters for PowerCenter

This section describes changes to PowerCenter adapters in version 10.2.

PowerExchange for Amazon Redshift

Effective in version 10.2, you must provide the schema name for the Amazon Redshift table to run mappings successfully.

Previously, mappings would run even if the public schema was selected.

For more information, see the *Informatica PowerExchange for Amazon Redshift 10.2 User Guide for PowerCenter*.

PowerExchange for Email Server

Effective in version 10.2, PowerExchange for Email Server installs with the Informatica services.

Previously, PowerExchange for Email Server had a separate installer.

For more information, see the *Informatica PowerExchange for Email Server 10.2 User Guide for PowerCenter*.

PowerExchange for JD Edwards EnterpriseOne

Effective in version 10.2, PowerExchange for JD Edwards EnterpriseOne installs with the Informatica services.

Previously, PowerExchange for JD Edwards EnterpriseOne had a separate installer.

For more information, see the *Informatica PowerExchange for JD Edwards EnterpriseOne 10.2 User Guide for PowerCenter*.

PowerExchange for JD Edwards World

Effective in version 10.2, PowerExchange for JD Edwards World installs with the Informatica services.

Previously, PowerExchange for JD Edwards World had a separate installer.

For more information, see the *Informatica PowerExchange for JD Edwards World 10.2 User Guide for PowerCenter*.

PowerExchange for LDAP

Effective in version 10.2, PowerExchange for LDAP installs with the Informatica services.

Previously, PowerExchange for LDAP had a separate installer.

For more information, see the *Informatica PowerExchange for LDAP 10.2 User Guide for PowerCenter*.

PowerExchange for Lotus Notes

Effective in version 10.2, PowerExchange for Lotus Notes installs with the Informatica services.

Previously, PowerExchange for Lotus Notes had a separate installer.

For more information, see the *Informatica PowerExchange for Lotus Notes 10.2 User Guide for PowerCenter*.

PowerExchange for Oracle E-Business Suite

Effective in version 10.2, PowerExchange for Oracle E-Business Suite installs with the Informatica services.

Previously, PowerExchange for Oracle E-Business Suite had a separate installer.

For more information, see the *Informatica PowerExchange for Oracle E-Business Suite 10.2 User Guide for PowerCenter*.

PowerExchange for SAP NetWeaver

Effective in version 10.2, Informatica does not package secure transports in a separate folder named `Secure` within the Informatica installer .zip file. Informatica packages both standard and secure transports in the following folders:

- Unicode cofiles: `Informatica installer zip file/saptrans/mySAP/UC/cofiles`
- Unicode data files: `Informatica installer zip file/saptrans/mySAP/UC/data`
- Non-Unicode cofiles: `Informatica installer zip file/saptrans/mySAP/NUC/cofiles`
- Non-Unicode data files: `Informatica installer zip file/saptrans/mySAP/NUC/data`

Previously, Informatica packaged the secure transports in the following folders:

- Unicode cofiles: `Informatica installer zip file/saptrans/mySAP/UC/Secure/cofiles`
- Unicode data files: `Informatica installer zip file/saptrans/mySAP/UC/Secure/data`
- Non-Unicode cofiles: `Informatica installer zip file/saptrans/mySAP/NUC/Secure/cofiles`
- Non-Unicode data files: `Informatica installer zip file/saptrans/mySAP/NUC/Secure/data`

For more information, see the *Informatica PowerExchange for SAP NetWeaver 10.2 User Guide for PowerCenter*.

PowerExchange for Siebel

Effective in version 10.2, PowerExchange for Siebel installs with the Informatica services.

Previously, PowerExchange for Siebel had a separate installer.

For more information, see the *Informatica PowerExchange for Siebel 10.2 User Guide for PowerCenter*.

Security

This section describes changes to security features in 10.2.

SAML Authentication

Effective in version 10.2, you must configure Security Assertion Markup Language (SAML) authentication at the domain level, and on all gateway nodes within the domain.

Previously, you had to configure SAML authentication at the domain level only.

For more information, see the "SAML Authentication for Informatica Web Applications" chapter in the *Informatica 10.2 Security Guide*.

Transformations

This section describes changed transformation behavior in 10.2.

Informatica Transformations

This section describes the changes to the Informatica transformations in 10.2.

Address Validator Transformation

This section describes the changes to the Address Validator transformation.

The Address Validator transformation contains the following updates to address functionality:

All Countries

Effective in version 10.2, the Address Validator transformation uses version 5.11.0 of the Informatica Address Verification software engine. The engine enables the features that Informatica adds to the Address Validator transformation in version 10.2.

Previously, the transformation used version 5.9.0 of the Informatica Address Verification software engine.

Japan

Effective in version 10.2, you can configure a single mapping to return the Choumei Aza code for a current address in Japan. To return the code, select the Current Choumei Aza Code JP port. You can use the code to find the current version of any legacy address that Japan Post recognizes.

Previously, you used the New Choumei Aza Code JP port to return incremental changes to the Choumei Aza code for an address. The transformation did not include the Current Choumei Aza Code JP port. You needed to configure two or more mappings to verify a current Choumei Aza code and the corresponding address.

United Kingdom

Effective in version 10.2, you can configure the Address Validator transformation to return postal, administrative, and traditional county information from the Royal Mail Postcode Address File. The transformation returns the information on the Province ports.

Previously, the transformation returned postal county information when the information was postally relevant.

The following table shows the ports that you can select for each information type:

County Information Type	Address Element
Postal	Province 1
Administrative	Province 2
Traditional	Province 3

Updated Certification Standards in Multiple Countries

Effective in version 10.2, Informatica supports the following certification standards for address verification software:

- Address Matching Approval System (AMAS) from Australia Post. Updated to Cycle 2017.
- SendRight certification from New Zealand Post. Updated to Cycle 2017.
- Software Evaluation and Recognition Program (SERP) from Canada Post. Updated to Cycle 2017.

Informatica continues to support the current versions of the Coding Accuracy Support System (CASS) standards from the United States Postal Service and the Service National de L'Adresse (SNA) standard from La Poste of France.

For more information, see the *Informatica 10.2 Developer Transformation Guide* and the *Informatica 10.2 Address Validator Port Reference*.

For comprehensive information about the updates to the Informatica Address Verification software engine from version 5.9.0 through version 5.11.0, see the *Informatica Address Verification 5.11.0 Release Guide*.

Expression Transformation

Effective in version 10.2, you can configure the Expression transformation to be an active transformation on the Spark engine by using a window function or an aggregate function with windowing properties.

Previously, the Expression transformation could only be a passive transformation.

For more information, see the *Big Data Management 10.2 Administrator Guide*.

Normalizer Transformation

Effective in version 10.2, the option to disable Generate First Level Output Groups is no longer available in the advanced properties of the Normalizer transformation.

Previously, you could select this option to suppress the generation of first level output groups.

For more information, see the *Informatica Big Data Management 10.2 Developer Transformation Guide*.

Workflows

This section describes changed workflow behavior in version 10.2.

Informatica Workflows

This section describes the changes to Informatica workflow behavior in 10.2.

Workflow Variables in Task Instance Notifications

Effective in version 10.2, the workflow variable *$taskEvent.startOwner* changes name to *$taskEvent.owner*. The usage of the variable does not change in version 10.2.

For more information, see the "Human Task" chapter in the *Informatica 10.2 Developer Workflow Guide*.

CHAPTER 7

Release Tasks (10.2)

This chapter includes the following topic:

PowerExchange Adapters

This section describes release tasks for PowerExchange adapters in version 10.2.

PowerExchange Adapters for PowerCenter

This section describes release tasks for PowerCenter adapters in version 10.2.

PowerExchange for Amazon Redshift

Effective in version 10.2, for existing mappings where public schema is selected, ensure that the schema name is correct and works for the Redshift table. The public schema might not work for all the tables.

For more information, see the *Informatica 10.2 PowerExchange for Amazon Redshift User Guide for PowerCenter*

PowerExchange for Amazon S3

Effective in version 10.2, when you upgrade from 9.5.1 or 9.6.1, the upgrade process does not retain all property values in the connection. After you upgrade, you must reconfigure the following properties:

Property	Description
Access Key	The access key ID used to access the Amazon account resources. Required if you do not use AWS Identity and Access Management (IAM) authentication. **Note:** Ensure that you have valid AWS credentials before you create a connection.
Secret Key	The secret access key used to access the Amazon account resources. This value is associated with the access key and uniquely identifies the account. You must specify this value if you specify the access key ID. Required if you do not use AWS Identity and Access Management (IAM) authentication.
Master Symmetric Key	Optional. Provide a 256-bit AES encryption key in the Base64 format when you enable client-side encryption. You can generate a key using a third-party tool. If you specify a value, ensure that you specify the encryption type as client side encryption in the target session properties.

For more information, see the *Informatica 10.2 PowerExchange for Amazon S3 User Guide for PowerCenter*

PowerExchange for Microsoft Dynamics CRM

When you upgrade from an earlier version, you must copy the `.jar` files in the installation location of 10.2.

- For the client, if you upgrade from 9.x to 10.2, copy the `local_policy.jar`, `US_export_policy.jar`, and cacerts files from the following 9.x installation folder `<Informatica installation directory>\clients\java\jre\lib\security` to the following 10.2 installation folder `<Informatica installation directory>\clients\java\32bit\jre\lib\security`.
  If you upgrade from 10.x to 10.2, copy the `local_policy.jar`, `US_export_policy.jar`, and cacerts files from the following 10.x installation folder `<Informatica installation directory>\clients\java\32bit\jre\lib\security` to the corresponding 10.2 folder.

- For the server, copy the `local_policy.jar`, `US_export_policy.jar`, and cacerts files from the `<Informatica installation directory>java/jre/lib/security` folder of the previous release to the corresponding 10.2 folder.

When you upgrade from an earlier version, you must copy the `msdcrm` folder in the installation location of 10.2.

- For the client, copy the `msdcrm` folder from the `<Informatica installation directory>\clients\PowerCenterClient\client\bin\javalib` folder of the previous release to the corresponding 10.2 folder.

- For the server, copy the `msdcrm` folder from the `<Informatica installation directory>/server/bin/javalib` folder of the previous release to the corresponding 10.2 folder.

PowerExchange for SAP NetWeaver

Effective in version 10.2, Informatica implemented the following changes in PowerExchange for SAP NetWeaver support for PowerCenter:

Dropped Support for the CPI-C Protocol

Effective in version 10.2, Informatica dropped support for the CPI-C protocol.

Use the RFC or HTTP protocol to generate and install ABAP programs while reading data from SAP tables.

If you upgrade ABAP mappings that were generated with the CPI-C protocol, you must complete the following tasks:

1. Regenerate and reinstall the ABAP program by using stream (RFC/HTTP) mode.
2. Create a System user or a communication user with the appropriate authorization profile to enable dialog-free communication between SAP and Informatica.

For more information, see the *Informatica PowerExchange for SAP NetWeaver 10.2 User Guide for PowerCenter*.

Dropped Support for ABAP Table Reader Standard Transports

Effective in version 10.2, Informatica dropped support for the ABAP table reader standard transports. Informatica will not ship the standard transports for ABAP table reader. Informatica will ship only secure transports for ABAP table reader.

If you upgrade from an earlier version, you must delete the standard transports and install the secure transports.

For more information, see the *Informatica PowerExchange for SAP NetWeaver 10.2 Transport Versions Installation Notice*.

Added Support for HTTP Streaming for ABAP Table Reader Mappings

Effective in version 10.2, when you run ABAP mappings to read data from SAP tables, you can configure HTTP streaming.

To use HTTP stream mode for upgraded ABAP mappings, perform the following tasks:

1. Regenerate and reinstall the ABAP program in stream mode.
2. Create an SAP ABAP HTTP streaming connection.
3. Configure the session to use the SAP streaming reader, an SAP ABAP HTTP streaming connection, and an SAP R/3 application connection.

Note: If you configure HTTP streaming, but do not regenerate and reinstall the ABAP program in stream mode, the session fails.

Part III: Version 10.1.1

This part contains the following chapters:

CHAPTER 8

New Features, Changes, and Release Tasks (10.1.1 HotFix 1)

This chapter includes the following topics:

New Products (10.1.1 HotFix 1)

This section describes new products in version 10.1.1 HotFix 1.

PowerExchange for Cloud Applications

Effective in version 10.1.1 HotFix 1, you can use PowerExchange for Cloud Applications to connect to Informatica Cloud from PowerCenter. You can read data from or write data to data sources for which connections are available in Informatica Cloud. It is not required to have the PowerExchange for the respective cloud application in PowerCenter.

For more information, see the *Informatica PowerExchange for Cloud Applications 10.1.1 HotFix 1 User Guide*.

New Features (10.1.1 HotFix 1)

This section describes new features in version 10.1.1 HotFix 1.

Command Line Programs

This section describes new commands in version 10.1.1 HotFix 1.

infacmd dis Commands (10.1.1 HF1)

The following table describes new infacmd dis commands:

Command	Description
disableMappingValidationEnvironment	Disables the mapping validation environment for mappings that are deployed to the Data Integration Service.
enableMappingValidationEnvironment	Enables a mapping validation environment for mappings that are deployed to the Data Integration Service.
setMappingExecutionEnvironment	Specifies the mapping execution environment for mappings that are deployed to the Data Integration Service.

For more information, see the "Infacmd dis Command Reference" chapter in the *Informatica 10.1.1 HotFix1 Command Reference.*

infacmd mrs Commands (10.1.1 HF1)

The following table describes new infacmd mrs commands:

Command	Description
disableMappingValidationEnvironment	Disables the mapping validation environment for mappings that you run from the Developer tool.
enableMappingValidationEnvironment	Enables a mapping validation environment for mappings that you run from the Developer tool.
setMappingExecutionEnvironment	Specifies the mapping execution environment for mappings that you run from the Developer tool.

For more information, see the "Infacmd mrs Command Reference" chapter in the *Informatica 10.1.1 HotFix1 Command Reference.*

infacmd ps Command

The following table describes a new infacmd ps command:

Command	Description
restoreProfilesAndScorecards	Restores profiles and scorecards from a previous version to version 10.1.1 HotFix 1.

For more information, see the "infacmd ps Command Reference" chapter in the Informatica 10.1.1 HotFix 1 Command Reference.

Informatica Analyst

This section describes new Analyst tool features in version 10.1.1 HotFix 1.

Profiles and Scorecards

This section describes new Analyst tool features for profiles and scorecards.

Invalid Rows Worksheet

Effective in version 10.1.1 HotFix1, scorecard export results include invalid source rows after you choose the **Data** > **All** option in the **Export data to a file** dialog box.

For more information about scorecards, see the "Scorecards in Informatica Analyst" chapter in the *Informatica 10.1.1 HotFix1 Data Discovery Guide*.

PowerCenter

This section describes new PowerCenter features in version 10.1.1 HotFix 1.

Pushdown Optimization for Greenplum

Effective in version 10.1.1 HotFix 1, when the connection type is ODBC, the PowerCenter Integration Service can push TRUNC(DATE), CONCAT(), and TO_CHAR(DATE) functions to Greenplum using source-side and full pushdown optimization.

For more information, see the *Informatica PowerCenter 10.1.1 HotFix 1 Advanced Workflow Guide*.

Pushdown Optimization for Microsoft Azure SQL Data Warehouse

Effective in version 10.1.1 HotFix 1, when the connection type is ODBC, you can configure source-side or full pushdown optimization to push the transformation logic to Microsoft Azure SQL Data Warehouse.

For more information, see the *Informatica PowerCenter 10.1.1 HotFix 1 Advanced Workflow Guide*.

PowerExchange Adapters

This section describes new PowerExchange adapter features in version 10.1.1 HotFix 1.

PowerExchange Adapters for PowerCenter®

This section describes new PowerCenter adapter features in version 10.1.1 HotFix 1.

PowerExchange for Amazon Redshift

This section describes new PowerExchange for Amazon Redshift features in version 10.1.1 HotFix 1:

- You can read data from or write data to the following regions:
    - Asia Pacific (Mumbai)
    - Canada (Central)
    - US East (Ohio)
- PowerExchange for Amazon Redshift supports the asterisk pushdown operator (*) that can be pushed to the Amazon Redshift database by using source-side, target-side, or full pushdown optimization.
- For client-side and server-side encryption, you can configure the customer master key ID generated by AWS Key Management Service (AWS KMS) in the connection.

For more information, see the *Informatica 10.1.1 HotFix 1 PowerExchange for Amazon Redshift User Guide for PowerCenter*.

PowerExchange for Amazon S3

This section describes new PowerExchange for Amazon S3 features in version 10.1.1 HotFix 1:

- You can read data from or write data to the following regions:
  - Asia Pacific (Mumbai)
  - Canada (Central)
  - US East (Ohio)
- For client-side and server-side encryption, you can configure the customer master key ID generated by AWS Key Management Service (AWS KMS) in the connection.
- When you write data to the Amazon S3 buckets, you can compress the data in GZIP format.
- You can override the Amazon S3 folder path when you run a mapping.

For more information, see the *Informatica PowerExchange for Amazon S3 10.1.1 HotFix 1 User Guide for PowerCenter*.

PowerExchange for Microsoft Azure Blob Storage

Effective in version 10.1.1 HotFix 1, you can use append blob type target session property to write data to Microsoft Azure Blob Storage.

For more information, see the *Informatica PowerExchange for Microsoft Azure Blob Storage 10.1.1 HotFix 1 User Guide*.

PowerExchange for Microsoft Azure SQL Data Warehouse

Effective in version 10.1.1 HotFix 1, you can use the following target session properties with PowerExchange for Microsoft Azure SQL Data Warehouse:

- Update as Update. The PowerCenter Integration Service updates all rows as updates.
- Update else Insert. The PowerCenter Integration Service updates existing rows and inserts other rows as if marked for insert.
- Delete. The PowerCenter Integration Service deletes the specified records from Microsoft Azure SQL Data Warehouse.

For more information, see the *Informatica PowerExchange for Microsoft Azure SQL Data Warehouse 10.1.1 HotFix 1 User Guide for PowerCenter*.

PowerExchange for Microsoft Dynamics CRM

Effective in version 10.1.1 HotFix 1, you can use the following target session properties with PowerExchange for Microsoft Dynamics CRM:

- Add row reject reason. Select to include the reason for rejection of rows to the reject file.
- Alternate Key Name. Indicates whether the column is an alternate key for an entity. Specify the name of the alternate key. You can use alternate key in update and upsert operations.

For more information, see the *Informatica PowerExchange for Microsoft Dynamics CRM 10.1.1 HotFix 1 User Guide for PowerCenter*.

PowerExchange for SAP NetWeaver

Effective in version 10.1.1 HotFix 1, PowerExchange for SAP NetWeaver supports the SSTRING data type when you read data from SAP tables through ABAP. The SSTRING data type is represented as SSTR in PowerCenter.

For more information, see the *Informatica PowerExchange for SAP NetWeaver 10.1.1 HotFix 1 User Guide*.

Changes (10.1.1 HotFix 1)

This section describes changes in version 10.1.1 HotFix 1.

Support Changes

Effective in version 10.1.1 HF1, the following changes apply to Informatica support for third-party platforms and systems:

Big Data Management Hadoop Distributions

The following table lists the supported Hadoop distribution versions and changes in 10.1.1 HotFix 1:

Distribution	Supported Versions	10.1.1 HotFix 1 Changes
Amazon EMR	5.4	To enable support for Amazon EMR 5.4, apply EBF-9585 to Big Data Management 10.1.1 Hot Fix 1. Big Data Management version 10.1.1 Update 2 supports Amazon EMR 5.0.
Azure HDInsight	3.5	Added support for version 3.5.
Cloudera CDH	5.8, 5.9, 5.10, 5.11	Added support for versions 5.10, 5.11.
Hortonworks HDP	2.3, 2.4, 2.5, 2.6	Added support for version 2.6.
IBM BigInsights	4.2	No change.
MapR	5.2.0 MEP binary v. 1.0	No change.

To see a list of the latest supported versions, see the Product Availability Matrix on the Informatica Customer Portal: https://network.informatica.com/community/informatica-network/product-availability-matrices.

CHAPTER 9

New Features, Changes, and Release Tasks (10.1.1 Update 2)

This chapter includes the following topics:

New Products (10.1.1 Update 2)

This section describes new products in version 10.1.1 Update 2.

PowerExchange for MapR-DB

Effective in version 10.1.1 Update 2, you can use PowerExchange for MapR-DB to read data from and write data to MapR-DB binary tables.

PowerExchange for MapR-DB uses the HBase API to connect to MapR-DB. To connect to a MapR-DB table, you must create an HBase connection in which you must specify the database type as MapR-DB. You must create an HBase data object read or write operation, and add it to a mapping to read or write data.

You can validate and run mappings in the native environment or on the Blaze engine in the Hadoop environment.

For more information, see the *Informatica PowerExchange for MapR-DB 10.1.1 Update 2 User Guide*.

New Features (10.1.1 Update 2)

This section describes new features in version 10.1.1 Update 2.

Big Data Management

This section describes new big data features in version 10.1.1 Update 2.

Truncate Hive table partitions on mappings that use the Blaze run-time engine

Effective in version 10.1.1 Update 2, you can truncate Hive table partitions on mappings that use the Blaze run-time engine.

For more information about truncating partitions in a Hive target, see the *Informatica 10.1.1 Update 2 Big Data Management User Guide.*

Filters for partitioned columns on the Blaze engine

Effective in version 10.1.1 Update 2, the Blaze engine can push filters on partitioned columns down to the Hive source to increase performance.

When a mapping contains a Filter transformation on a partitioned column of a Hive source, the Blaze engine reads only the partitions with data that satisfies the filter condition. To enable the Blaze engine to read specific partitions, the Filter transformation must be the next transformation after the source in the mapping.

For more information, see the *Informatica 10.1.1 Update 2 Big Data Management User Guide*.

OraOop support on the Spark engine

Effective in version 10.1.1 Update 2, you can configure OraOop to run Sqoop mappings on the Spark engine. When you read data from or write data to Oracle, you can configure the direct argument to enable Sqoop to use OraOop.

OraOop is a specialized Sqoop plug-in for Oracle that uses native protocols to connect to the Oracle database. When you configure OraOop, the performance improves.

For more information, see the *Informatica 10.1.1 Update 2 Big Data Management User Guide*.

Sqoop support for native Teradata mappings on Cloudera clusters

Effective in version 10.1.1 Update 2, if you use a Teradata PT connection to run a mapping on a Cloudera cluster and on the Blaze engine, the Data Integration Service invokes the Cloudera Connector Powered by Teradata at run time. The Data Integration Service then runs the mapping through Sqoop.

For more information, see the *Informatica 10.1.1 Update 2 PowerExchange for Teradata Parallel Transporter API User Guide*.

Scheduler support on Blaze and Spark engines

Effective in version 10.1.1 Update 2, the following schedulers are valid for Hadoop distributions on both Blaze and Spark engines:

- Fair Scheduler. Assigns resources to jobs such that all jobs receive, on average, an equal share of resources over time.
- Capacity Scheduler. Designed to run Hadoop applications as a shared, multi-tenant cluster. You can configure Capacity Scheduler with or without node labeling. Node label is a way to group nodes with similar characteristics.

For more information, see the Mappings in the Hadoop Environment chapter of the *Informatica 10.1.1 Update 2 Big Data Management User Guide*.

Support for YARN queues on Blaze and Spark engines

Effective in version 10.1.1 Update 2, you can direct Blaze and Spark jobs to a specific YARN scheduler queue. Queues allow multiple tenants to share the cluster. As you submit applications to YARN, the scheduler assigns them to a queue. You configure the YARN queue in the Hadoop connection properties.

For more information, see the Mappings in the Hadoop Environment chapter of the *Informatica 10.1.1 Update 2 Big Data Management User Guide.*

Hadoop security features on IBM BigInsights 4.2

Effective in version 10.1.1 Update 2, you can use the following Hadoop security features on the IBM BigInsights 4.2 Hadoop distribution:

- Apache Knox
- Apache Ranger
- HDFS Transparent Encryption

For more information, see the *Informatica 10.1.1 Update 2 Big Data Management Security Guide*.

SSL/TLS security modes

Effective in version 10.1.1 Update 2, you can use the SSL and TLS security modes on the Cloudera and HortonWorks Hadoop distributions, including the following security methods and plugins:

- Kerberos authentication
- Apache Ranger
- Apache Sentry
- Name node high availability
- Resource Manager high availability

For more information, see the *Informatica 10.1.1 Update 2 Big Data Management Installation and Configuration Guide*.

Hive sources and targets on Amazon S3

Effective in version 10.1.1 Update 2, Big Data Management supports reading and writing to Hive on Amazon S3 buckets for clusters configured with the following Hadoop distributions:

- Amazon EMR
- Cloudera
- HortonWorks
- MapR
- BigInsights

For more information, see the *Informatica 10.1.1 Update 2 Big Data Management User Guide*.

Enterprise Information Catalog

This section describes new features in Enterprise Information Catalog version 10.1.1 Update 2.

File System resource

Effective in version 10.1.1 Update 2, you can create a **File System** resource to import metadata from files in Windows and Linux file systems.

For more information, see the *Informatica 10.1.1 Update 2 Live Data Map Administrator Guide*.

Apache Ranger-enabled clusters

Effective in version 10.1.1 Update 2, you can deploy Enterprise Information Catalog on Apache Ranger-enabled clusters. Apache Ranger provides a security framework to manage the security of the clusters.

Enhanced SSH support for deploying Informatica Cluster Service

Effective in version 10.1.1 Update 2, you can deploy Informatica Cluster Service on hosts where Centrify is enabled. Centrify integrates with an existing Active Directory infrastructure to manage user authentication on remote Linux hosts.

Intelligent Data Lake

This section describes new Intelligent Data Lake features in version 10.1.1 Update 2.

Hadoop ecosystem

Effective in version 10.1.1 Update 2, you can use following Hadoop distributions as a Hadoop data lake:

- Cloudera CDH 5.9
- Hortonworks HDP 2.3, 2.4, and 2.5
- Azure HDInsight 3.5
- Amazon EMR 5.0
- IBM BigInsights 4.2

Using MariaDB for the Data Preparation Service

Effective in version 10.1.1 Update 2, you can use MariaDB 10.0.28 for the Data Preparation Service repository.

Viewing column-level lineage

Effective in version 10.1.1 Update 2, data analysts can view lineage of individual columns in a table corresponding to activities such as data asset copy, import, export, publication, and upload.

SSL/TLS support

Effective in version 10.1.1 Update 2, you can integrate Intelligent Data Lake with Cloudera 5.9 clusters that are SSL/TLS enabled.

PowerExchange Adapters for Informatica

This section describes new Informatica adapter features in version 10.1.1 Update 2.

PowerExchange for Amazon Redshift

Effective in version 10.1.1 Update 2, you can select multiple schemas for Amazon Redshift objects.

For more information, see the *Informatica 10.1.1 Update 2 PowerExchange for Amazon Redshift User Guide*.

Changes (10.1.1 Update 2)

This section describes changes in version 10.1.1 Update 2.

Support Changes

This section describes the support changes in version 10.1.1 Update 2.

Distribution support changes for Big Data Management

The following table lists the supported Hadoop distribution versions and changes in 10.1.1 Update 2:

Distribution	Supported Versions	10.1.1 Update 2 Changes
Amazon EMR	5.0.0	No change.
Azure HDInsight	3.5 *	Added support for version 3.5 Dropped support for version 3.4.
Cloudera CDH	5.8, 5.9, 5.10 *	Added support for version 5.10.
Hortonworks HDP	2.3, 2.4, 2.5	Added support for versions 2.3 and 2.4.
IBM BigInsights	4.2	No change.
MapR	5.2	Reinstated support. Added support for version 5.2. Dropped support for version 5.1.

*Azure HDInsight 3.5 and Cloudera CDH 5.10 are available for technical preview. Technical preview functionality is supported but is not production-ready. Informatica recommends that you use in non-production environments only.

For a complete list of Hadoop support, see the Product Availability Matrix on Informatica Network: https://network.informatica.com/community/informatica-network/product-availability-matrices

Dropped support for Teradata Connector for Hadoop (TDCH) and Teradata PT objects on the Blaze engine

Effective in version 10.1.1 Update 2, Informatica dropped support for Teradata Connector for Hadoop (TDCH) on the Blaze engine. The configuration for Sqoop connectivity in 10.1.1 Update 2 depends on the Hadoop distribution:

IBM BigInsights and MapR

You can configure Sqoop connectivity through the JDBC connection. For information about configuring Sqoop connectivity through JDBC connections, see the *Informatica 10.1.1 Update 2 Big Data Management User Guide*.

Cloudera CDH

You can configure Sqoop connectivity through the Teradata PT connection and the Cloudera Connector Powered by Teradata.

1. Download the Cloudera Connector Powered by Teradata .jar files and copy them to the node where the Data Integration Service runs. For more information, see the *Informatica 10.1.1 Update 2 PowerExchange for Teradata Parallel Transporter API User Guide*.

2. Move the configuration parameters that you defined in the `InfaTDCHConfig.txt` file to the **Additional Sqoop Arguments** field in the Teradata PT connection. See the Cloudera Connector Powered by Teradata documentation for a list of arguments that you can specify.

Hortonworks HDP

You can configure Sqoop connectivity through the Teradata PT connection and the Hortonworks Connector for Teradata.

1. Download the Hortonworks Connector for Teradata .jar files and copy them to the node where the Data Integration Service runs. For more information, see the *Informatica 10.1.1 Update 2 PowerExchange for Teradata Parallel Transporter API User Guide*.

2. Move the configuration parameters that you defined in the `InfaTDCHConfig.txt` file to the **Additional Sqoop Arguments** field in the Teradata PT connection. See the Hortonworks Connector for Teradata documentation for a list of arguments that you can specify.

Note: You can continue to use TDCH on the Hive engine through Teradata PT connections.

Deprecated support of Sqoop connectivity through Teradata PT data objects and Teradata PT connections

Effective in version 10.1.1 Update 2, Informatica deprecated Sqoop connectivity through Teradata PT data objects and Teradata PT connections for Cloudera CDH and Hortonworks. Support will be dropped in a future release.

To read data from or write data to Teradata by using TDCH and Sqoop, Informatica recommends that you configure Sqoop connectivity through JDBC connections and relational data objects.

Big Data Management

This section describes the changes to big data in version 10.1.1 Update 2.

Sqoop

Effective in version 10.1.1 Update 2, you can no longer override the user name and password in a Sqoop mapping by using the --username and --password arguments. Sqoop uses the values that you configure in the **User Name** and **Password** fields of the JDBC connection.

For more information, see the *Informatica 10.1.1 Update 2 Big Data Management User Guide*.

Enterprise Information Catalog

This section describes the changes to the Enterprise Information Catalog in version 10.1.1 Update 2.

Asset path

Effective in version 10.1.1 Update 2, you can view the path to the asset in the Asset Details view along with other general information about the asset.

For more information, see the *Informatica 10.1.1 Update 2 Enterprise Information Catalog User Guide*.

Business terms in the Profile Results section

Effective in version 10.1.1 Update 2, the profile results section for tabular assets also includes business terms. Previously, the profile results section included column names, data types, and data domains.

For more information, see the *Informatica 10.1.1 Update 2 Enterprise Information Catalog User Guide*.

URLs as attribute values

Effective in version 10.1.1 Update 2, if you had configured a custom attribute to allow you to enter URLs as the attribute value, you can assign multiple URLs as attribute values to a technical asset.

For more information, see the *Informatica 10.1.1 Update 2 Enterprise Information Catalog User Guide*.

Detection of CSV file headers

> Effective in version 10.1.1 Update 2, you can configure the following resources to automatically detect headers for CSV files from which you extract metadata:
>
> - Amazon S3
>
> - HDFS
>
> - File System
>
> For more information, see the *Informatica 10.1.1 Update 2 Live Data Map Administrator Guide*.

Amazon Redshift resource

> Effective in version 10.1.1 Update 2, you can import multiple schemas for an Amazon Redshift resource.
>
> For more information, see the *Informatica 10.1.1 Update 2 Live Data Map Administrator Guide*.

Profiling for Hive resource on Data Integration Service

> Effective in version 10.1.1 Update 2, you can run Hive resources on Data Integration Service for profiling.
>
> For more information, see the *Informatica 10.1.1 Update 2 Live Data Map Administrator Guide*.

PowerExchange Adapters for Informatica

This section describes changes to Informatica adapters in version 10.1.1 Update 2.

PowerExchange for Amazon Redshift

Effective in version 10.1.1 Update 2, you can select multiple schemas for Amazon Redshift objects. To select multiple schemas, leave the **Schema** field blank in the connection properties. In earlier releases, selecting schema was mandatory and you could select only one schema.

If you upgrade to version 10.1.1 Update 2, the PowerExchange for Redshift mappings created in earlier versions must have the relevant schema name in the connection property. Else, mappings fail when you run them on version 10.1.1 Update 2.

For more information, see the *Informatica 10.1.1 Update 2 PowerExchange for Amazon Redshift User Guide*.

CHAPTER 10

New Features, Changes, and Release Tasks (10.1.1 Update 1)

This chapter includes the following topics:

New Features (10.1.1 Update 1)

This section describes new features in version 10.1.1 Update 1.

Big Data Management

This section describes new big data features in version 10.1.1 Update 1.

Sqoop Support for Native Teradata Mappings

Effective in version 10.1.1 Update 1, if you use a Teradata PT connection to run a mapping on a Hortonworks cluster and on the Blaze engine, the Data Integration Service invokes the Hortonworks Connector for Teradata at run time. The Data Integration Service then runs the mapping through Sqoop.

For more information, see the *Informatica 10.1.1 Update 1 PowerExchange for Teradata Parallel Transporter API User Guide*.

SQL Override Support for Native Teradata Mappings

Effective in version 10.1.1 Update 1, if you use a Teradata PT connection to run a mapping on a Hortonworks cluster and on the Blaze engine, you can configure an SQL override query. You can also parameterize the SQL override query.

For more information, see the *Informatica 10.1.1 Update 1 PowerExchange for Teradata Parallel Transporter API User Guide*.

Changes (10.1.1 Update 1)

This section describes changes in version 10.1.1 Update 1.

PowerExchange Adapters for Informatica

This section describes PowerExchange adapter changes in version 10.1.1 Update 1.

PowerExchange for Amazon S3

Effective in version 10.1.1 Update 1, PowerExchange for Amazon S3 has the following advanced properties for an Amazon S3 data object read and write operation:

- Folder Path
- Download S3 File in Multiple Parts
- Staging Directory

Previously, the advanced properties for an Amazon S3 data object read and write operation were:

- S3 Folder Path
- Enable Download S3 Files in Multiple Parts
- Local Temp Folder Path

For more information, see the *Informatica 10.1.1 Update 1 PowerExchange for Amazon S3 User Guide*.

Release Tasks (10.1.1 Update 1)

This section describes the release tasks for version 10.1.1 Update 1.

PowerExchange Adapters for Informatica

This section describes PowerExchange adapter release tasks for version 10.1.1 Update 1.

PowerExchange for Teradata Parallel Transporter API

Effective in version 10.1.1 Update 1, if you use a Teradata PT connection to run a mapping on a Hortonworks cluster and on the Blaze engine, the Data Integration Service invokes the Hortonworks Connector for Teradata at run time. The Data Integration Service then runs the mapping through Sqoop.

If you had configured Teradata Connector for Hadoop (TDCH) to run Teradata mappings on the Blaze engine and installed 10.1.1 Update 1, the Data Integration Service ignores the TDCH configuration. You must perform the following upgrade tasks to run Teradata mappings on the Blaze engine:

1. Install 10.1.1 Update 1.
2. Download the Hortonworks Connector for Teradata JAR files.
3. Move the configuration parameters that you defined in the `InfaTDCHConfig.txt` file to the **Additional Sqoop Arguments** field in the Teradata PT connection. See the Hortonworks for Teradata Connector documentation for a list of arguments that you can specify.

Note: If you had configured TDCH to run Teradata mappings on the Blaze engine and on a distribution other than Hortonworks, do not install 10.1.1 Update 1. You can continue to use version 10.1.1 to run mappings with TDCH on the Blaze engine and on a distribution other than Hortonworks.

For more information, see the *Informatica 10.1.1 Update 1 PowerExchange for Teradata Parallel Transporter API User Guide*.

CHAPTER 11

New Products (10.1.1)

This chapter includes the following topics:

Intelligent Streaming

With the advent of big data technologies, organizations are looking to derive maximum benefit from the velocity of data, capturing it as it becomes available, processing it, and responding to events in real time. By adding real-time streaming capabilities, organizations can leverage the lower latency to create a complete, up-to-date view of customers, deliver real-time operational intelligence to customers, improve fraud detection, reduce security risk, improve physical asset management, improve total customer experience, and generally improve their decision-making processes by orders of magnitude.

In 10.1.1, Informatica introduces Intelligent Streaming, a new product to help IT derive maximum value from real-time queues by streaming data, processing it, and extracting meaningful business value in near real time. Customers can process diverse data types and from non-traditional sources, such as website log file data, sensor data, message bus data, and machine data, in flight and with high degrees of accuracy.

Intelligent Streaming is built as a capability extension of Informatica's Intelligent Data Platform and provides the following benefits for IT:

- Create and run streaming (continuous-processing) mappings.
- Collect events from real-time queues such as Apache Kafka and JMS.
- Transform the data, create business rules for the transformed data, detect real-time patterns, and drive automated responses or alerts.
- Provide management and monitoring capabilities of streams at runtime.
- Provide at-least-once delivery guarantees.
- Granulate lifecycle controls based on number of rows processed or time of execution.
- Reuse and maintain event processing logic, including batch mappings (after some modifications).

Intelligent Streaming has the following features:

Capture and Transport Stream Data

You can stream the following types of data from sources such as Kafka or JMS, in JSON, XML, or Avro formats:

- Application and infrastructure log data

- Change data capture (CDC) from relational databases

- Clickstreams from web servers

- Social media event streams

- Time-series data from IoT devices

- Message bus data

- Programmable logic controller (PLC) data

- Point of sale data from devices

In addition, Informatica customers can leverage Informatica's Vibe Data Stream (licensed separately) to collect and ingest data in real time, for example, data from sensors, and machine logs, to a Kafka queue. Intelligent Streaming can then process this data.

Refine, Enrich, Analyze, and Process Stream Data

Use the underlying processing platform to run the following complex data transformations in real time without coding or scripting:

- Window Transformation for Streaming use cases with the option of sliding and tumbling windows.

- Filter, Expression, Union, Router, Aggregate, Joiner, Lookup, Java, and Sorter transformations can now be used with Streaming mappings and are executed on Spark Streaming.

- Lookup transformations can be used with Flat file, HDFS, Sqoop, and Hive.

Publish Data

You can stream data to different types of targets, such as Kafka, HDFS, NoSQL databases, and enterprise messaging systems.

Intelligent Streaming is built on the Informatica Big Data Platform platform and extends the platform to provide streaming capabilities. Intelligent Streaming uses Spark Streaming to process streamed data. It uses YARN to manage the resources on a Spark cluster more efficiently and uses third-parties distributions to connect to and push job processing to a Hadoop environment.

Use Informatica Developer (the Developer tool) to create streaming mappings. Use the Hadoop run-time environment and the Spark engine to run the mapping. You can configure high availability to run the streaming mappings on the Hadoop cluster.

For more information about Intelligent Streaming, see the *Informatica Intelligent Streaming User Guide*.

PowerExchange Adapters

PowerExchange Adapters for Informatica

This section describes new Informatica adapters in version 10.1.1.

PowerExchange for Amazon S3

Effective in version 10.1.1, you can create an Amazon S3 connection to specify the location of Amazon S3 sources and targets you want to include in a data object. You can use the Amazon S3 connection in data object read and write operations. You can validate and run mappings in the native environment or on the Blaze engine in the Hadoop environment.

For more information, see the *Informatica PowerExchange for Amazon S3 10.1.1 User Guide*.

CHAPTER 12

New Features (10.1.1)

This chapter includes the following topics:

Application Services

This section describes new application service features in version 10.1.1.

Analyst Service

Effective in version 10.1.1, you can configure an Analyst Service to store all audit data for exception management tasks in a single database. The database stores a record of the work that users perform on Human task instances in the Analyst tool that the Analyst Service specifies.

Set the database connection and the schema for the audit tables on the Human task properties of the Analyst Service in the Administrator tool. After you specify a connection and schema, use the **Actions** menu options in the Administrator tool to create the audit database contents. Or, use the infacmd as commands to set the database and schema and to create the audit database contents. To set the database and the schema, run infacmd as updateServiceOptions. To create the database contents, run infacmd as createExceptionAuditTables

If you do not specify a connection and schema, the Analyst Service creates audit tables for each task instance in the database that stores the task instance data.

For more information, see the *Informatica 10.1.1 Application Service Guide* and the *Informatica 10.1.1 Command Reference*.

Big Data

This section describes new big data features in version 10.1.1.

Blaze Engine

Effective in version 10.1.1, the Blaze engine has the following new features:

Hive Sources and Targets on the Blaze Engine

Effective in version 10.1.1, Hive sources and targets have the following additional support on the Blaze engine:

- Hive decimal data type values with precision 38
- Quoted identifiers in Hive table names, column names, and schema names
- Partitioned Hive tables as targets
- Bucketed Hive tables as source and targets
- SQL overrides for Hive sources
- Table locking for Hive sources and targets
- Create or replace target tables for Hive targets
- Truncate target table for Hive targets and Hive partitioned tables

For more information, see the "Mapping Objects in the Hadoop Environment" chapter in the *Informatica Big Data Management® 10.1.1 User Guide*.

Transformation Support on the Blaze Engine

Effective in version 10.1.1, transformations have the following additional support on the Blaze engine:

- Lookup transformation. You can use SQL overrides and filter queries with Hive lookup sources.
- Sorter transformation. Global sorts are supported when the Sorter transformation is connected to a flat file target. To maintain global sort order, you must enable the Maintain Row Order property in the flat file target. If the Sorter transformation is midstream in the mapping, then rows are sorted locally.
- Update Strategy transformation. The Update Strategy transformation is supported with some restrictions.

For more information, see the "Mapping Objects in the Hadoop Environment" chapter in the *Informatica Big Data Management 10.1.1 User Guide*.

Blaze Engine Monitoring

Effective in Version 10.1.1, more detailed statistics about mapping jobs are available in the Blaze Summary Report. In the Blaze Job Monitor, a green summary report button appears beside the names of successful grid tasks which opens the Blaze Summary Report.

The Blaze Summary Report contains the following information about a mapping job:

- Time taken by individual segments. A pie chart of segments within the grid task.

- Mapping properties. A table containing basic information about the mapping job.
- Tasklet execution time. A time series graph of all tasklets within the selected segment.
- Selected tasklet information. Source and target row counts and cache information for each individual tasklet.

Note: The Blaze Summary Report is in beta. It contains most of the major features, but is not yet complete.

Blaze Engine Logs

Effective in version 10.1.1, the following error logging enhancements are available on the Blaze engine:

- Execution statistics are available in the LDTM log when the log tracing level is set to verbose initialization or verbose data. The log includes the following mapping execution details:
  - Start time, end time, and state of each task
  - Blaze Job Monitor URL
  - Number of total, succeeded, and failed/cancelled tasklets
  - Number of processed and rejected rows for sources and targets
  - Data errors, if any, for transformations in each executed segment
- The LDTM log includes the following transformation statistics:
  - Number of output rows for sources and targets
  - Number of error rows for sources and targets
- The session log also displays a list of all segments within the grid task with corresponding links to the Blaze Job Monitor. Click on a link to see the execution details of that segment.

For more information, see the "Monitoring Mappings in a Hadoop Environment" chapter in the *Informatica Big Data Management 10.1.1 User Guide*.

Installation and Configuration

This section describes new features related to big data installation and configuration.

Address Reference Data Installation

Effective in version 10.1.1, Informatica Big Data Management installs with a shell script that you can use to install address reference data files. The script installs the reference data files on the compute nodes that you specify.

When you run an address validation mapping in a Hadoop environment, the reference data files must reside on each compute node on which the mapping runs. Use the script to install the reference data files on multiple nodes in a single operation.

The shell script name is `copyRefDataToComputeNodes.sh`.

Find the script in the following directory in the Informatica Big Data Management installation:

```
[Informatica installation directory]/tools/dq/av
```

When you run the script, you can enter the following information:

- The current location of the reference data files.
- The directory to which the script installs the files.
- The location of the file that contains the compute node names.
- The user name of the user who runs the script.

If you do not enter the information, the script uses a series of default values to identify the file locations and the user name.

For more information, see the *Informatica Big Data Management 10.1.1 Installation and Configuration Guide*.

Hadoop Configuration Manager in Silent Mode

Effective in version 10.1.1, you can use the Hadoop Configuration Manager in silent mode to configure Big Data Mangement.

For more information about configuring Big Data Management in silent mode, see the *Informatica Big Data Management 10.1.1 Installation and Configuration Guide*.

Installation in an Ambari Stack

Effective in version 10.1.1, you can use the Ambari configuration manager to install Big Data Management as a service in an Ambari stack.

For more information about installing Big Data Management in an Ambari stack, see the *Informatica 10.1.1 Big Data Management Installation and Configuration Guide*.

Script to Populate HDFS in HDInsight Clusters

Effective in version 10.1.1, you can use a script to populate the HDFS file system on an Azure HDInsight cluster when you configure the cluster for Big Data Management.

For more information about using the script to populate the HDFS file system, see the *Informatica Big Data Management 10.1.1 Installation and Configuration Guide*.

Spark Engine

Effective in version 10.1.1, the Spark engine has the following new features:

Binary Data Types

Effective in version 10.1.1, the Spark engine supports binary data type for the following functions:

- DEC_BASE64
- ENC_BASE64
- MD5
- UUID4
- UUID_UNPARSE
- CRC32
- COMPRESS
- DECOMPRESS (ignores precision)
- AES Encrypt
- AES Decrypt

Note: The Spark engine does not support binary data type for the join and lookup conditions.

For more information, see the "Function Reference" chapter in the *Informatica Big Data Management 10.1.1 User Guide*.

Transformation Support on the Spark Engine

Effective in version 10.1.1, transformations have the following additional support on the Spark engine:

- The Java transformation is supported with some restrictions.
- The Lookup transformation can access a Hive lookup source.

For more information, see the "Mapping Objects in the Hadoop Environment" chapter in the *Informatica Big Data Management 10.1.1 User Guide*.

Run-time Statistics for Spark Engine Job Runs

Effective in version 10.1.1, you can view summary and detailed statistics for mapping jobs run on the Spark engine.

You can view the following Spark summary statistics in the **Summary Statistics** view:

- Source. The name of the mapping source file.
- Target. The name of the target file.
- Rows. The number of rows read for source and target.

The **Detailed Statistics** view displays a graph of the row counts for Spark engine job runs.

For more information, see the "Mapping Objects in the Hadoop Environment" chapter in the *Informatica Big Data Management 10.1.1 User Guide*.

Security

This section describes new big data security features in version 10.1.1.

Fine-Grained SQL Authorization Support for Hive Sources

Effective in version 10.1.1, you can configure a Hive connection to observe fine-grained SQL authorization when a Hive source table uses this level of authorization. Enable the **Observe Fine Grained SQL Authorization** option in the Hive connection to observe row and column-level restrictions that are configured for Hive tables and views.

For more information, see the Authorization section in the "Introduction to Big Data Management Security" chapter of the *Informatica 10.1.1 Big Data Management Security Guide*.

Spark Engine Security Support

Effective in version 10.1.1, the Spark engine supports the following additional security systems:

- Apache Sentry on Cloudera CDH clusters
- Apache Ranger on Hortonworks HDP clusters
- HDFS Transparent Encryption on Hadoop distributions that the Spark engine supports
- Operating system profiles on Hadoop distributions that the Spark engine supports

For more information, see the "Introduction to Big Data Management Security" chapter in the *Informatica Big Data Management 10.1.1 Security Guide*.

Sqoop

Effective in version 10.1.1, you can use the following new features when you configure Sqoop:

- You can run Sqoop mappings on the Blaze engine.
- You can run Sqoop mappings on the Spark engine to read data from or write data to Oracle databases.
- When you run Sqoop mappings on the Blaze and Spark engines, you can configure partitioning. You can also run the mappings on a Hadoop cluster that uses Kerberos authentication.
- When you run Sqoop mappings on the Blaze engine to read data from or write data to Teradata, you can use the following specialized connectors:
  - Cloudera Connector Powered by Teradata
  - Hortonworks Connector for Teradata

  These specialized connectors use native protocols to connect to the Teradata database.

For more information, see the *Informatica 10.1.1 Big Data Management User Guide*.

Business Glossary

This section describes new Business Glossary features in version 10.1.1.

Export Rich Text as Plain Text

Effective in version 10.1.1, you can export rich text glossary content as plain text. The export option is available in the glossary export wizard and in the command line program.

For more information, see the "Glossary Administration " chapter in the *Informatica 10.1.1 Business Glossary Guide*.

Include Rich Text Content for Conflicting Assets

Effective in version 10.1.1, you can choose to import properties that are formatted as rich text or are of long string data type, from the import file, when the Analyst tool detects conflicting assets.

The import option is available in the glossary import wizard and in the command line program.

For more information, see the "Glossary Administration" chapter in the *Informatica 10.1.1 Business Glossary Guide*.

Command Line Programs

This section describes new commands in version 10.1.1.

infacmd as Commands

The following table describes new infacmd as commands:

Command	Description
CreateExceptionAuditTables	Creates the audit tables for the Human task instances that the Analyst Service specifies.
DeleteExceptionAuditTables	Deletes the audit tables for the Human task instances that the Analyst Service specifies.

The following table describes new options for infacmd as commands:

Command	Description
UpdateServiceOptions	- HumanTaskDataIntegrationService.exceptionDbName Identifies the database to store the audit trail tables for exception management tasks. - HumanTaskDataIntegrationService.exceptionSchemaName Identifies the schema to store the audit trail tables for exception management tasks.

For more information, see the "Infacmd as Command Reference" chapter in the *Informatica 10.1.1 Command Reference*.

infacmd dis command

The following table describes new infacmd dis command:

Command	Description
replaceMappingHadoopRuntimeConnections	Replaces the Hadoop connection of all mappings in deployed applications with another Hadoop connection. The Data Integration Service uses the Hadoop connection to connect to the Hadoop cluster to run mappings in the Hadoop environment.

For more information, see the "infacmd dis Command Reference" chapter in the *Informatica 10.1.1 Command Reference*.

infacmd mrs command

The following table describes new infacmd mrs command:

Command	Description
replaceMappingHadoopRuntimeConnections	Replaces the Hadoop connection of all mappings in the repository with another Hadoop connection. The Data Integration Service uses the Hadoop connection to connect to the Hadoop cluster to run mappings in the Hadoop environment.

For more information, see the "infacmd mrs Command Reference" chapter in the *Informatica 10.1.1 Command Reference*.

pmrep Commands

The following table describes an updated option for a pmrep command:

Command	Description
Validate	Contains the following updated option: -n (object_name). Required. Name of the object to validate. Do not use this option if you use the -i argument. When you validate a non-reusable session, include the workflow name. Enter the workflow name and the session name in the following format: \<workflow name>.\<session instance name> When you validate a non-reusable session in a non-reusable worklet, enter the workflow name, worklet name, and session name in the following format: \<workflow name>.\<worklet name>.\<session instance name>

For more information, see the "pmrep Command Reference" chapter in the *Informatica 10.1.1 Command Reference*.

Enterprise Information Catalog

This section describes new features in Enterprise Information Catalog version 10.1.1.

Business Glossary Integration

Effective in version 10.1.1, Analyst tool business glossaries are fully integrated with Enterprise Information Catalog.

You can perform the following tasks with business glossary assets:

View business glossary assets in the catalog.

> You can search for and view the full details for a business term, category, or policy in Enterprise Information Catalog. When you view the details for a business term, Enterprise Information Catalog also displays the glossary assets, technical assets, and other assets, such as Metadata Manager objects, that the term is related to.

> When you view a business glossary asset in the catalog, you can open the asset in the Analyst tool business glossary for further analysis.

Associate an asset with a business term.

> You can associate a business term with a technical asset to make an asset easier to understand and identify in the catalog. For example, you associate business term "Movie Details" with a relational table named "mv_dt." Enterprise Information Catalog displays the term "Movie Details" next to the asset name in the search results, in the Asset Details view, and optionally, in the lineage and impact diagram.

> When you associate a term with an asset, Enterprise Information Catalog provides intelligent recommendations for the association based on data domain discovery.

For more information about business glossary assets, see the "View Assets" chapter in the *Informatica 10.1.1 Enterprise Information Catalog User Guide*.

Column Similarity Profiling

Effective in version 10.1.1, you can configure and perform column similarity profiling. Column similarity profiling implies preparing metadata extracted from data sources for discovering similar columns in your enterprise data. You can then attach data domains to similar columns for faster and efficient searches for similar data in Enterprise Information Catalog.

Enterprise Information Catalog supports column similarity profiling for the following resource scanners:

- Amazon Redshift
- Amazon S3
- Salesforce
- HDFS
- Hive
- IBM DB2
- IBM DB2 for z/OS
- IBM Netezza
- JDBC
- Microsoft SQL Server
- Oracle
- Sybase
- Teradata
- SAP

Data Domains and Data Domain Groups

Effective in version 10.1.1, you can create data domains and data domain groups in Enterprise Information Catalog. You can group logical data domains in a data domain group.

A data domain is a predefined or user-defined Model repository object based on the semantics of column data or a column name. Examples include Social Security number, phone number, and credit card number.

You can create data domains based on data rules or column name rules defined in the Informatica Analyst Tool or the Informatica Developer Tool. Alternatively, you can create data domains based on existing columns in the catalog. You can define proximity rules to configure inference for new data domains from existing data domains configured in the catalog.

Lineage and Impact Analysis

Effective in version 10.1.1, lineage and impact diagrams have expanded functionality. The Lineage and Impact view also contains a tabular impact summary that lists the assets that impact and are impacted by the asset that you are studying.

The Lineage and Impact view has the following enhancements:

Diagram enhancements

The lineage and impact diagram has the following enhancements:

- By default, the lineage and impact diagram displays the origins, the asset that you are studying, and the destinations for the data. You can use the slider controls to reveal intermediate assets one at-a-time by distance from the seed asset or to fully expand the diagram. You can also expand all assets within a particular data flow path.

- You can display the child assets of the asset that you are studying, all the way down to the column or field level. When you drill-down on an asset, the diagram displays the child assets that you select and the assets to which the child assets are linked.

- You can display the business terms that are associated with the technical assets in the diagram.

- You can print the diagram and export it to a scalable vector graphics (.svg) file.

Impact analysis

When you open the Lineage and Impact view for an asset, you can switch from the diagram view to the tabular asset summary. The tabular asset summary lists all of the assets that impact and are impacted by the asset that you are studying. You can export the asset summary to a Microsoft Excel file to create reports or further analyze the data.

For more information about lineage and impact analysis, see the "View Lineage and Impact" chapter in the *Informatica 10.1.1 Enterprise Information Catalog User Guide*.

Permissions for Users and User Groups

Effective in version 10.1.1, you can configure permissions for users and user groups on resources configured in Enterprise Information Catalog. You can specify permissions to view the resource metadata in Enterprise Information Catalog or view and enrich the resource metadata in Enterprise Information Catalog. You can also deny permissions to view or enrich resource metadata in Enterprise Information Catalog for specific users and user groups.

New Resource Types

Effective in version 10.1.1, you can create resources for the following data source types:

Oracle Business Intelligence

Extract metadata from the Business intelligence tool from Oracle that includes analysis and reporting capabilities.

Informatica Master Data Management

Extract metadata about critical information within an organization from Informatica Master Data Management.

Microsoft SQL Server Integration Service

Extract metadata about data integration and workflow applications from Microsoft SQL Server Integration Service.

SAP

Extract metadata from SAP application platform that integrates multiple business applications and solutions.

Hive on Amazon Elastic MapReduce

Extract metadata from files in Amazon Elastic MapReduce using a Hive resource.

Hive on Azure HDInsight

Extract metadata from files in Azure HDInsight using a Hive resource.

Synonym Definition Files

Effective in version 10.1.1, you can upload synonym definition files to Enterprise Information Catalog. Synonym definition files include synonyms defined for table names, column names, data domains and other assets in the catalog. You can search for the assets in the Enterprise Information Catalog using the defined synonyms.

Universal Connectivity Framework

Effective in version 10.1.1, Enterprise Information Catalog introduces the Universal Connectivity Framework. Using the framework, you can build custom resources to extract metadata from a range of data sources supported by MITI.

Informatica Analyst

This section describes new Analyst tool features in version 10.1.1.

Profiles

This section describes new Analyst tool features for profiles and scorecards.

Drilldown on Scorecards

Effective in version 10.1.1, when you click a data series or data point in the scorecard dashboard, the scorecards that map to the data series or data point appears in the assets list pane.

For more information about scorecards, see the "Scorecards in Informatica Analyst" chapter in the *Informatica 10.1.1 Data Discovery Guide*.

Informatica Installation

This section describes new installation features in version 10.1.1.

Informatica Upgrade Advisor

Effective in version 10.1.1, you can run the Informatica Upgrade Advisor to check for conflicts and deprecated services in the domain before you perform an upgrade.

For more information about the upgrade advisor, see the *Informatica Upgrade Guides*.

Intelligent Data Lake

This section describes new Intelligent Data Lake features in version 10.1.1.

Data Preview for Tables in External Sources

Effective in version 10.1.1, you can preview sample data for external (outside Hadoop data lake) tables if these sources are cataloged. The administrator needs to configure JDBC connections with Sqoop and provide the analysts with requisite permissions. The analyst can connect to the data source using these connections to view the data from assets that are not in the data lake.

For more information, see the "Discover Data" chapter in the *10.1.1 Intelligent Data Lake User Guide*.

Importing Data From Tables in External Sources

Effective in version 10.1.1, you can import data from tables in external sources (outside Hadoop data lake), such as Oracle and Teradata, into the data lake if these sources are already cataloged. The administrator needs to configure JDBC connections with Sqoop to the external sources and provide access to the analyst. The analyst can use these connections to preview the data asset and import into the lake based on their needs.

For more information, see the "Discover Data" chapter in the *10.1.1 Intelligent Data Lake User Guide*.

Exporting Data to External Targets

Effective in version 10.1.1, you can export a data asset or a publication to external targets (outside Hadoop data lake), such as Oracle and Teradata. The administrator needs to configure the JDBC connections with Sqoop to the external sources and provide access to the analyst. The analyst can use these connections to export the data asset to the external database.

For more information, see the "Discover Data" chapter in the *10.1.1 Intelligent Data Lake User Guide*.

Configuring Sampling Criteria for Data Preparation

Effective in version 10.1.1, you can specify sampling criteria that best suits your needs for data preparation for a given data asset. You can choose to include only a few columns during preparation and filter the data, choose number of rows to sample, and select Random or First N rows as sample.

For more information, see the "Prepare Data" chapter in the *10.1.1 Intelligent Data Lake User Guide*.

Performing a Lookup on Worksheets

Effective in version 10.1.1, you can perform a lookup. Use the lookup function to lookup a key column in another sheet and fetch values in corresponding other columns in that looked up sheet.

For more information, see the "Prepare Data" chapter in the *10.1.1 Intelligent Data Lake User Guide*.

Downloading as a TDE File

Effective in version 10.1.1, you can download data in data lake assets as a TDE file. You can directly open the downloaded file in Tableau. You can search for any data asset and download it as a CSV file or TDE file.

For more information, see the "Discover Data" chapter in the *10.1.1 Intelligent Data Lake User Guide*.

Sentry and Ranger Support

Effective in version 10.1.1, Intelligent Data Lake supports Sentry and Ranger on Cloudera and Hortonworks. Ranger and Sentry offer a centralized security framework to manage granular level access control on Cloudera and Hortonworks. You can create authorization rules or policies to control the access of data. Sentry and Ranger support SQL-based authorization for data lake assets.

Mappings

This section describes new mapping features in version 10.1.1.

Informatica Mappings

This section describes new Informatica mappings features in version 10.1.1.

Export Parameters to a Parameter File

You can export a mapping parameter file or a workflow parameter file from the Developer tool. You can export a parameter file that contains mapping parameters or workflow parameters that you define in the Developer tool. The Developer tool creates a parameter file in .xml format. Export parameters from the mapping **Parameters** tab or from the workflow **Parameters** tab. Use the parameter file when you run deployed mappings or workflows.

For more information, see the "Mapping Parameters" chapter in the *Informatica Developer 10.1.1 Mapping Guide* or the "Workflow Parameters" chapter in the *Informatica Developer 10.1.1 Workflow Guide*.

Metadata Manager

This section describes new Metadata Manager features in version 10.1.1.

Dataset Extraction for Cloudera Navigator Resources

Effective in version 10.1.1, Metadata Manager can extract HDFS datasets from Cloudera Navigator. Metadata Manager displays the datasets in the metadata catalog within the HDFS Datasets logical group.

For more information about Cloudera Navigator resources, see the "Database Management Resources" chapter in the *Informatica 10.1.1 Metadata Manager Administrator Guide*.

Mapping Extraction for Informatica Platform Resources

Effective in version 10.1.1, Informatica Platform resources can extract metadata for mappings in deployed workflows.

Informatica Platform resources that are based on version 10.1.1 applications can extract metadata for mappings in deployed workflows in addition to mappings that are deployed directly to the application.

When Metadata Manager extracts a mapping in a deployed workflow, it adds the workflow name and Mapping task name to the mapping name as a prefix. Metadata Manager displays the mapping in the metadata catalog within the Mappings logical group.

For more information about Informatica Platform resources, see the "Data Integration Resources" chapter in the *Informatica 10.1.1 Metadata Manager Administrator Guide*.

PowerExchange Adapters

This section describes new PowerExchange adapter features in version 10.1.1

PowerExchange® Adapters for Informatica

This section describes new Informatica adapter features in version 10.1.1.

PowerExchange for Amazon Redshift

Effective in version 10.1.1, you can enable PowerExchange for Amazon Redshift to run a mapping on the Blaze engine. When you run the mapping, the Data Integration Service pushes the mapping to a Hadoop cluster and processes the mapping on the Blaze engine, which significantly increases the performance.

For more information, see the *Informatica PowerExchange for Amazon Redshift 10.1.1 User Guide*.

PowerExchange for Cassandra

Effective in version 10.1.1, PowerExchange for Cassandra supports the following features:

- You can use the following advanced ODBC driver configurations with PowerExchange for Cassandra:
  - Load balancing policy. Determines how the queries are distributed to nodes in a Cassandra cluster based on the specified DC Aware or Round-Robin policy.
  - Filtering. Limits the connections of the drivers to a predefined set of hosts.
- You can enable the following arguments in the ODBC driver to optimize the performance:
  - Token Aware. Improves the query latency and reduces load on the Cassandra node.
  - Latency Aware. Ignores the slow performing Cassandra nodes while sending queries.
  - Null Value Insertion. Enables you to specify null values in an INSERT statement.
  - Case Sensitive. Enables you to specify schema, table, and column names in a case-sensitive fashion.
- You can process Cassandra sources and targets that contain the date, smallint, and tinyint data types

For more information, see the *Informatica PowerExchange for Cassandra 10.1.1 User Guide*.

PowerExchange for HBase

Effective in version 10.1.1, you can enable PowerExchange for HBase to run a mapping on a Blaze or Spark engine. When you run the mapping, the Data Integration Service pushes the mapping to a Hadoop cluster and processes the mapping on the selected engine, which significantly increases the performance.

For more information, see the *Informatica PowerExchange for HBase 10.1.1 User Guide*.

PowerExchange for Hive

Effective in version 10.1.1, you can configure the Lookup transformation on Hive data objects in mappings in the native environment.

For more information, see the *Informatica PowerExchange for Hive 10.1.1 User Guide*.

PowerExchange Adapters for PowerCenter®

This section describes new PowerCenter adapter features in version 10.1.1.

PowerExchange for Amazon Redshift

Effective in version 10.1.1, you can perform the following tasks with PowerExchange for Amazon Redshift:

- You can configure partitioning for Amazon Redshift sources and targets. You can configure the partition information so that the PowerCenter Integration Service determines the number of partitions to create at run time.
- You can include a Pipeline Lookup transformation in a mapping.
- The PowerCenter Integration Service can push expression, aggregator, operator, union, sorter, and filter functions to Amazon Redshift sources and targets when the connection type is ODBC and the ODBC Subtype is selected as Redshift.
- You can configure advanced filter properties in a mapping.
- You can configure pre-SQL and post-SQL queries for source and target objects in a mapping.
- You can configure a Source transformation to select distinct rows from the Amazon Redshift table and sort data.
- You can parameterize source and target table names to override the table name in a mapping.
- You can define an SQL query for source and target objects in a mapping to override the default query. You can enter an SQL statement supported by the Amazon Redshift database.

For more information, see the *Informatica 10.1.1 PowerExchange for Amazon Redshift User Guide for PowerCenter*.

PowerExchange for Cassandra

Effective in version 10.1.1, PowerExchange for Cassandra supports the following features:

- You can use the following advanced ODBC driver configurations with PowerExchange for Cassandra:
  - Load balancing policy. Determines how the queries are distributed to nodes in a Cassandra cluster based on the specified DC Aware or Round-Robin policy.
  - Filtering. Limits the connections of the drivers to a predefined set of hosts.
- You can enable the following arguments in the ODBC driver to optimize the performance:
  - Token Aware. Improves the query latency and reduces load on the Cassandra node.
  - Latency Aware. Ignores the slow performing Cassandra nodes while sending queries.
  - Null Value Insertion. Enables you to specify null values in an INSERT statement.
  - Case Sensitive. Enables you to specify schema, table, and column names in a case-sensitive fashion.
- You can process Cassandra sources and targets that contain the date, smallint, and tinyint data types.

For more information, see the *Informatica PowerExchange for Cassandra 10.1.1 User Guide for PowerCenter*.

PowerExchange for Vertica

Effective in version 10.1.1, PowerExchange for Vertica supports compressing data in GZIP format. When you use bulk mode to write large volumes of data to a Vertica target, you can configure the session to create a staging file. On UNIX operating systems, when you enable file staging, you can also compress the data in a GZIP format. By compressing the data, you can reduce the size of data that is transferred over the network and improve session performance.

To compress data, you must re-register the PowerExchange for Vertica plug-in with the PowerCenter repository.

For more information, see the *Informatica PowerExchange for Vertica 10.1.1 User Guide for PowerCenter*.

Security

This section describes new security features in version 10.1.1.

Custom Kerberos Libraries

Effective in version 10.1.1, you can configure custom or native database clients and Informatica processes within an Informatica domain to use custom Kerberos libraries instead of the default Kerberos libraries that Informatica uses.

For more information, see the "Kerberos Authentication Setup" chapter in the *Informatica 10.1.1 Security Guide*.

Scheduler Service Support in Kerberos-Enabled Domains

Effective in version 10.1.1, you can use the Scheduler Service to run mappings, workflows, profiles and scorecards in a domain that uses Kerberos authentication.

Single Sign-on for Informatica Web Applications

Effective in version 10.1.1, you can configure single sign-on (SSO) using Security Assertion Markup Language (SAML) to log into the Administrator tool, the Analyst tool and the Monitoring tool.

Security Assertion Markup Language is an XML-based data format for exchanging authentication and authorization information between a service provider and an identity provider. In an Informatica domain, the Informatica web application is the service provider. Microsoft Active Directory Federation Services (AD FS) 2.0 is the identity provider, which authenticates web application users with your organization's LDAP or Active Directory identity store.

For more information, see the "Single Sign-on for Informatica Web Applications" chapter in the *Informatica 10.1.1 Security Guide*.

Transformations

This section describes new transformation features in version 10.1.1.

Informatica Transformations

This section describes new features in Informatica transformations in version 10.1.1.

Address Validator Transformation

This section describes the new Address Validator transformation features.

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

All Countries

Effective in version 10.1.1, you can add the Count Number port to an output address. The Count Number port value indicates the position of each address in a set of suggestions that the transformation returns in interactive mode or suggestion list mode.

For example, the Count Number port returns the number 1 for the first address in the set. The port returns the number 2 for the second address in the set. The number increments by 1 for each address that address validation returns.

Find the Count Number port in the Status Info port group.

China

Multi-language address parsing and verification

Effective in version 10.1.1, you can configure the Address Validator transformation to return the street descriptor and street directional information in a valid China address in a transliterated Latin script (Pinyin) or in English. The transformation returns the other elements in the address in the Hanzi script.

To specify the output language, set the Preferred Language advanced property on the transformation.

Single-line verification of China addresses in suggestion list mode

Effective in version 10.1.1, you can configure the Address Validator transformation to return valid suggestions for a China address that you enter on a single line in fast completion mode. To enter an address on a single line, select a Complete Address port from the Multiline port group. Enter the address in the Hanzi script.

When you enter a partial address, the transformation returns one or more address suggestions for the address that you enter. When you enter a complete valid address, the transformation returns the valid version of the address from the reference database.

Ireland

Multi-language address parsing and verification

Effective in version 10.1.1, you can configure the Address Validator transformation to read and write the street, locality, and county information for an Ireland address in the Irish language.

An Post, the Irish postal service, maintains the Irish-language information in addition to the English-language addresses. You can include Irish-language street, locality, and county information in an input address and retrieve the valid English-language version of the address. You can enter an English-language address and retrieve an address that includes the street, locality, and county information in the Irish language. Address validation returns all other information in English.

To specify the output language, set the Preferred Language advanced property on the transformation.

Rooftop geocoordinates in Ireland addresses

Effective in version 10.1.1, you can configure the Address Validator transformation to return rooftop geocoordinates for an address in Ireland.

To return the geocoordinates, add the Geocoding Complete port to the output address. Find the Geocoding Complete port in the Geocoding port group. To specify Rooftop geocoordinates, set the Geocode Data Type advanced property on the transformation.

Support for preferred descriptors in Ireland addresses

Effective in version 10.1.1, you can configure the Address Validator transformation to return the short or long forms of the following elements in the English language:

- Street descriptors

- Directional values

To specify a preference for the elements, set the Global Preferred Descriptor advanced property on the transformation,

Note: The Address Validator transformation writes all street information to the street name field in an Irish-language address.

Italy

Effective in version 10.1.1, you can configure the Address Validator transformation to add the ISTAT code to a valid Italy address. The ISTAT code contains characters that identify the province, municipality, and region to which the address belongs. The Italian National Institute of Statistics (ISTAT) maintains the ISTAT codes.

To add the ISTAT code to an address, select the ISTAT Code port. Find the ISTAT Code port in the IT Supplementary port group.

Japan

Geocoding enrichment for Japan addresses

Effective in version 10.1.1, you can configure the Address Validator transformation to return standard geocoordinates for addresses in Japan.

The transformation can return geocoordinates at multiple levels of accuracy. When a valid address contains information to the Ban level, the transformation returns house number-level geocoordinates. When a valid address contains information to the Chome level, the transformation returns street-level geocoordinates. If an address does not contain Ban or Chome information, Address Verification returns locality-level geocoordinates.

To return the geocoordinates, add the Geocoding Complete port to the output address. Find the Geocoding Complete port in the Geocoding port group.

Single-line verification of Japan addresses in suggestion list mode

Effective in version 10.1.1, you can configure the Address Validator transformation to return valid suggestions for a Japan address that you enter on a single line in suggestion list mode. You can retrieve suggestions for an address that you enter in the Kanji script or the Kana script. To enter an address on a single line, select a Complete Address port from the Multiline port group.

When you enter a partial address, the transformation returns one or more address suggestions for the address that you enter. When you enter a complete valid address, the transformation returns the valid version of the address from the reference database.

South Korea

Support for Revised Romanization transliteration in South Korea addresses

Effective in version 10.1.1, the Address Validator transformation can use the Revised Romanization system to transliterate an address between Hangul and Latin character sets. To specify a character set for output addresses from South Korea, use the Preferred Script advanced property.

Updates to post code verification in South Korea addresses

Effective in version 10.1.1, the Address Validator transformation adds a five-digit post code to a fully valid input address that does not include a post code. The five-digit post code represents the current post code format in use in South Korea. The transformation can add the five-digit post code to a fully valid lot-based address and a fully valid street-based address.

To verify addresses in the older, lot-based format, use the Matching Extended Archive advanced property.

Spain

Effective in version 10.1.1, you can configure the Address Validator transformation to add the INE code to a valid Spain address. The INE code contains characters that identify the province, municipality, and street in the address. The National Institute of Statistics (INE) in Spain maintains the INE codes.

To add an INE code to an address, select one or more of the following ports:

- INE Municipality Code
- INE Province Code
- INE Street Code

Find the INE Code ports in the ES Supplementary port group.

United States

Support for CASS Cycle O requirements

Effective in version 10.1.1, the Address Validator transformation adds features that support the proposed requirements of the Coding Accuracy Support System (CASS) Cycle O standard.

To prepare for the Cycle O standard, the transformation includes the following features:

- Private mailbox and commercial mail receiving agency identification

  The United States Postal Service updates the CASS requirements for private mailbox (PMB) addresses and commercial mail receiving agency (CMRA) addresses in Cycle O. To meet the Cycle O standard, the Address Validator transformation adds PMB as a prefix before a private mailbox number in a CMRA address. If a pound sign (#) precedes a private mailbox number in the address, the transformation converts the pound sign to PMB. To comply with the Cycle O standard, the transformation does not use the PMB number to verify Delivery Point Validation (DPV) data for an address.

- DPV PBSA Indicator port for post office box street address (PBSA) identification

  The United States Postal Service can recognize post office box addresses in a street address format. To identify PBSA addresses in an address set, use the DPV PBSA Indicator port. Find the DPV PBSA Indicator port in the US Specific port group.

  For example, the following address identifies post office box number 3094 at a post office on South Center Street:

  ```
  131 S Center St Unit 3094
  Collierville TN 38027-0419
  ```

- DPV ZIP Code Validation port for Form 3553 completion

  The DPV ZIP Code Validation port indicates whether an address is valid for inclusion in the total address count on CASS Form 3553. If an address passes delivery point validation but does not include a deliverable ZIP+4 Code, you cannot include the address in the total address count. Find the DPV ZIP Code Validation port in the US Specific port group.

Improved parsing of non-standard first-line data in United States addresses

Effective in version 10.1.1, the Address Validation transformation parses non-standard mailbox data into sub-building elements. The non-standard data might identify a college campus mailbox or a courtroom at a courthouse.

Support for global preferred descriptors in United States addresses

Effective in version 10.1.1, you can return the short or long forms of the following elements in a United States address:

- Street descriptors

- Directional values
- Sub-building descriptors

To specify the format of the elements that the transformation returns, set the Global Preferred Descriptor advanced property on the transformation.

For more information, see the *Informatica 10.1.1 Developer Transformation Guide* and the *Informatica 10.1.1 Address Validator Port Reference*.

Write Transformation

Effective in version 10.1.1, when you create a Write transformation from an existing transformation in a mapping, you can specify the type of link for the input ports of the Write transformation.

You can link ports by name. Also, in a dynamic mapping, you can link ports by name, create a dynamic port based on a mapping flow, or link ports at run time based on a link policy.

For more information, see the "Write Transformation" chapter in the *Informatica 10.1.1 Developer Transformation Guide*.

Web Services

This section describes new web services features in version 10.1.1.

Informatica Web Services

This section describes new Informatica web service features in version 10.1.1.

REST Web Services

You can create an Informatica REST web service that returns data to a web service client in JSON or XML format.

An Informatica REST web service is a web service that receives an HTTP request to perform a GET operation. A GET operation retrieves data. The REST request is a simple URI string from an internet browser. The client limits the web service output data by adding filter parameters to the URI.

Define a REST web service resource in the Developer tool. A REST web service resource contains the definition of the REST web service response message and the mapping that returns the response. When you create an Informatica REST web service, you can define the resource from a data object or you can manually define the resource.

Workflows

This section describes new workflow features in version 10.1.1.

Informatica Workflows

This section describes new features in Informatica workflows in version 10.1.1.

Terminate Event

Effective in version 10.1.1, you can add a Terminate event to a workflow. A Terminate event defines a point before the End event at which the workflow can end. A workflow can contain one or more Terminate events.

A workflow terminates if you connect a task or a gateway to a Terminate event and the task output satisfies a condition on the sequence flow. The Terminate event aborts the workflow before any further task in the workflow can run.

Add a Terminate event to a workflow if the workflow data can reach a point at which there is no need to run additional tasks. For example, you might add a Terminate event to end a workflow that contains a Mapping task and a Human task. Connect the Mapping task to an Exclusive gateway, and then connect the gateway to a Human task and to a Terminate event. If the Mapping task generates exception record data for the Human task, the workflow follows the sequence flow to the Human task. If the Mapping task does not generate exception record data, the workflow follows the sequence flow to the Terminate event.

For more information, see the *Informatica 10.1.1 Developer Workflow Guide*.

User Permissions on Human Tasks

Effective in version 10.1.1, you can set user permissions on Human task data. The permissions specify the data that users can view and the types of action that users can perform in Human task instances in the Analyst tool. You can set the permissions within a step in a Human task when you design a workflow. The permissions apply to all users who can view or edit a task instance that the step defines.

By default, Analyst tool users can view all data and perform any action in the task instances that they work on.

You can set viewing permissions and editing permissions. The viewing permissions define the data that the Analyst tool displays for the task instances that the step defines. The editing permissions define the actions that users can take to update the task instance data. Viewing permissions take precedence over editing permissions. If you grant editing permissions on a column and you do not grant viewing permissions on the column, Analyst tool users cannot edit the column data.

For more information, see the *Informatica 10.1.1 Developer Workflow Guide*.

Workflow Variables in Human Task Instance Notifications

Effective in version 10.1.1, you can use workflow variables to write information about a Human task instance to an email notification. The variables record information about the task instance when a user completes, escalates, or reassigns a task instance.

To display the list of variables, open the Human task and select the step that defines the Human task instances. On the **Notifications** view, select the message body of the email notification and press the $+CTRL+SPACE keys.

The notification can display the following variables:

$taskEvent.eventTime

The time that the workflow engine performs the user instruction to escalate, reassign, or complete the task instance.

$taskEvent.startOwner

The owner of the task instance at the time that the workflow engine escalates or completes the task. Or, the owner of the task instance after the engine reassigns the task instance.

$taskEvent.status

The task instance status after the engine performs the user instruction to escalate, reassign, or complete the task instance. The status names are READY and IN_PROGRESS.

$taskEvent.taskEventType

The type of instruction that the engine performs. The variable values are escalate, reassign, and complete.

$taskEvent.taskId

The task instance identifier that the Analyst tool displays.

For more information, see the *Informatica 10.1.1 Developer Workflow Guide*.

CHAPTER 13

Changes (10.1.1)

This chapter includes the following topics:

Support Changes

This section describes support changes in version 10.1.1 HotFix 2.

Big Data Management Hive Engine

Effective in version 10.1.1, Informatica dropped support for HiveServer2 which the Hive engine uses to run mappings.

Previously, the Hive engine supported the Hive driver and HiveServer2 to run mappings in the Hadoop environment. HiveServer2 and the Hive driver convert HiveQL queries to MapReduce or Tez jobs that are processed on the Hadoop cluster.

If you install Big Data Management 10.1.1 or upgrade to version 10.1.1, the Hive engine uses the Hive driver when you run the mappings. The Hive engine no longer supports HiveServer2 to run mappings in the Hadoop environment. Hive sources and targets that use the HiveServer2 service on the Hadoop cluster are still supported.

To run mappings in the Hadoop environment, Informatica recommends that you select all run-time engines. The Data Integration Service uses a proprietary rule-based methodology to determine the best engine to run the mapping.

For information about configuring the run-time engines for your Hadoop distribution, see the *Informatica Big Data Management 10.1.1 Installation and Configuration Guide*. For information about mapping objects that the run-time engines support, see the *Informatica Big Data Management 10.1.1 User Guide*.

Support Changes - Big Data Management Hadoop Distributions

The following table lists the supported Hadoop distribution versions and changes in Big Data Management 10.1.1:

At release date, version 10.1.1 supports the following Hadoop distributions:

- Azure HDInsight v. 3.4
- Cloudera CDH v. 5.8
- IBM BigInsights v. 4.2
- Hortonworks HDP v. 2.5
- Amazon EMR v. 5.0

To see a list of the latest supported versions, see the Product Availability Matrix on the Informatica Customer Portal: `https://network.informatica.com/community/informatica-network/product-availability-matrices`.

MapR Support

Effective in version 10.1.1, Informatica deferred support for Big Data Management on a MapR cluster. To run mappings on a MapR cluster, use Big Data Management 10.1. Informatica plans to reinstate support in a future release.

Some references to MapR remain in documentation in the form of examples. Apply the structure of these examples to your Hadoop distribution.

Amazon EMR Support

Effective in version 10.1.1, you can install Big Data Management in the Amazon EMR environment. You can choose from the following installation methods:

- Download and install from an RPM package. When you install Big Data Management in an Amazon EMR environment, you install Big Data Management elements on a local machine to run the Model Repository Service, Data Integration Service, and other services.
- Install an Informatica instance in the Amazon cloud environment. When you create an implementation of Big Data Management in the Amazon cloud, you bring online virtual machines where you install and run Big Data Management.

For more information about installing and configuring Big Data Management on Amazon EMR, see the *Informatica Big Data Management 10.1.1 Installation and Configuration Guide*.

Big Data Management Spark Support

Effective in version 10.1.1, you can configure the Spark engine on all supported Hadoop distributions. You can configure Big Data Management to use one of the following Spark versions based on the Hadoop distribution that you use:

- Cloudera Spark 1.6 and Apache Spark 2.0.1 for Cloudera cdh5u8 distribution.
- Apache Spark 2.0.1 for all Hadoop distributions.

For more information, see the *Informatica Big Data Management 10.1.1 Installation and Configuration Guide*.

Data Analyzer

Effective in version 10.1.1, Informatica dropped support for Data Analyzer. Informatica recommends that you use a third-party reporting tool to run PowerCenter and Metadata Manager reports. You can use the recommended SQL queries for building all the reports shipped with earlier versions of PowerCenter.

Operating System

Effective in version 10.1.1, Informatica added support for the following operating systems:

Solaris 11

Windows 10 for Informatica Clients

PowerExchange for SAP NetWeaver

Effective in version 10.1.1, Informatica implemented the following changes in PowerExchange for SAP NetWeaver support:

Support Change	Level of Support	Comments
Analytic Business Components	Dropped support	Effective in version 10.1.1, Informatica dropped support for the Analytic Business Components (ABC) functionality. You cannot use objects in the ABC repository to read and transform SAP data. Informatica will not ship the ABC transport files.
SAP R/3 version 4.7	Dropped support	Effective in version 10.1.1, Informatica dropped support for SAP R/3 4.7 systems. Upgrade to SAP ECC version 5.0 or later.

Reporting and Dashboards Service

Effective in version 10.1.1, Informatica dropped support for the Reporting and Dashboards Service. Informatica recommends that you use a third-party reporting tool to run PowerCenter and Metadata Manager reports. You can use the recommended SQL queries for building all the reports shipped with earlier versions of PowerCenter.

Reporting Service

Effective in version 10.1.1, Informatica dropped support for the Reporting Service. Informatica recommends that you use a third-party reporting tool to run PowerCenter and Metadata Manager reports. You can use the recommended SQL queries for building all the reports shipped with earlier versions of PowerCenter.

Big Data

This section describes the changes to big data in version 10.1.1.

Functions Supported in the Hadoop Environment

Effective in 10.1.1, the following support changes affect functions in the Hadoop environment:

Function	Description	Changes
AES_DECRYPT	Returns decrypted data to string format.	Supported on the Spark engine. Previously supported only on the Blaze and Hive engines.
AES_ENCRYPT	Returns data in encrypted format.	Supported on the Spark engine. Previously supported only on the Blaze and Hive engines.
COMPRESS	Compresses data using the zlib 1.2.1 compression algorithm.	Supported on the Spark engine. Previously supported only on the Blaze and Hive engines.
CRC32	Returns a 32-bit Cyclic Redundancy Check (CRC32) value.	Supported on the Spark engine. Previously supported only on the Blaze and Hive engines.
DECOMPRESS	Decompresses data using the zlib 1.2.1 compression algorithm.	Supported with restrictions on the Spark engine. Previously supported only on the Blaze and Hive engines.
DEC_BASE64	Decodes a base 64 encoded value and returns a string with the binary data representation of the data.	Supported on the Spark engine. Previously supported only on the Blaze and Hive engines.
ENC_BASE64	Encodes data by converting binary data to string data using Multipurpose Internet Mail Extensions (MIME) encoding.	Supported on the Spark engine. Previously supported only on the Blaze and Hive engines.
MD5	Calculates the checksum of the input value. The function uses Message-Digest algorithm 5 (MD5).	Supported on the Spark engine. Previously supported only on the Blaze and Hive engines.
UUID4	Returns a randomly generated 16-byte binary value that complies with variant 4 of the UUID specification described in RFC 4122.	Supported on the Spark engine without restrictions. Previously supported on the Blaze engine without restrictions and on the Spark and Hive engines with restrictions.
UUID_UNPARSE	Converts a 16-byte binary value to a 36-character string representation as specified in RFC 4122.	Supported on the Spark engine without restrictions. Previously supported on the Blaze engine without restrictions and on the Spark and Hive engines with restrictions.

Hadoop Configuration Manager

Effective in version 10.1.1, the Big Data Management Configuration Utility has the following changes:

- The utility is renamed to the Hadoop Configuration Manager.

- The Hadoop Configuration Manager supports configuring Big Data Management on Azure HDInsight clusters in addition to other Hadoop clusters.

For more information about the Hadoop Configuration Manager, see the *Informatica Big Data Management 10.1.1 Installation and Configuration Guide*.

Business Glossary

This section describes the changes to Business Glossary in version 10.1.1

Export File Restriction

Effective in version 10.1.1, the Business Glossary export in the Analyst tool and command line has the following changed behavior:

Truncation of characters in a Microsoft Excel export file cell

When you export Glossary assets that contain more than 32,767 characters in one Microsoft Excel cell, the Analyst tool automatically truncates the characters in the cell to a value lesser than 32,763.

Microsoft Excel supports only up to 32,767 characters in a cell. Previously, when you exported a glossary, Microsoft Excel truncated long text properties that contained more than 32,767 characters in a cell, causing loss of data without any warning.

For more information about Export and Import, see the "Glossary Administration" chapter in the *Informatica 10.1.1 Business Glossary Guide*.

Data Integration Service

This section describes changes to the Data Integration Service in version 10.1.1.

Execution Options in the Data Integration Properties

Effective in version 10.1.1, you no longer need to restart the Data Integration Service when you edit the following Data Integration Services properties:

- Cache Directory
- Home Directory
- Maximum Parallelism
- Rejected Files Directory
- Source Directory
- State Store
- Target Directory
- Temporary Directories

Previously, you had to restart the Data Integration Service when you edited these properties.

Data Types

This section describes changes to data types in version 10.1.1.

Informatica Data Types

This section describes changes to transformation data types in the Developer tool.

Double Data Type

Effective in version 10.1.1, you can edit the precision and scale for double data types. The scale must be less than or equal to the precision.

Previously, the precision was set to 15 and the scale was set to 0.

For more information, see the "Data Type Reference" appendix in the *Informatica 10.1.1 Developer Tool Guide*.

Informatica Analyst

This section describes changes to the Analyst tool in version 10.1.1.

Profiles

This section describes new Analyst tool features for profiles.

Run-time Environment

Effective in version 10.1.1, after you choose the Hive option as the run-time environment, select a Hadoop connection to run the profiles.

Previously, after you choose the Hive option as the run-time environment, you selected a Hive connection to run the profiles.

For more information about run-time environment, see the "Column Profiles in Informatica Analyst" chapter in the *Informatica 10.1.1 Data Discovery Guide*.

Informatica Developer

This section describes changes to the Developer tool in version 10.1.1.

Profiles

This section describes new Developer tool features for profiles.

Run-time Environment

Effective in version 10.1.1, after you choose the Hive option as the run-time environment, select a Hadoop connection to run the profiles.

Previously, after you choose the Hive option as the run-time environment, you selected a Hive connection to run the profiles.

For more information about run-time environment, see the "Data Object Profiles" chapter in the *Informatica 10.1.1 Data Discovery Guide*.

Mappings

This section describes changes to mappings in version 10.1.1.

Informatica Mappings

This section describes the changes to the Informatica mappings in version 10.1.1.

Reorder Generated Ports in a Dynamic Port

Effective in version 10.1.1, you can change the order of generated ports based on the following options:

- The order of ports in the group or dynamic port of the upstream transformation.
- The order of input rules for the dynamic port.
- The order of ports in the nearest transformation with static ports.

Default is to reorder based on the ports in the upstream transformation.

Previously, you could reorder generated ports based on the order of input rules for the dynamic port.

For more information, see the "Dynamic Mappings" chapter in the *Informatica 10.1.1 Developer Mapping Guide*.

Enterprise information Catalog

This section describes changes to Enterprise Information Catalog in version 10.1.1.

HDFS Scanner Enhancement

Effective in version 10.1.1, you can extract metadata from flat file types using the HDFS resource scanner.

Relationships View

Effective in version 10.1.1, you can view business terms, related glossary assets, related technical assets, and similar columns for the selected asset.

Previously, you could view asset relationships such as columns, data domains, tables, and views.

For more information about relationships view, see the "View Relationships" chapter in the *Informatica 10.1.1 Enterprise Information Catalog User Guide*.

Metadata Manager

This section describes changes to Metadata Manager in version 10.1.1.

Cloudera Navigator Resources

Effective in version 10.1.1, Cloudera Navigator resources have the following behavior changes:

Incremental loading changes

Incremental loading for Cloudera Navigator resources is disabled by default. Previously, incremental loading was enabled by default.

When incremental loading is enabled, Metadata Manager performs a full metadata load when the Cloudera administrator invokes a purge operation in Cloudera Navigator after the last successful metadata load.

Additionally, there are new guidelines that explain when you might want to disable incremental loading.

Search query changes

You can use the search query to exclude entity types besides HDFS entities from the metadata load. For example, you can use the search query to exclude YARN or Oozie job executions.

Data lineage changes

To reduce complexity of the data lineage diagram, Metadata Manager has the following changes:

- Metadata Manager no longer displays data lineage for Hive query template parts. You can run data lineage analysis on Hive query templates instead.
- For partitioned Hive tables, Metadata Manager displays data lineage links between each column in the table and the parent directory that contains the related HDFS entities. Previously, Metadata Manager displayed a data lineage link between each column and each related HDFS entity.

For more information about Cloudera Navigator resources, see the "Database Management Resources" chapter in the *Informatica 10.1.1 Metadata Manager Administrator Guide*.

Netezza Resources

Effective in version 10.1.1, Metadata Manager supports multiple schemas for Netezza resources.

Netezza resources have the following behavior changes:

- When you create or edit a Netezza resource, you select the schemas from which to extract metadata. You can select one or multiple schemas.
- Metadata Manager organizes Netezza objects in the metadata catalog by schema. The database does not appear in the metadata catalog.
- When you configure connection assignments to Netezza, you select the schema to which you want to assign the connection.

Because of these changes, Netezza resources behave like other types of relational resources.

Previously, when you created or edited a Netezza resource, you could not select the schemas from which to extract metadata. If you created a resource from a Netezza database that included multiple schemas, Metadata Manager ignored the schema information. Metadata Manager organized Netezza objects in the metadata catalog by database. When you configured connection assignments to Netezza, you selected the database to which to assign the connection.

For more information about Netezza resources, see the "Database Management Resources" chapter in the *Informatica 10.1.1 Metadata Manager Administrator Guide*.

PowerExchange Adapters

This section describes changes to PowerExchange adapters in version 10.1.1.

PowerExchange Adapters for Informatica

This section describes changes to Informatica adapters in version 10.1.1.

PowerExchange for Hive

Effective in version 10.1.1, PowerExchange for Hive has the following connection modes for Hive Connection:

- Access Hive as a source or target
- Use Hive to run mappings in Hadoop cluster

Previously, the connection modes were:

- Access HiveServer2 to run mappings
- Access Hive CLI to run mappings

For more information, see the *Informatica 10.1.1 PowerExchange for Hive User Guide*.

PowerExchange for Tableau

Effective in version 10.1.1, PowerExchange for Tableau has the following changes:

- PowerExchange for Tableau installs with Informatica 10.1.1.
  Previously, PowerExchange for Tableau had a separate installer.
- When you configure a target operation to publish a Tableau Data Extract (TDE) file, you can use the append operation in the advanced properties to add data to an existing TDE file in Tableau Server and Tableau Online.
  Previously, you could configure the append operation to publish the TDE file only to Tableau Desktop.

For more information, see the *Informatica 10.1.1 PowerExchange for Tableau User Guide*.

PowerExchange Adapters for PowerCenter

This section describes changes to PowerCenter adapters in version 10.1.1.

PowerExchange for Essbase

Effective in version 10.1.1, PowerExchange for Essbase installs with PowerCenter.

Previously, PowerExchange for Essbase had a separate installer.

For more information, see the *Informatica 10.1.1 PowerExchange for Essbase User Guide for PowerCenter*.

PowerExchange for Greenplum

Effective in version 10.1.1, PowerExchange for Greenplum installs with PowerCenter.

Previously, PowerExchange for Greenplum had a separate installer.

For more information, see the *Informatica 10.1.1 PowerExchange for Greenplum User Guide for PowerCenter*.

PowerExchange for Microsoft Dynamics CRM

Effective in version 10.1.1, PowerExchange for Microsoft Dynamics CRM installs with PowerCenter.

Previously, PowerExchange for Microsoft Dynamics CRM had a separate installer.

For more information, see the *Informatica 10.1.1 PowerExchange for Microsoft Dynamics CRM User Guide for PowerCenter*.

PowerExchange for Tableau

Effective in version 10.1.1, PowerExchange for Tableau has the following changes:

- PowerExchange for Tableau installs with PowerCenter.
  Previously, PowerExchange for Tableau had a separate installer.

- When you configure a target operation to publish a Tableau Data Extract (TDE) file, you can configure the append operation in the session properties to add data to an existing TDE file in Tableau Server and Tableau Online.
  Previously, you could configure the append operation to publish the TDE file only to Tableau Desktop.

For more information, see the *Informatica 10.1.1 PowerExchange for Tableau User Guide for PowerCenter*.

Transformations

This section describes changed transformation behavior in version 10.1.1.

InformaticaTransformations

This section describes the changes to the Informatica transformations in version 10.1.1.

Address Validator Transformation

Effective in version 10.1.1, the Address Validator transformation uses version 5.9.0 of the Informatica Address Verification software engine. The engine enables the features that Informatica adds to the Address Validator transformation in version 10.1.1.

Previously, the transformation used version 5.8.1 of the engine.

For more information, see the *Informatica 10.1.1 Developer Transformation Guide* and the *Informatica 10.1.1 Address Validator Port Reference*.

Workflows

This section describes changed workflow behavior in version 10.1.1.

Informatica Workflows

This section describes the changes to Informatica workflow behavior in version 10.1.1.

Nested Inclusive Gateways

Effective in version 10.1.1, you can add one or more pairs of gateways to a sequence flow between two Inclusive gateways or two Exclusive gateways.

Previously, you invalidated the workflow if you added a pair of gateways to a sequence flow between two Inclusive gateways.

For more information, see the *Informatica 10.1.1 Developer Workflow Guide*.

Documentation

This section describes documentation changes in version 10.1.1.

Metadata Manager Documentation

Effective in version 10.1.1, the *Informatica Metadata Manager Repository Reports Reference* is obsolete because Informatica dropped support for the Reporting and Dashboards Service and for JasperReports Server.

PowerExchange for SAP NetWeaver Documentation

Effective in version 10.1.1, the following guides are obsolete because Informatica dropped support for the Analytic Business Components functionality:

- *Informatica PowerExchange for SAP NetWeaver Analytic Business Components Guide*
- *Informatica PowerExchange for SAP NetWeaver Analytic Business Components Transport Version Installation Notice*

CHAPTER 14

Release Tasks (10.1.1)

This chapter includes the following topic:

- Metadata Manager, 180

Metadata Manager

This section describes release tasks for Metadata Manager in version 10.1.1.

Business Intelligence Resources

Effective in version 10.1.1, the **Worker Threads** configuration property for some Business Intelligence resources is replaced with the **Multiple Threads** configuration property. If you set the Worker Threads property in the previous version of Metadata Manager, set the Multiple Threads property to the same value after you upgrade.

Update the value of the Multiple Threads property for the following resources:

- Business Objects
- Cognos
- Oracle Business Intelligence Enterprise Edition
- Tableau

The Multiple Threads configuration property controls the number of worker threads that the Metadata Manager Agent uses to extract metadata asynchronously. If you do not update the Multiple Threads property after upgrade, the Metadata Manager Agent calculates the number of worker threads. The Metadata Manager Agent allocates between one and six threads based on the JVM architecture and the number of available CPU cores on the machine that runs the Metadata Manager Agent.

For more information about the Multiple Threads configuration property, see the "Business Intelligence Resources" chapter in the *Informatica 10.1.1 Metadata Manager Administrator Guide*.

Cloudera Navigator Resources

Effective in version 10.1, you must configure the Java heap size for the Cloudera Navigator server and the maximum heap size for the Metadata Manager Service. If you do not correctly configure the heap sizes, the metadata load can fail.

Set the Java heap size for the Cloudera Navigator Server to at least 2 GB. If the heap size is not sufficient, the resource load fails with a connection refused error.

Set the maximum heap size for the Metadata Manager Service to at least 4 GB. If you perform simultaneous resource loads, increase the maximum heap size by at least 1 GB for each resource load. For example, to load two Cloudera Navigator resources simultaneously, increase the maximum heap size by 2 GB. Therefore, you would set the **Max Heap Size** property for the Metadata Manager Service to at least 6144 MB (6 GB). If the maximum heap size is not sufficient, the load fails with an out of memory error.

For more information about Cloudera Navigator resources, see the "Database Management Resources" chapter in the *Informatica 10.1.1 Metadata Manager Administrator Guide*.

Tableau Resources

Effective in version 10.1.1, the Tableau model has minor changes. Therefore, you must purge and reload Tableau resources after you upgrade.

For more information about Tableau resources, see the "Business Intelligence Resources" chapter in the *Informatica 10.1.1 Metadata Manager Administrator Guide*.

Part IV: Version 10.1

This part contains the following chapters:

CHAPTER 15

New Products (10.1)

This chapter includes the following topics:

Intelligent Data Lake

With the advent of big data technologies, many organizations are adopting a new information storage model called data lake to solve data management challenges. The data lake model is being adopted for diverse use cases, such as business intelligence, analytics, regulatory compliance, and fraud detection.

A data lake is a shared repository of raw and enterprise data from a variety of sources. It is often built over a distributed Hadoop cluster, which provides an economical and scalable persistence and compute layer. Hadoop makes it possible to store large volumes of structured and unstructured data from various enterprise systems within and outside the organization. Data in the lake can include raw and refined data, master data and transactional data, log files, and machine data.

Organizations are also looking to provide ways for different kinds of users to access and work with all of the data in the enterprise, within the Hadoop data lake as well data outside the data lake. They want data analysts and data scientists to be able to use the data lake for ad-hoc self-service analytics to drive business innovation, without exposing the complexity of underlying technologies or the need for coding skills. IT and data governance staff want to monitor data related user activities in the enterprise. Without strong data management and governance foundation enabled by intelligence, data lakes can turn into data swamps.

In version 10.1, Informatica introduces Intelligent Data Lake, a new product to help customers derive more value from their Hadoop-based data lake and make data available to all users in the organization.

Intelligent Data Lake is a collaborative self-service big data discovery and preparation solution for data analysts and data scientists. It enables analysts to rapidly discover and turn raw data into insight and allows IT to ensure quality, visibility, and governance. With Intelligent Data Lake, analysts to spend more time on analysis and less time on finding and preparing data.

Intelligent Data Lake provides the following benefits:

- Data analysts can quickly and easily find and explore trusted data assets within the data lake and outside the data lake using semantic search and smart recommendations.
- Data analysts can transform, cleanse, and enrich data in the data lake using an Excel-like spreadsheet interface in a self-service manner without the need for coding skills.
- Data analysts can publish data and share knowledge with the rest of the community and analyze the data using their choice of BI or analytic tools.

- IT and governance staff can monitor user activity related to data usage in the lake.
- IT can track data lineage to verify that data is coming from the right sources and going to the right targets.
- IT can enforce appropriate security and governance on the data lake
- IT can operationalize the work done by data analysts into a data delivery process that can be repeated and scheduled.

Intelligent Data Lake has the following features:

Search

- Find the data in the lake as well as in the other enterprise systems using smart search and inference-based results.
- Filter assets based on dynamic facets using system attributes and custom defined classifications.

Explore

- Get an overview of assets, including custom attributes, profiling statistics for data quality, data domains for business content, and usage information.
- Add business context information by crowd-sourcing metadata enrichment and tagging.
- Preview sample data to get a sense of the data asset based on user credentials.
- Get lineage of assets to understand where data is coming from and where it is going and to build trust in the data.
- Know how the data asset is related to other assets in the enterprise based on associations with other tables or views, users, reports and data domains.
- Progressively discover additional assets with lineage and relationship views.

Acquire

- Upload personal delimited files to the lake using a wizard-based interface.
  Hive tables are automatically created for the uploads in the most optimal format.
- Create, append to, or overwrite assets for uploaded data.

Collaborate

- Organize work by adding data assets to projects.
- Add collaborators to projects with different roles, such as co-owner, editor, or viewer, and with different privileges.

Recommendations

- Improve productivity by using recommendations based on the behavior and shared knowledge of other users.
- Get recommendations for alternate assets that can be used in a project.
- Get recommendations for additional assets that can be used a project.
- Recommendations change based on what is in the project.

Prepare

- Use excel-like environment to interactively specify transformation using sample data.
- See sheet-level and column-level overviews, including value distributions and numeric and date distributions.
- Add transformations in the form of recipe steps and see the results immediately on the sheets.

- Perform column-level data cleansing and data transformation using string, math, date, logical operations.

- Perform sheet-level operations to combine, merge, aggregate, or filter data.

- Refresh the sample in the worksheet if the data in the underlying tables change.

- Derive sheets from existing sheets and get alerts when parent sheets change.

- All transformation steps are stored in the recipe which can be played back interactively.

Publish

- Use the power of the underlying Hadoop system to run large-scale data transformation without coding or scripting.

- Run data preparation steps on actual large data sets in the lake to create new data assets.

- Publish the data in the lake as a Hive table in the desired database.

- Create, append, or overwrite assets for published data.

Data Asset Operations

- Export data from the lake to a CSV file.

- Copy data into another database or table.

- Delete the data asset if allowed by user credentials.

My Activities

- Keep track of upload activities and their status.

- Keep track of publications and their status.

- View log files in case of errors and share with IT administrators if needed.

IT Monitoring

- Keep track of user, data asset and project activities by building reports on top of the audit database.

- Find information such as the top active users, the top datasets by size, prior updates, most reused assets, and the most active projects.

IT Operationalization

- Operationalize the ad-hoc work done by analysts.

- User Informatica Developer to customize and optimize the Informatica Big Data Management mappings translated from the recipes that analysts create.

- Deploy, schedule, and monitor the Informatica Big Data Management mappings to ensure that data assets are delivered at the right time to the right destinations.

- Make sure that the entitlements for access to various databases and tables in the data lake are according to security policies.

PowerExchange Adapters

PowerExchange Adapters for Informatica

This section describes new Informatica adapters in version 10.1.

PowerExchange for Amazon Redshift

Effective in version 10.1, you can use PowerExchange for Amazon Redshift to read data from and write data to Amazon Redshift. You can import Amazon Redshift business entities as read and write data objects to create and run mappings to extract data from or load data to an Amazon Redshift entity.

For more information, see the *Informatica PowerExchange for Amazon Redshift 10.1 User Guide*.

PowerExchange for Microsoft Azure Blob Storage

Effective in version 10.1, you can use PowerExchange for Microsoft Azure Blob Storage to read data from and write data to Microsoft Azure Blob Storage. You can create a Microsoft Azure Blob Storage connection to read or write Microsoft Azure Blob Storage data into a Microsoft Azure Blob Storage data object. You can validate and run mappings in native and Hadoop environments.

For more information, see the *Informatica PowerExchange for Microsoft Azure Blob Storage 10.1 User Guide*.

PowerExchange for Microsoft Azure SQL Data Warehouse

Effective in version 10.1, you can use PowerExchange for Microsoft Azure SQL Data Warehouse to read data from and write data to Microsoft Azure SQL Data Warehouse. You can validate and run mappings in native and Hadoop environments.

For more information, see the *Informatica PowerExchange for Microsoft Azure SQL Data Warehouse 10.1 User Guide*.

CHAPTER 16

New Features (10.1)

This chapter includes the following topics:

Application Services

This section describes new application services features in version 10.1.

System Services

This section describes new system service features in version 10.1.

Scheduler Service for Profiles and Scorecards

Effective in version 10.1, you can use the Scheduler Service to schedule profile runs and scorecard runs to run at a specific time or intervals.

For more information about schedules, see the "Schedules" chapter in the *Informatica 10.1 Administrator Guide*.

Set the Time Zone for a Schedule

Effective in version 10.1, when you choose a date and time to run a schedule, you also choose the time zone. When you set the time zone, you ensure that the job runs at the time you expect it to run, no matter where the Data Integration Service is running.

For more information about schedules, see the "Schedules" chapter in the *Informatica 10.1 Administrator Guide*.

Big Data

This section describes new big data features in version 10.1.

Hadoop Ecosystem

Support in Big Data Management 10.1

Effective in version 10.1, Informatica supports the following updated versions of Hadoop distrbutions:

- Azure HDInsight 3.3
- Cloudera CDH 5.5
- MapR 5.1

For the full list of Hadoop distributions that Big Data Management 10.1 supports, see the *Informatica Big Data Management 10.1 Installation and Configuration Guide*.

Hadoop Security Systems

Effective in version 10.1, Informatica supports the following security systems on the Hadoop ecosystem:

- Apache Knox
- Apache Ranger
- Apache Sentry
- HDFS Transparent Encryption

Limitations apply to some combinations of security system and Hadoop distribution platform. For more information on Informatica support for these technologies, see the *Informatica Big Data Management 10.1 Security Guide*.

Spark Runtime Engine

Effective in version 10.1, you can push mappings to the Apache Spark engine in the Hadoop environment.

Spark is an Apache project with a run-time engine that can run mappings on the Hadoop cluster. Configure the Hadoop connection properties specific to the Spark engine. After you create the mapping, you can validate it and view the execution plan in the same way as the Blaze and Hive engines.

When you push mapping logic to the Spark engine, the Data Integration Service generates a Scala program and packages it into an application. It sends the application to the Spark executor that submits it to the Resource Manager on the Hadoop cluster. The Resource Manager identifies resources to run the application. You can monitor the job in the Administrator tool.

For more information about using Spark to run mappings, see the *Informatica Big Data Management 10.1 User Guide*.

Sqoop Connectivity for Relational Sources and Targets

Effective in version 10.1, you can use Sqoop to process data between relational databases and HDFS through MapReduce programs. You can use Sqoop to import and export data. When you use Sqoop, you do not need to install the relational database client and software on any node in the Hadoop cluster.

To use Sqoop, you must configure Sqoop properties in a JDBC connection and run the mapping in the Hadoop environment. You can configure Sqoop connectivity for relational data objects, customized data objects, and logical data objects that are based on a JDBC-compliant database. For example, you can configure Sqoop connectivity for the following databases:

- Aurora
- IBM DB2
- IBM DB2 for z/OS
- Greenplum
- Microsoft SQL Server
- Netezza
- Oracle
- Teradata

You can also run a profile on data objects that use Sqoop in the Hive run-time environment.

For more information, see the *Informatica 10.1 Big Data Management User Guide*.

Transformation Support on the Blaze Engine

Effective in version 10.1, the following transformations are supported on the Blaze engine:

- Address Validator
- Case Converter
- Comparison
- Consolidation
- Data Processor
- Decision
- Key Generator
- Labeler

- Match

- Merge

- Normalizer

- Parser

- Sequence Generator

- Standardizer

- Weighted Average

The Address Validator, Consolidation, Data Processor, Match, and Sequence Generator transformations are supported with restrictions.

Effective in version 10.1, the following transformations have additional support on the Blaze engine:

- Aggregator. Supports pass-through ports.

- Lookup. Supports unconnected Lookup transformation.

For more information, see the "Mapping Objects in a Hadoop Environment" chapter in the *Informatica Big Data Management 10.1 User Guide*.

Business Glossary

This section describes new Business Glossary features in version 10.1.

Inherit Glossary Content Managers to All Assets

Effective in version 10.1, the Analyst tool assigns the data steward and owner that you assign to a glossary to all the assets in the glossary.

For more information, see the "Glossary Content Management" chapter in the *Informatica 10.1 Business Glossary Guide*.

Bi-directional Custom Relationships

Effective in version 10.1, you can create bi-directional custom relationships. You can view the direction of related assets in the relationship view diagram. In a bi-directional custom relationship, you provide the name for the relationships in both directions.

For more information, see the "Finding Glossary Content" chapter in the *Informatica 10.1 Business Glossary Guide*.

Custom Colors in the Relationship View Diagram

Effective in version 10.1, you can define the color of the line that connects related assets in the relationship view diagram.

For more information, see the "Glossary Administration" chapter in the *Informatica 10.1 Business Glossary Guide*.

Connectivity

This section describes new connectivity features in version 10.1.

Schema Names in IBM DB2 Connections

Effective in version 10.1, when you use an IBM DB2 connection to import a table in the Developer tool or the Analyst tool, you can specify one or more schema names from which you want to import the table. Use the ischemaname attribute in the metadata connection string URL to specify the schema names. Use the pipe (|) character to separate multiple schema names.

For example, enter the following syntax in the metadata connection string URL:

```
jdbc:informatica:db2://<host name>:<port>;DatabaseName=<database
name>;ischemaname=<schema_name1>|<schema_name2>|<schema_name3>
```

This feature is also available in 9.6.1 HotFix 4. It is not available in 10.0.

For more information, see the *Informatica 10.1 Developer Tool Guide* and *Informatica 10.1 Analyst Tool Guide*.

Command Line Programs

This section describes new commands in version 10.1.

infacmd bg Commands

The following table describes new infacmd bg commands:

Command	Description
listGlossary	Lists the business glossaries in the Analyst tool.
exportGlossary	Exports the business glossaries available in the Analyst tool.
importGlossary	Imports business glossaries from .xlsx or .zip files that were exported from the Analyst tool.

infacmd dis Commands

The following table describes the new infacmd dis commands:

Command	Description
ListApplicationPermissions	Lists the permissions that a user or group has for an application.
ListApplicationObjectPermissions	Lists the permissions that a user or group has for an application object such as mapping or workflow.
SetApplicationPermissions	Assigns permissions on an application to a user or a group.
SetApplicationObjectPermissions	Assigns permissions on an application object such as mapping or workflow to a user or a group.

For more information, see the "infacmd dis Command Reference" chapter in the *Informatica 10.1 Command Reference*.

infacmd ihs Commands

The following table describes new infacmd ihs commands:

Command	Description
BackupData	Backs up HDFS data in the internal Hadoop cluster to a .zip file.
UpgradeClusterService	Upgrades the Informatica Cluster Service configuration.
removeSnapshot	Removes existing HDFS snapshots so that you can run the infacmd ihs BackupData command successfully to back up HDFS data.

For more information, see the "infacmd ihs Command Reference" chapter in the *Informatica 10.1 Command Reference*.

infacmd isp Commands

The following table describes the new infacmd isp commands:

Command	Description
AssignDefaultOSProfile	Assigns a default operating system profile to a user or group.
ListDefaultOSProfiles	Lists the default operating system profiles for a user or group.
ListDomainCiphers	Displays one or more of the following cipher suite lists used by the Informatica domain or a gateway node: **Black list** User-specified list of cipher suites that the Informatica domain blocks. **Default list** List of cipher suites that Informatica supports by default. **Effective list** The list of cipher suites that the Informatica domain uses after you configure it with the infasetup updateDomainCiphers command. The effective list supports cipher suites in the default list and white list but blocks cipher suites in the black list. **White list** User-specified list of cipher suites that the Informatica domain can use in addition to the default list. You can specify which lists that you want to display.
UnassignDefaultOSProfile	Removes the default operating system profile that is assigned to a user or group.

The following table describes updated options for infacmd isp commands:

Command	Description
CreateOSProfile	The following options are added: - -DISProcessVariables - -DISEnvironmentVariables - -HadoopImpersonationUser - -HadoopImpersonationProperties - -UseLoggedInUserAsProxy - -ProductExtensionName - -ProductOptions Use these options to configure the operating system profile properties for the Data Integration Service.
UpdateOSProfile	The following options are added: - -DISProcessVariables - -DISEnvironmentVariables - -HadoopImpersonationUser - -HadoopImpersonationProperties - -UseLoggedInUserAsProxy - -ProductExtensionName - -ProductOptions Use these options to configure the operating system profile properties for the Data Integration Service.

For more information, see the "infacmd isp Command Reference" chapter in the *Informatica 10.1 Command Reference*.

infacmd ldm Commands

The following table describes new infacmd ldm commands:

Command	Description
backupData	Takes a snapshot of the HDFS directory and creates a .zip file of the snapshot in the local machine.
restoreData	Retrieves the HDFS data backup .zip file from the local system and restores data in the HDFS directory.
removeSnapshot	Removes the snapshot from the HDFS directory.
upgrade	Upgrades the Catalog Service.

For more information, see the "infacmd ldm Command Reference" chapter in the *Informatica 10.1 Command Reference*.

infacmd ms Commands

The following table describes new options for infacmd ms commands:

Command	Description
RunMapping	The command contains the following new option: - -osp. The operating system profile name if the Data Integration Service is enabled to use operating system profiles.

For more information, see the "infacmd ms Command Reference" chapter in the *Informatica 10.1 Command Reference*.

infacmd ps Commands

The following table describes new options for infacmd ps commands:

Command	Description
- Execute - executeProfile	The commands contain the following new option: - -ospn. The operating system profile name if the Data Integration Service is enabled to use operating system profiles.

For more information, see the "infacmd ps Command Reference" chapter in the *Informatica 10.1 Command Reference*.

infacmd sch Commands

The following table describes updated options for infacmd sch commands:

Command	Description
CreateSchedule	The following argument is added to the -RunnableObjects option: - -osProfileName. The operating system profile name if the Data Integration Service is enabled to use operating system profiles.
UpdateSchedule	The following argument is added to the -AddRunnableObjects option: - -osProfileName. The operating system profile name if the Data Integration Service is enabled to use operating system profiles.

For more information, see the "infacmd sch Command Reference" chapter in the *Informatica 10.1 Command Reference*.

infasetup Commands

The following table describes new infasetup commands:

Command	Description
ListDomainCiphers	Displays one or more of the following cipher suite lists used by the Informatica domain or a gateway node uses: **Black list** User-specified list of cipher suites that the Informatica domain blocks. **Default list** List of cipher suites that Informatica supports by default. **Effective list** The list of cipher suites that the Informatica domain uses after you configure it with the infasetup updateDomainCiphers command. The effective list supports cipher suites in the default list and white list but blocks cipher suites in the black list. **White list** User-specified list of cipher suites that the Informatica domain can use. You can specify which lists that you want to display.
updateDomainCiphers	Updates the cipher suites that the Informatica domain can use with a new effective list.

The following table describes updated options for infasetup commands:

Command	Description
- DefineDomain - DefineGatewayNode - DefineWorkerNode - UpdateGatewayNode - UpdateWorkerNode	The commands contain the following new options: - cipherWhiteList |-cwl - cipherWhiteListFile |-cwlf - cipherBlackList |-cbl - cipherBlackListFile |-cblf Use these options to configure cipher suites for an Informatica domain that uses secure communication within the domain or secure connections to web application services.

For more information, see the "infasetup Command Reference" chapter in the *Informatica 10.1 Command Reference*.

pmrep Commands

The following table describes a new pmrep command:

Command	Description
AssignIntegrationService	Assigns the PowerCenter Integration Service to the specified workflow.

The following table describes the updated option for a pmrep command:

Command	Description
CreateConnection	The command contains the following updated option: - -s. The connection type list includes FTP.

For more information, see the "pmrep Command Reference" chapter in the *Informatica 10.1 Command Reference*.

Documentation

This section describes new or updated guides with the Informatica documentation in version 10.1.

The Informatica documentation contains the following new guides:

Metadata Manager Command Reference

Effective in version 10.1, the *Metadata Manager Command Reference* contains information about all of the Metadata Manager command line programs. The *Metadata Manager Command Reference* is included in the online help for Metadata Manager. Previously, information about the Metadata Manager command line programs was included in the *Metadata Manager Administrator Guide*.

For more information, see the *Informatica 10.1 Metadata Manager Command Reference.*

Informatica Administrator Reference for Live Data Map®

Effective in Live Data Map version 2.0, the *Informatica Administrator Reference for Live Data Map* contains basic reference information on Informatica Administrator tasks that you need to perform in Live Data Map. The *Informatica Administrator Reference for Live Data Map* is included in the online help for Informatica Administrator.

For more information, see the *Informatica 2.0 Administrator Reference for Live Data Map*.

Exception Management

This section describes new exception management features in version 10.1.

Search and replace data values by data type

Effective in version 10.1, you can configure the options in an exception task to search and replace data values based on the data type. You can configure the options to search and replace data in any column that contains date, string, or numeric data.

When you specify a data type, the Analyst tool searches for the value that you enter in any column that uses the data type. You can find and replace any value that a string data column contains. You can perform case-sensitive searches on string data. You can search for a partial match or a complete match between the search value and the contents of a field in a string data column.

This feature is also available in 9.6.1 HotFix 4. It is not available in 10.0.

For more information, see the Exception Records chapter in the *Informatica 10.1 Exception Management Guide*.

Informatica Administrator

This section describes new Administrator tool features in version 10.1.

Domain View

Effective in 10.1, you can view historical statistics for CPU usage and memory usage in the domain.

You can view the CPU and memory statistics for usage for the last 60 minutes. You can toggle between the current statistics and the last 60 minutes. In the **Domain** view choose **Actions** > **Current** or **Actions** > **Last Hour Trend** in the **CPU Usage** panel or the **Memory Usage** panel.
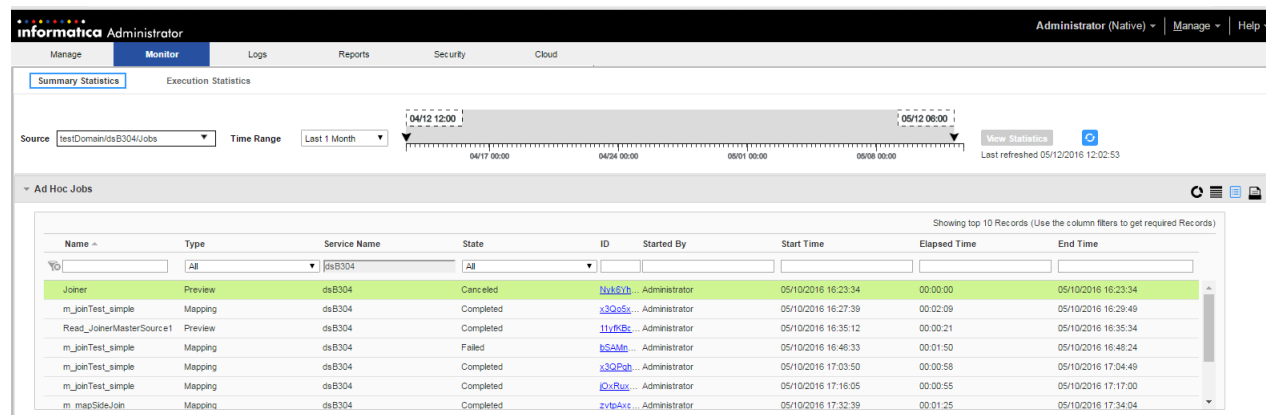
Monitoring

Effective in version 10.1, the Monitor tab in the Administrator tool has the following features:

Details view on the Summary Statistics view

The **Summary Statistics** view has a **Details** view. You can view information about jobs, export the list to a .csv file, and link to a job in the **Execution Statistics** view. To access the Details view, click **View Details**.

The following image shows the **Details** view:

Historical Statistics view.

When you select an Ad Hoc or a deployed mapping job in the **Contents** panel of the **Monitor** tab, the **Details** panel contains the **Historical Statistics** view. The **Historical Statistics** view shows averaged data from multiple runs for a specific job. For example, you can view the minimum, maximum, and average duration of the mapping job. You can view the average amount of CPU that the job consumes when it runs.

The following image shows the **Historical Statistics** view:
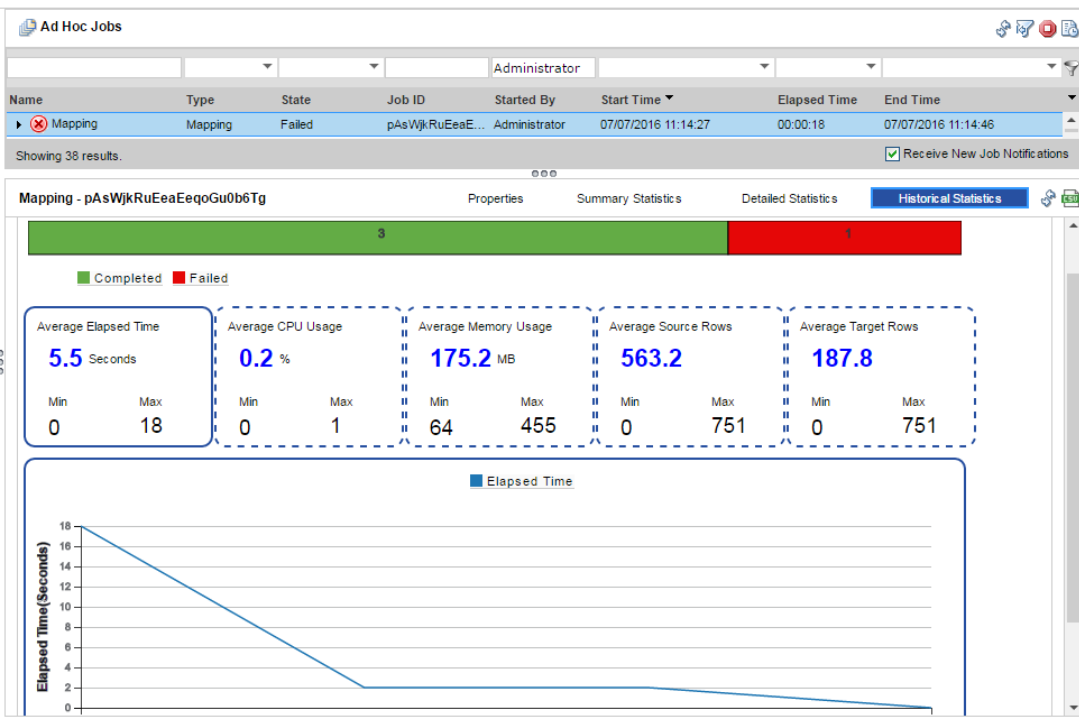
Informatica Analyst

This section describes new Analyst tool features in version 10.1.

Profiles

This section describes new Analyst tool features for profiles and scorecards.

Conformance Criteria

Effective in version 10.1, you can select a minimum number of conforming rows as conformance criteria for data domain discovery.

For more information about conformance criteria, see the "Data Domain Discovery in Informatica Analyst" chapter in the *Informatica 10.1 Data Discovery Guide*.

Exclude Nulls for Data Domain Discovery

Effective in version 10.1, you can exclude null values from the data set when you perform data domain discovery on a data source. When you select the minimum percentage of rows with the exclude null values option, the conformance percentage is the ratio of number of matching rows divided the total number of rows excluding the null values in the column.

For more information about exclude null values from data domain discovery option, see the "Data Domain Discovery in Informatica Analyst" chapter in the *Informatica 10.1 Data Discovery Guide*.

Run-time Environment

Effective in version 10.1, you can choose the Hadoop option as the run-time environment when you create or edit a column profile, data domain discovery profile, enterprise discovery profile, or scorecard. When you choose the Hadoop option, the Data Integration Service pushes the profile logic to the Blaze engine on the Hadoop cluster to run profiles.

For more information about run-time environment, see the "Data Object Profiles" chapter in the *Informatica 10.1 Data Discovery Guide*.

Scorecard Dashboard

Effective in version 10.1, you can view the following scorecard details in the scorecard dashboard:

- Total number of scorecards in the projects
- Scorecard run trend for the past six months
- Total number of data objects and the number of data objects that have scorecards
- Cumulative metrics trend for the past six months

For more information about scorecard dashboard, see the "Scorecards in Informatica Analyst" chapter in the *Informatica 10.1 Data Discovery Guide*.

Informatica Developer

This section describes new Informatica Developer features in version 10.1.

Generate Source File Name

Effective in 10.1, you can use the file name column option to return the source file name. You can configure the mapping to write the source file name to each source row.

For more information, see the *Informatica 10.1 Developer Tool Guide.*

Import from PowerCenter

Effective in version 10.1, you can import mappings that contain Netezza and Teradata objects from PowerCenter into the Developer tool and run the mappings in a native or Hadoop run-time environment.

For more information, see the *Informatica 10.1 Developer Mapping Guide*.

Copy Text Between Excel and the Developer Tool

Effective in version 10.1, you can copy text from Excel to the Developer tool or from the Developer tool to Excel. Copy text from Excel to the Developer tool to provide metadata for transformations. For example, you have designed a mapping in Excel that includes all transformations, their port names, data types, and transformation logic. In the Developer tool, you can copy the fields from Excel into the ports of empty transformations. Similarly, you can copy transformation ports from the Developer tool into Excel.

Logical Data Object Read and Write Mapping Editing

Effective in Informatica 10.1, you can use the logical data object editor to edit and change metadata in logical data object Read and Write mappings. For more information, see the "Logical View of Data" chapter in the *Informatica 10.1 Developer Tool Guide*.

DDL Query

Effective in version 10.1, when you choose to create or replace the target at run time, you can define a DDL query based on which the Data Integration Service must create or replace the target table at run time. You can define a DDL query for relational and Hive targets.

You can enter placeholders in the DDL query. The Data Integration Service substitutes the placeholders with the actual values at run time. For example, if a table contains 50 columns, instead of entering all the column names in the DDL query, you can enter a placeholder.

You can enter the following placeholders in the DDL query:

- INFA_TABLE_NAME
- INFA_COLUMN_LIST
- INFA_PORT_SELECTOR

You can also enter parameters in the DDL query.

For more information, see the *Informatica 10.1 Developer Mapping Guide*.

Profiles

This section describes new Developer tool features for profiles and scorecards.

Columns Profiles with Avro and Parquet Data Sources

Effective in version 10.1, you can create a column profile on an Avro or Parquet data source in HDFS.

For more information about column profiles on Avro and Parquet data sources, see the "Column Profiles on Semi-structured Data Sources" chapter in the *Informatica 10.1 Data Discovery Guide*.

Conformance Criteria

Effective in version 10.1, you can select a minimum number of conforming rows as conformance criteria for data domain discovery.

For more information about conformance criteria, see the "Data Domain Discovery in Informatica Developer" chapter in the *Informatica 10.1 Data Discovery Guide*.

Exclude Nulls for Data Domain Discovery

Effective in version 10.1, you can exclude null values from the data set when you perform data domain discovery on a data source. When you select the minimum percentage of rows with the exclude null values option, the conformance percentage is the ratio of number of matching rows divided by the total number of rows excluding the null values in the column.

For more information about exclude null values from data domain discovery option, see the "Data Domain Discovery in Informatica Developer" chapter in the *Informatica 10.1 Data Discovery Guide*.

Run-time Environment

Effective in version 10.1, you can choose the Hadoop option as the run-time environment when you create or edit a column profile, data domain discovery profile, enterprise discovery profile, or scorecard. When you

choose the Hadoop option, the Data Integration Service pushes the profile logic to the Blaze engine on the Hadoop cluster to run profiles.

For more information about run-time environment, see the "Data Object Profiles" chapter in the *Informatica 10.1 Data Discovery Guide*.

Informatica Development Platform

This section describes new features and enhancements to the Informatica Development Platform.

Informatica Connector Toolkit

Effective in version 10.1, you can use the following features in the Informatica Connector Toolkit:

Pre-defined type system

When you create a connector that uses REST APIs to connect to the data source, you can use pre-defined data types. You can use the following Informatica Platform data types:

- string
- integer
- bigInteger
- decimal
- double
- binary
- date

Procedure pattern

When you create a connector for Informatica Cloud, you can define native metadata objects for procedures in data sources. You can use the following options to define the native metadata object for a procedure:

Manually create the native metadata object

When you define the native metadata objects manually, you can specify the following details:

Metadata Component	Description
Procedure extension	Additional metadata information that you can specify for a procedure.
Parameter extension	Additional metadata information that you can specify for parameters.
Call capability attributes	Additional metadata information that you can specify to create a read or write call to a procedure.

Use swagger specifications

When you use swagger specifications to define the native metadata object, you can either use an existing swagger specification or you can generate a swagger specification by sampling the REST end point.

Edit common metadata

You can specify common metadata information for Informatica Cloud connectors, such as schema name and foreign key name.

Export the connector files for Informatica Cloud

After you design and implement the connector components, you can export the connector files for Informatica Cloud by specifying the plug-in ID and plug-in version.

Export the connector files for PowerCenter

After you design and implement the connector components, you can export the connector files for PowerCenter by specifying the PowerCenter version.

Live Data Map

This section describes new Live Data Map features in version 10.1.

Email Notifications

Effective in version 10.1, you can configure and receive email notifications on the Catalog Service status to closely monitor and troubleshoot the application service issues. You use the Email Service and the associated Model Repository Service to send email notifications.

For more information, see the *Informatica 10.1 Administrator Reference for Live Data Map*.

Keyword Search

Effective in version 10.1, you can use the following keywords to restrict the search results to specific types of assets:

- Table
- Column
- File
- Report

For example, if you want to search for all the tables with the term "customer" in them, type in "tables with customer" in the Search box. Enterprise Information Catalog lists all the tables that include the search term "customer" in the table name.

For more information, see the *Informatica 10.1 Enterprise Information Catalog User Guide*.

Profiling

Effective in version 10.1, Live Data Map can run profiles in the Hadoop environment. When you choose the Hadoop connection, the Data Integration Service pushes the profile logic to the Blaze engine on the Hadoop cluster to run profiles.

For more information, see the *Informatica 10.1 Live Data Map Administrator Guide*.

Scanners

Effective in version 10.1, you can extract metadata from the following sources:

- Amazon Redshift
- Amazon S3

- Custom Lineage
- HDFS
- Hive
- Informatica Cloud
- MicroStrategy

For more information, see the *Informatica 10.1 Live Data Map Administrator Guide*.

Mappings

This section describes new mapping features in version 10.1.

Informatica Mappings

This section describes new features for Informatica mappings in version 10.1.

Generate a Mapplet from Connected Transformations

Effective in version 10.1, you can generate a mapplet from a group of connected transformations in a mapping. Use the mapplet as a template to add to multiple mappings that connect to different sources and targets.

Generate a Mapping or Logical Data Object from an SQL Query

Effective in version 10.1, you can generate a mapping or a logical data object from an SQL query in the Developer tool.

To generate a mapping or logical data object from an SQL query, click **File** > **New** > **Mapping from SQL Query**. Enter a SQL query or select the location of the text file with an SQL query that you want to convert to a mapping. You can also generate a logical data object from an SQL query that contains only SELECT statements.

For more information about generating a mapping or a logical data object from an SQL query, see the *Informatica 10.1 Developer Mapping Guide*.

Metadata Manager

This section describes new Metadata Manager features in version 10.1.

Universal Resources

Effective in version 10.1, you can create universal resources to extract metadata from some metadata sources for which Metadata Manager does not package a model. For example, you can create a universal resource to extract metadata from an Apache Hadoop Hive Server, QlikView, or Talend metadata source.

To extract metadata from these sources, you first create an XConnect that represents the metadata source type. The XConnect includes the model for the metadata source. You then create one or more resources that

are based on the model. The universal resources that you create behave like packaged resources in Metadata Manager.

For more information about universal resources, see the "Universal Resources" chapter in the *Informatica 10.1 Metadata Manager Administrator Guide*.

Incremental Loading for Oracle and Teradata Resources

Effective in version 10.1, you can enable incremental loading for Oracle resources and for Teradata resources. An incremental load causes Metadata Manager 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.

To enable incremental loading for an Oracle resource or for a Teradata resource, enable **Incremental load** option in the resource configuration properties. This option is disabled by default.

For more information about incremental loading for Oracle and Teradata resources, see the "Database Management Resources" chapter in the *Informatica 10.1 Metadata Manager Administrator Guide*.

Hiding Resources in the Summary View

Effective in version 10.1, you can prevent a resource and its child objects from being displayed in the summary view of data lineage diagrams. To hide a resource, enable the **Hide in Summary Lineage** option on the **Properties** page of the resource configuration properties. This option is available for all resource types. It is disabled by default.

You can hide objects such as staging databases from data lineage diagrams. If you want to view the hidden objects, you can switch from the summary view to the detail view through the task bar.

For more information about the summary view of data lineage diagrams, see the "Working with Data Lineage" chapter in the *Informatica 10.1 Metadata Manager User Guide*.

Creating an SQL Server Integration Services Resource from Multiple Package Files

Effective in version 10.1, you can create a Microsoft SQL Server Integration Services resource that extracts metadata from packages in separate package (.dtsx) files. The package files must be in the same directory.

To create a resource that extracts metadata from packages in different package files, specify the directory that contains the package files in the **Directory** resource configuration property.

For more information about creating and configuring Microsoft SQL Server Integration Services resources, see the "Database Management Resources" chapter in the *Informatica 10.1.1 Metadata Manager Administrator Guide*.

Metadata Manager Command Line Programs

Effective in version 10.1, Metadata Manager has a new command line program. The mmXConPluginUtil command line program generates the image mapping information or the plug-in for a universal XConnect.

The following table describes the mmXConPluginUtil commands:

Command Name	Description
generateImageMapping	Generates the image mapping information for a universal XConnect.
generatePlugin	Generates the plug-in for a universal XConnect.

For more information about the mmXConPluginUtil command line program, see the "mmXConPluginUtil" chapter in the *Informatica 10.1 Metadata Manager Command Reference*.

Application Properties

Effective in version 10.1 you can configure new application properties in the Metadata Manager imm.properties file. This feature is also available in 9.6.1 HotFix 4. It is not available in 10.0.

The following table describes new Metadata Manager application properties in imm.properties:

Property	Description
xconnect.custom.failLoadOnErrorCount	Maximum number of errors that the Metadata Manager Service can encounter before the custom resource load fails.
xconnect.io.print.batch.errors	Number of errors that the Metadata Manager Service writes to the in memory cache and to the mm.log file in one batch when you load a custom resource.

For more information about the imm.properties file, see the "Metadata Manager Properties Files" appendix in the *Informatica 10.1 Metadata Manager Administrator Guide*.

Migrate Business Glossary Audit Trail History and Links to Technical Metadata

Effective in version 10.1, you can migrate audit trail history and links to technical metadata when you export business glossaries. You can import the audit trail history and links in the Analyst tool.

This feature is also available in 9.6.1 HotFix 4. It is not available in 10.0.

For more information, see the *Informatica 10.1 Upgrading from Version 9.5.1 Guide*.

PowerCenter

This section describes new PowerCenter features in version 10.1.

Create a Source Definition from a Target Definition

Effective in version 10.1, you can create a source definition from a target definition. You can drag the target definitions into the Source Analyzer to create source definitions.

For more information, see the *Informatica 10.1 PowerCenter Designer Guide*.

Create an FTP Connection Type from the Command Line

Effective in version 10.1, you can create an FTP connection with the *pmrep CreateConnection* command.

For more information, see the "pmrep Command Reference" chapter in the *Informatica 10.1 Command Reference*.

Pushdown Optimization for Greenplum

Effective in version 10.1, the PowerCenter Integration Service can push transformation logic to Greenplum sources and targets when the connection type is ODBC.

For more information, see the *Informatica PowerCenter 10.1 Advanced Workflow Guide*.

PowerExchange Adapters

This section describes new PowerExchange adapter features in version 10.1.

PowerExchange Adapters for Informatica

This section describes new Informatica adapter features in version 10.1.

PowerExchange for HDFS

Effective in version 10.1, you can use PowerExchange for HDFS to read Avro and Parquet data files from and write Avro and Parquet data files to HDFS and local file system without using a Data Processor transformation.

For more information, see the *Informatica PowerExchange for HDFS 10.1 User Guide*.

PowerExchange for Hive

Effective in version 10.1, you can use char and varchar data types in mappings. You can also select different Hive databases when you create a data object and a mapping.

For more information, see the *Informatica PowerExchange for Hive 10.1 User Guide*.

PowerExchange for Teradata Parallel Transporter API

Effective in version 10.1, you can enable Teradata Connector for Hadoop (TDCH) to run a Teradata mapping on a Blaze engine. When you run the mapping, the Data Integration Service pushes the mapping to a Hadoop cluster and processes the mapping on a Blaze engine, which significantly increases the performance.

For more information, see the *Informatica PowerExchange for Teradata Parallel Transporter API 10.1 User Guide*.

PowerExchange Adapters for PowerCenter

This section describes new PowerCenter adapter features in version 10.1.

PowerExchange for Greenplum

Effective in version 10.1, you can configure Kerberos authentication for native Greenplum connections.

This feature is also available in 9.6.1 HotFix 4. It is not available in 10.0.

For more information, see the "Greenplum Sessions and Workflows" chapter in the *Informatica 10.1 PowerExchange for Greenplum User Guide for PowerCenter*.

Security

This section describes new security features in version 10.1.

Custom Cipher Suites

Effective in version 10.1, you can customize the cipher suites that the Informatica domain uses for secure communication within the domain and secure connections to web application services. You can create a whitelist and blacklist to enable or block specific ciphersuites. This feature is also available in 9.6.1 HotFix 4. It is not available in 10.0.

The Informatica domain uses an effective list of cipher suites that uses the cipher suites in the default and whitelists but blocks cipher suites in the blacklist.

For more information, see the "Domain Security" chapter in the *Informatica 10.1 Security Guide*.

Operating System Profiles

Effective in version 10.1, if the Data Integration Service runs on UNIX or Linux, you can create operating system profiles and configure the Data Integration Service to use operating system profiles. Use operating system profiles to increase security and to isolate the run-time user environment in Informatica products such as Big Data Management, Data Quality, and Intelligent Data Lake.

The Data Integration Service uses operating system profiles to run mappings, profiles, scorecards, and workflows. The operating system profile contains the operating system user name, service process variables, Hadoop impersonation properties, the Analyst Service properties, environment variables, and permissions. The Data Integration Service runs the mapping, profile, scorecard, or workflow with the system permissions of the operating system user and the properties defined in the operating system profile.

For more information about operating system profiles, see the "Users and Groups" chapter in the *Informatica 10.1 Security Guide*.

Application and Application Object Permissions

Effective in version 10.1, you can assign permissions to control the level of access that a user or group has on applications and application objects such as mappings and workflows.

For more information about application and application object permissions, see the "Permissions" chapter in the *Informatica 10.1 Security Guide*.

Transformations

This section describes new transformation features in version 10.1.

Informatica Transformations

This section describes new features in Informatica transformation in version 10.1.

Address Validator Transformation

This section describes the new Address Validator transformation features.

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

Ireland

Effective in version 10.1, you can return the eircode for an address in Ireland. An eircode is a seven-character code that uniquely identifies an Ireland address. The eircode system covers all residences, public buildings, and business premises and includes apartment addresses and addresses in rural townlands.

To return the eircode for an address, select a Postcode port or a Postcode Complete port.

France

Effective in version 10.1, address validation uses the Hexaligne 3 repository of the National Address Management Service to certify a France address to the SNA standard.

The Hexaligne 3 data set contains additional information on delivery point addresses, including sub-building details such as building names and residence names.

Germany

Effective in version 10.1, you can retrieve the three-digit street code part of the *Frachtleitcode* or Freight Code as an enrichment to a valid Germany addresses. The street code identifies the street within the address.

To retrieve the street code as an enrichment to verified Germany addresses, select the Street Code DE port. Find the port in the DE Supplementary port group.

Informatica adds the Street Code DE port in version 10.1.

South Korea

Effective in version 10.1, you can verify older, lot-based addresses and addresses with older, six-digit post codes in South Korea. You can verify and update addresses that use the current format, the older format, and a combination of the current and older formats. A current South Korea address has a street-based format and includes a five-digit post code. A non-current address has a lot-based format and includes a six-digit post code.

To verify a South Korea address in an older format and to change the information to another format, use the Address Identifier KR ports. You update the address information in two stages. First, run the address validation mapping in batch or interactive mode and select the Address Identifier KR output port. Then, run the address validation mapping in address code lookup mode and select the Address Identifier KR input port. Find the Address Identifier KR input port in the Discrete port group. Find the Address Identifier KR output port in the KR Supplementary port group.

To verify that the Address Validator transformation can read and write the address data, add the Supplementary KR Status port to the transformation.

Informatica adds the Address Identifier KR ports, the Supplementary KR Status port, and the KR Supplementary port group in version 10.1.

Effective in version 10.1, you can retrieve South Korea address data in the Hangul script and in a Latin script.

United Kingdom

Effective in version 10.1, you can retrieve delivery point type data and organization key data for a United Kingdom address. The delivery point type is a single-character code that indicates whether the address points to a residence, a small organization, or a large organization. The organization key is an eight-digit code that the Royal Mail assigns to small organizations.

To add the delivery point type to a United Kingdom address, use the Delivery Point Type GB port. To add the organization key to a United Kingdom address, use the Organization Key GB port. Find the ports in the UK Supplementary port group. To verify that the Address Validator transformation can read and write the data, add the Supplementary UK Status port to the transformation.

Informatica adds the Delivery Point Type GB port and the Organization Key GB port in version 10.1.

These features are also available in 9.6.1 HotFix 4. They are not available in 10.0.

For more information, see the *Informatica 10.1 Address Validator Port Reference*.

Data Processor Transformation

This section describes new Data Processor transformation features.

REST API

An application can call the Data Transformation REST API to run a Data Transformation service.

For more information, see the *Informatica 10.1 Data Transformation REST API User Guide*.

XmlToDocument_45 Document Processor

The **XmlToDocument_45** document processor converts XML data to document formats, such as PDF or Excel. This component uses the **Business Intelligence and Reporting Tool** (BIRT) version 4.5 Eclipse add-on. Document processors for older versions of BIRT are also available.

For more information, see the *Informatica10.1 Data Transformation User Guide*.

Relational to Hierarchical Transformation

This section describes the Relational to Hierarchical transformation that you create in the Developer tool.

The Relational to Hierarchical transformation is an optimized transformation introduced in version 10.1 that converts relational input to hierarchical output.

For more information, see the *Informatica 10.1 Developer Transformation Guide*.

Workflows

This section describes new workflow features in version 10.1.

PowerCenter Workflows

This section describes new features in PowerCenter workflows in version 10.1.

Assign Workflows to the PowerCenter Integration Service

Effective in version 10.1, you can assign a workflow to the PowerCenter Integration Service with the *pmrep AssignIntegrationService* command.

For more information, see the "pmrep Command Reference" chapter in the *Informatica 10.1 Command Reference*.

CHAPTER 17

Changes (10.1)

This chapter includes the following topics:

Support Changes

Effective in version 10.1, Informatica announces the following support changes:

Informatica Installation

Effective in version 10.1, Informatica implemented the following change in operating system:

Support Change	Level of Support	Comments
SUSE 11	Added support	Effective in version 10.1, Informatica added support for SUSE Linux Enterprise Server 11.

Reporting Service (Deprecated)

Effective in version 10.1, Informatica deprecated the Reporting Service. Informatica will drop support for the Reporting Service in a future release. The Reporting Service custom roles are deprecated.

If you upgrade to version 10.1, you can continue to use the Reporting Service. You can continue to use Data Analyzer. Informatica recommends that you begin using a third-party reporting tool before Informatica drops

support. You can use the recommended SQL queries for building all the reports shipped with earlier versions of PowerCenter.

If you install version 10.1, you cannot create a Reporting Service. You cannot use Data Analyzer. You must use a third-party reporting tool to run PowerCenter and Metadata Manager reports.

For information about the PowerCenter Reports, see the *Informatica PowerCenter Using PowerCenter Reports Guide*. For information about the PowerCenter repository views, see the *Informatica PowerCenter Repository Guide*. For information about the Metadata Manager repository views, see the *Informatica Metadata Manager View Reference*.

Reporting and Dashboard Service (Deprecated)

Effective in version 10.1, Informatica deprecated the Reporting and Dashboards Service. Informatica will drop support for the Reporting and Dashboards Service in a future release.

If you upgrade to version 10.1, you can continue to use the Reporting and Dashboards Service. Informatica recommends that you begin using a third-party reporting tool before Informatica drops support. You can use the recommended SQL queries for building all the reports shipped with earlier versions of PowerCenter.

If you install version 10.1, you cannot create a Reporting and Dashboards Service. You must use a third-party reporting tool to run PowerCenter and Metadata Manager reports.

For information about the PowerCenter Reports, see the *Informatica PowerCenter Using PowerCenter Reports Guide*. For information about the PowerCenter repository views, see the *Informatica PowerCenter Repository Guide*. For information about the Metadata Manager repository views, see the *Informatica Metadata Manager View Reference*.

Application Services

This section describes changes to application services in version 10.1

System Services

This section describes changes to system services in version 10.1.

Email Service for Scorecard Notifications

Effective in version 10.1, scorecard notifications use the email server that you configure on the Email Service.

Previously, scorecard notifications used the email server that you configured on the domain.

For more information about the Email Service, see the "System Services" chapter in the *Informatica 10.1 Application Service Guide*.

Big Data

This section describes changes to big data features.

Effective in version 10.1, Informatica Big Data Management ships the JCE policy file and installs it when you run the installer.

Previously, you had to download and manually install the JCE policy file for AES encryption.

Business Glossary

This section describes changes to Business Glossary in version 10.1.

Custom Relationships

Effective in version 10.1, you can create custom relationships in the **Manage Glossary Relationships** workspace. Under **Manage** click **Glossary Relationships** to open the **Manage Glossary Relationships** workspace.

Previously, you had to edit the glossary template to create custom relationships.

For more information, see the "Glossary Administration" chapter in the *Informatica 10.1 Business Glossary Guide*.

Bi-Directional Default Relationships

Effective in version 10.1, the default business term relationships are bi-directional.

Previously, the default relationships were uni-directional.

For more information, see the "Finding Glossary Content" chapter in the *Informatica 10.1 Business Glossary Guide*.

Governed By Relationship

Effective in version 10.1, you can no longer create a "governed by" relationship between terns. The "governed by" relationship can only be used between a policy and a term.

Previously, you could create a "governed by" relationship between terms.

For more information, see the *Informatica 10.1 Business Glossary Guide*.

Glossary Workspace

Effective in version 10.1, in the **Glossary** workspace, the Analyst tool displays multiple Glossary assets in separate tabs.

Previously, the Analyst tool displayed only one Glossary asset in the **Glossary** workspace.

For more information, see the "Finding Glossary Content" chapter in the *Informatica 10.1 Business Glossary Guide*.

Business Glossary Desktop

Effective in version 10.1, you can install Business Glossary Desktop on the OS X operating system.

Previously, Business Glossary Desktop was available only for Windows.

For more information, see the *Informatica 10.1 Business Glossary Desktop Installation and Configuration Guide*.

Kerberos Authentication for Business Glossary Command Program

Effective in version 10.1, Business Glossary command program is supported in a domain that uses Kerberos authentication.

Previously, Business Glossary command program was not supported in a domain that uses Kerberos authentication.

For more information, see the "infacmd bg Command Reference" chapter in the *Informatica 10.1 Command Reference*.

Command Line Programs

This section describes changes to commands in version 10.1.

infacmd isp Commands

The following table describes the deprecated infacmd isp commands:

Command	Description
BackupDARepositoryContents	Backs up content for a Data Analyzer repository to a binary file. When you back up the content, the Reporting Service saves the Data Analyzer repository including the repository objects, connection information, and code page information.
CreateDARepositoryContents	Creates content for a Data Analyzer repository. You add repository content when you create the Reporting Service or delete the repository content. You cannot create content for a repository that already includes content.
CreateReportingService	Creates a Reporting Service in the domain.
DeleteDARepositoryContents	Deletes repository content from a Data Analyzer repository. When you delete repository content, you also delete all privileges and roles assigned to users for the Reporting Service.
RestoreDARepositoryContents	Restores content for a Data Analyzer repository from a binary file. You can restore metadata from a repository backup file to a database. If you restore the backup file on an existing database, you overwrite the existing content.
UpdateReportingService	Updates or creates the service and lineage options for the Reporting Service.

Command	Description
UpgradeDARepositoryCont ents	Upgrades content for a Data Analyzer repository.
UpgradeDARepositoryUser s	Upgrades users and groups in a Data Analyzer repository. When you upgrade the users and groups in the Data Analyzer repository, the Service Manager moves them to the Informatica domain.

For more information, see the "infacmd isp Command Reference" chapter in the *Informatica 10.1 Command Reference*.

Exception Management

This section describes the changes to exception management in version 10.1.

Default search and replace operations in an exception task

Effective in version 10.1, you can configure the options in an exception task to find and replace data values in one or more columns. You can specify a single column, or you can specify any column that uses a string, date, or numeric data type. By default, a find and replace operation applies to all columns that contain string data.

Previously, a find and replace operation ran by default on all of the data in the task. In version 10.1, you cannot configure a find and replace operation to run on all of the data in the task.

For more information, see the Exception Records chapter in the *Informatica 10.1 Exception Management Guide*.

Informatica Developer

This section describes the changes to the Developer tool in version 10.1.

Keyboard Shortcuts

Effective in version 10.1, the shortcut key to select the next area is **CTRL + Tab** followed by pressing the **Tab** button three times.

Previously, the shortcut key was Ctrl+Tab followed by Ctrl+Tab.

For more information, see the "Keyboard Shortcuts" appendix in the *Informatica 10.1.1 Developer Tool Guide*.

Live Data Map

This section describes changes to Live Data Map in version 10.1.

Enterprise Information Catalog

This section describes the changes to Enterprise Information Catalog.

Home Page

Effective in version 10.1, the home page displays the trending search, top 50 assets, and recently viewed assets. Trending search refers to the terms that were searched the most in the catalog in the last week. The top 50 assets refer to the assets with the most number of relationships with other assets in the catalog.

Previously, the Enterprise Information Catalog home page displayed the search field, the number of resources that Live Data Map scanned metadata from, and the total number of assets in the catalog.

For more information about the Enterprise Information Catalog home page, see the "Getting Started with Informatica Enterprise Information Catalog" chapter in the *Informatica 10.1 Enterprise Information Catalog User Guide*.

Asset Overview

Effective in version 10.1, you can view the schema name associated with an asset in the **Overview** tab.

Previously, the Overview tab for an asset did not display the associated schema name.

For more information about assets in Enterprise Information Catalog, see the *Informatica 10.1 Enterprise Information Catalog User Guide*.

Live Data Map Administrator Home Page

Effective in version 10.1, the Start workspace displays the total number of assets in the catalog, unused resources, and unassigned connections in addition to many other monitoring statistics.

Previously, the Live Data Map Administrator home page displayed several monitoring statistics, such as number of resources for each resource type, task distribution, and predictive job load.

For more information about Live Data Map Administrator home page, see the "Using Live Data Map Administrator" chapter in the *Informatica 10.1 Live Data Map Administrator Guide*.

Metadata Manager

This section describes changes to Metadata Manager in version 10.1.

Microsoft SQL Server Integration Services Resources

Effective in version 10.1, Metadata Manager organizes SQL Server Integration Services objects in the metadata catalog according to the connections in which the objects are used. The metadata catalog does not contain a separate folder for each package. To select an object such as a table or column in the metadata catalog, navigate to the object through the source or target connection in which the object is used.

Previously, Metadata Manager organized SQL Server Integration Services objects by connection and by package. The metadata catalog contained a Connections folder in addition to a folder for each package.

For more information about SQL Server Integration Services resources, see the "Data Integration Resources" chapter in the *Informatica 10.1 Metadata Manager Administrator Guide*.

Certificate Validation for Command Line Programs

Effective in version 10.1, when you configure a secure connection for the Metadata Manager web application, the Metadata Manager command line programs do not accept security certificates that have errors. The property that controls whether a command line program can accept security certificates that have errors is removed. This feature is also available in 9.6.1 HotFix 4. It is not available in 10.0.

Previously, the Security.Authentication.Level property in the MMCmdConfig.properties file controlled certificate validation for mmcmd or mmRepoCmd. You could configure the property to either accept all certificates or accept only certificates that do not have errors.

Because the command line programs no longer accept security certificates that have errors, the Security.Authentication.Level property is obsolete. The property no longer appears in the MMCmdConfig.properties files for mmcmd or mmRepoCmd.

For more information about certificate validation for mmcmd and mmRepoCmd, see the "Metadata Manager Command Line Programs" chapter in the *Informatica 10.1 Metadata Manager Administrator Guide*.

PowerCenter

This section describes changes to PowerCenter in version 10.1.

Operating System Profiles

Effective in version 10.1, the OS Profile tab in the Security page of the Administrator tool is renamed to the **Operating System Profiles** tab. To create operating system profiles, go to the Security Actions menu and click **Create Operating System Profile**. You can also assign a default operating system profile to users and groups when you create an operating system profile. Previously, the Security Actions menu had an Operating System Profiles Configuration option.

For more information about managing operating system profiles, see the "Users and Groups" chapter in the *Informatica 10.1 Security Guide*.

Security

This section describes changes to security in version 10.1.

Transport Layer Security (TLS)

Effective in version 10.1, Informatica uses TLS v1.1 and v1.2 to encrypt traffic. Additionally, Informatica disabled support for TLS v1.0 and lower.

The changes affect secure communication within the Informatica domain, secure connections to web application services, and connections from the Informatica domain to an external destination.

This feature is also available in 9.6.1 HotFix 4. It is not available in 10.0.

Permissions

Effective in version 10.1, the following Model repository objects have permission changes:

- Applications, mappings, and workflows. All users in the domain are granted all permissions.
- SQL data services and web services. Users with effective permissions are assigned direct permissions.

The changes affect the level of access that users and groups have to these objects.

After you upgrade, you might need to review and change the permissions to ensure that users have appropriate permissions on objects.

For more information, see the "Permissions" chapter in the *Informatica 10.1 Security Guide*.

Transformations

This section describes changed transformation behavior in version 10.1.

Informatica Transformations

This section describes the changes to the Informatica transformations in version 10.1.

Address Validator Transformation

This section describes the changes to the Address Validator transformation.

The Address Validator transformation contains the following updates to address functionality:

Address validation engine upgrade

Effective in version 10.1, the Address Validator transformation uses version 5.8.1 of the Informatica Address Verification software engine. The engine enables the features that Informatica adds to the Address Validator transformation in version 10.1.

Previously, the transformation used version 5.7.0 of the Informatica AddressDoctor software engine.

Product name change

Informatica Address Verification is the new name of Informatica AddressDoctor. Informatica AddressDoctor became Informatica Address Verification in version 5.8.0.

Changes to geocode options for United Kingdom addresses

Effective in version 10.1, you can select Rooftop as a geocode data property to retrieve rooftop-level geocodes for United Kingdom addresses.

Previously, you selected the Arrival Point geocode data property to retrieve rooftop-level geocodes for United Kingdom addresses.

If you upgrade a repository that includes an Address Validator transformation, you do not need to reconfigure the transformation to specify the Rooftop geocode property. If you specify rooftop geocodes and the Address Validator transformation cannot return the geocodes for an address, the transformation does not return any geocode data.

Support for unique property reference numbers in United Kingdom input data

Effective in version 10.1, the Address Validator transformation has a UPRN GB input port and a UPRN GB output port.

Previously, the transformation had a UPRN GB output port.

Use the input port to retrieve a United Kingdom address for a unique property reference number that you enter. Use the UPRN GB output port to retrieve the unique property reference number for a United Kingdom address.

These features are also available in 9.6.1 HotFix 4. They are not available in 10.0.

For more information, see the *Informatica 10.1 Address Validator Port Reference*.

Data Processor Transformation

This section describes the changes to the Data Processor transformation.

Excel 2013

Effective in version 10.1, the ExcelToXml_03_07_10 document processor can process Excel 2013 files. You can use the document processor in a Data Processor transformation as a pre-processor that converts the format of a source document before a transformation.

For more information, see the *Informatica 10.1 Data Transformation User Guide*.

Performance Improvement with Avro or Parquet Input

A Data Processor transformation receives Avro or Parquet data input from a complex file reader object. Effective in version 10.1, you can configure the complex file reader settings to optimize performance for Avro or Parquet input.

For more information, see the *Informatica10.1 Data Transformation User Guide*.

Performance Improvement with COBOL Input in the Hadoop Environment

Effective in version 10.1, you can configure the complex file reader settings to optimize performance when processing large COBOL files in a Hadoop environment. Use a regular expression to define how to split record processing for an appropriate COBOL input file.

For more information, see the *Informatica10.1 Data Transformation User Guide*.

Exception Transformations

Effective in version 10.1, you can configure a Bad Record Exception transformation and a Duplicate Record Exception transformation to create exception tables in a non-default database schema.

Previously, you configured the transformations to create exception tables in the default schema on the database.

This feature is also available in 9.6.1 HotFix 4. It is not available in 10.0.

For more information, see the *Informatica 10.1 Developer Transformation Guide*.

Workflows

This section describes changed workflow behavior in version 10.1.

Informatica Workflows

This section describes the changes to Informatica workflow behavior in version 10.1.

Parallel Execution of Human Tasks

Effective in version 10.1, the Data Integration Service can run Human tasks on multiple sequence flows in a workflow in parallel. To create the parallel sequence flows, add Inclusive gateways to the workflow in the Developer tool. Add one or more Human tasks to each sequence flow between the Inclusive gateways.

Previously, you added one or more Human tasks to a single sequence flow between Inclusive gateways.

For more information, see the *Informatica 10.1 Developer Workflow Guide*.

CHAPTER 18

Release Tasks (10.1)

This chapter includes the following topics:

Metadata Manager

This section describes release tasks for Metadata Manager in version 10.1.

Informatica Platform Resources

Effective in version 10.1, to extract metadata from an Informatica 10.0 application that is deployed to a Data Integration Service, you must install the version 10.0 Command Line Utilities. Install the utilities in a directory that the 10.1 Metadata Manager Service can access. For best performance, extract the files to a directory on the machine that runs the Metadata Manager Service.

When you configure the resource, you must also enter the file path to the 10.0 Informatica Command Line Utilities installation directory in the **10.0 Command Line Utilities Directory** property.

For more information about Informatica Platform resources, see the "Data Integration Resources" chapter in the *Informatica 10.1 Metadata Manager Administrator Guide*.

Verify the Truststore File for Command Line Programs

Effective in version 10.1, when you configure a secure connection for the Metadata Manager web application, the Metadata Manager command line programs do not accept security certificates that have errors. The property that controls whether a command line program can accept security certificates that have errors is removed. This feature is also available in 9.6.1 HotFix 4. It is not available in 10.0.

The Security.Authentication.Level property in the MMCmdConfig.properties file controlled certificate validation for mmcmd or mmRepoCmd. You could set the property to one of the following values:

- NO_AUTH. The command line program accepts the digital certificate, even if the certificate has errors.
- FULL_AUTH. The command line program does not accept a security certificate that has errors.

The NO_AUTH setting is no longer valid. The command line programs now only accept security certificates that do not contain errors.

If a secure connection is configured for the Metadata Manager web application, and you previously set the Security.Authentication.Level property to NO_AUTH, you must now configure a truststore file. To configure

mmcmd or mmRepoCmd to use a truststore file, edit the MMCmdConfig.properties file that is associated with mmcmd or mmRepoCmd. Set the TrustStore.Path property to the path and file name of the truststore file.

For more information about the MMCmdConfig.properties files for mmcmd and mmRepoCmd, see the "Metadata Manager Command Line Programs" chapter in the *Informatica 10.1 Metadata Manager Administrator Guide*.

Security

This section describes release tasks for security features in version 10.1.

Permissions

After you upgrade to 10.1, the following Model repository objects have permission changes:

- Applications, mappings, and workflows. All users in the domain are granted all permissions.
- SQL data services and web services. Users with effective permissions are assigned direct permissions.

The changes affect the level of access that users and groups have to these objects.

After you upgrade, review and change the permissions on applications, mappings, workflows, SQL data services, and web services to ensure that users have appropriate permissions on objects.

For more information, see the "Permissions" chapter in the *Informatica 10.1 Security Guide*.