Informatica®

Informatica®
10.5.2.1

What's New and Changed (10.5.2.1)

Informatica What's New and Changed (10.5.2.1)
10.5.2.1
July 2022

© Copyright Informatica LLC 2003, 2023

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

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

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

Subject to your opt-out rights, the software will automatically transmit to Informatica in the USA information about the computing and network environment in which the Software is deployed and the data usage and system statistics of the deployment. This transmission is deemed part of the Services under the Informatica privacy policy and Informatica will use and otherwise process this information in accordance with the Informatica privacy policy available at https://www.informatica.com/in/privacy-policy.html. You may disable usage collection in Administrator tool.

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

See patents at https://www.informatica.com/legal/patents.html.

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

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

Publication Date: 2023-07-24

Table of Contents

Preface 28

Informatica Resources 28

- Informatica Network 28
- Informatica Knowledge Base 28
- Informatica Documentation 28
- Informatica Product Availability Matrices 29
- Informatica Velocity 29
- Informatica Marketplace 29
- Informatica Global Customer Support 29

Part I: Version 10.5.2 - 10.5.2.1.x 30

Chapter 1: Version 10.5.2.1 - 10.5.2.1.x 31

Version 10.5.2.1.2 31

Version 10.5.2.1.1 32

Version 10.5.2.1 32

- Important Notices (10.5.2.1) 32
- What's New (10.5.2.1) 32
- What's Changed (10.5.2.1) 34

Chapter 2: 10.5.2 Important Notices 35

Support Changes 35

- Technical Preview 35
- Dropped Support 36

Release Tasks 36

- PowerExchange for Microsoft Dynamics CRM 36

Informatica Installer 36

- Informatica Deployment Manager 37
- Azure PostgreSQL 37

Enterprise Data Catalog 37

- MetaDex Component Names Updated 37

Chapter 3: 10.5.2 What's New 38

Data Engineering 38

- Databricks 38
- Data Transformation 39
- Kudu Dynamic Mapping 39

Data Privacy Management 39

- Subject Registry 39

Data Quality 39

Data Transformation. . . . 39
Data Transformation ExcelToDataXml document processor. . . . 40
Enterprise Data Catalog. . . . 40
Bulk certification of assets. . . . 40
Enrichment Migration Utility. . . . 40
MetaDex Job Reports. . . . 40
Scanners. . . . 40
Profiling Enhancements. . . . 41
Resource Enhancements. . . . 41
REST APIs. . . . 42
Search Results. . . . 42
Reverse Proxy Support. . . . 42
View foreign key relationships. . . . 43
Informatica Installer. . . . 43
Azure PostgreSQL. . . . 43
PowerExchange Adapters. . . . 43
PowerExchange Adapters for Informatica. . . . 43
PowerExchange Adapters for PowerCenter. . . . 44

Chapter 4: 10.5.2 What's Changed. . . . 45

Data Engineering. . . . 45
Time zone information in logs. . . . 45
User Authentication Error Logging. . . . 45
Data Privacy Management. . . . 46
Data Store Connectivity. . . . 46
Risk Score Information. . . . 46
Default Dashboard Settings. . . . 46
Enterprise Data Catalog. . . . 47
MetaDex Server Configuration File Change. . . . 47
MetaDex Server Memory. . . . 47
Tableau Extension Download. . . . 47
Enterprise Data Catalog Plug-in Download. . . . 47
MITI Windows Agent. . . . 47
Scanner Utility. . . . 48
Informatica Developer. . . . 48
Enabling Nullable Values. . . . 48
PowerCenter. . . . 48
Time zone information in the logs. . . . 48

Part II: Version 10.5.1 - 10.5.1.1. . . . 49

Chapter 5: Version 10.5.1.1. . . . 50

What's New (10.5.1.1). . . . 50

Data Engineering Streaming. . . . 50
Data Privacy Management. . . . 50
Enterprise Data Catalog. . . . 51
What's Changed (10.5.1.1). . . . 51
Enterprise Data Catalog. . . . 51

Chapter 6: 10.5.1 Important Notices. . . . 52

Support Changes. . . . 52
Technical Preview. . . . 52
Dropped Support. . . . 53
Deprecation. . . . 53
Release Tasks. . . . 53
PowerExchange for Microsoft Dynamics CRM. . . . 53
Informatica Installer. . . . 54
Documentation. . . . 54
Upgrade Paths 54
Informatica Deployment Manager. . . . 55
Enterprise Data Catalog. . . . 55

Chapter 7: 10.5.1 What's New. . . . 56

Data Engineering Streaming. . . . 56
Transformations in Databricks. . . . 56
Intelligent Structure Models on Databricks 56
Data Privacy Management. . . . 57
Security Policy Violation Tags. . . . 57
Security Policy Task Assignment. . . . 57
DSAR Report Template. . . . 57
Microsoft Azure SQL Server Support. . . . 57
Sources. . . . 57
Data Transformation. . . . 58
Data Transformation HIX Library. . . . 58
Enterprise Data Catalog. . . . 58
Advanced Scanners. . . . 58
Sudo Requirements for the Gateway User. . . . 59
SELinux in Enforcing Mode. . . . 59
Support for RHEL 7.9 in FIPS mode. . . . 59
Data Asset Analytics. . . . 59
Sort Search Results. . . . 59
Informatica Data Quality. . . . 60
Informatica Transformations. . . . 60
Address Validator Transformation. . . . 60
PowerExchange Adapters. . . . 60
PowerExchange Adapters for Informatica. . . . 60

Security (10.5.1). 61
Kerberos Constrained Delegation. 61
Test Data Management. 61
Hive Data Types. 61

Chapter 8: 10.5.1 What's Changed. 62

Data Privacy Management 62
Sensitive Field Information Export. 62
Informatica Transformations. 62
Address Validator Transformation. 62
Intelligent Structure Models. 63
Transformation of Dates and Times. 64
Enterprise Data Catalog. 64
SSL-enabled Informatica Cluster Service and associated services. 64
Custom SSL Certificate Requirements. 64
Scanner Binaries for Informatica Platform Resources. 64
Data Asset Analytics. 64

Part III: Versions 10.5 - 10.5.0.1. 66

Chapter 9: Version 10.5.0.1. 67

Important Notices. 67
Support Changes. 67
What's New. 67
Data Privacy Management. 67
Advanced Scanners. 68
What's Changed. 68
Enterprise Data Catalog 68
PowerExchange Adapters for Informatica. 70

Chapter 10: 10.5 Important Notices. 71

Support Changes 71
Technical Preview 71
Deprecation. 73
Deferment 74
New Products. 74
PowerExchange for Kudu for Informatica. 74
Informatica Installer. 75
Upgrade Paths. 75
Data Engineering. 76
Data Privacy Management. 76
Domain. 77
Enterprise Data Catalog. 77

Informatica Deployment Manager. . . . 78
PowerCenter . . . 78
Release Tasks. . . . 78
Test Data Management. . . . 78

Chapter 11: 10.5 What's New. . . . 80

CI/CD. . . . 80
Command Line Programs. . . . 83
infacmd cluster Commands. . . . 84
infacmd ics Commands. . . . 84
infacmd isp Commands. . . . 84
infacmd ldm Commands. . . . 84
infacmd ms Commands. . . . 85
infasetup Commands. . . . 85
infasetup defineDomain. . . . 89
Data Engineering Integration. . . . 89
EXTRACT_STRUCT Function. . . . 89
File Manager for Cloud File Preprocessing. . . . 90
Mapping Audits. . . . 90
Profile on the Databricks Cluster. . . . 90
Sensitive Data Recommendations and Insights by CLAIRE. . . . 91
Warm Pool Support for Ephemeral Clusters on Databricks. . . . 91
Data Engineering Quality. . . . 91
Databricks Integration. . . . 91
Data Engineering Streaming. . . . 92
Cassandra. . . . 92
DataProc. . . . 92
Google Cloud Storage Target on Google Dataproc. . . . 92
Google PubSub. . . . 92
High Precision Data Types. . . . 92
Kudu. . . . 93
Python Transformation in Databricks. . . . 93
Data Privacy Management. . . . 93
Customize the Brand Logo. . . . 93
Data Domains. . . . 93
PostgreSQL Support. . . . 93
Privacy Dashboard. . . . 94
Sources. . . . 94
Data Transformation. . . . 94
Enterprise Data Catalog. . . . 95
Advanced Scanners. . . . 95
Cloning a Resource. . . . 95
Enhanced Home Page. . . . 96

Enhanced Filter Panel. . . . 96
Enhanced Search Results. . . . 96
New Walkthroughs. . . . 96
Notifications Page Enhancements. . . . 97
Profile on the Databricks Cluster. . . . 97
Resource Enhancements. . . . 97
Enterprise Data Preparation. . . . 98
Informatica Transformations. . . . 98
Data Masking Transformation. . . . 99
Lookup Transformation 99
Sequence Generator Transformation on Databricks. . . . 99
Informatica Workflows. . . . 99
Intelligent Structure Models. . . . 99
Metadata Manager. . . . 100
Support for Azure Microsoft SQL Server. . . . 100
PowerCenter 100
Ciphers Supported for PowerCenter Services. . . . 100
Connectivity. . . . 100
Transformations 101
PowerExchange Adapters. . . . 101
PowerExchange Adapters for Informatica. . . . 101
PowerExchange Adapters for PowerCenter 104
Security. . . . 105
Encryption Strengthening. . . . 105
SAML Functionality 106
Test Data Management. . . . 106
PostgreSQL Support. . . . 106
Sort Entries in a Relational Dictionary. . . . 106
Test Data Generation for XSD Sources. . . . 106

Chapter 12: 10.5 What's Changed. . . . 108

Command Line Programs. . . . 108
infasetup 108
Data Engineering 109
Documentation 109
Performance. . . . 109
Processing Invalid Values. . . . 109
Data Privacy Management 109
Compressed Files in Scans. . . . 109
Data Store Properties. . . . 110
Deletion of Subject Requests. . . . 110
Subject Registry Configuration File. . . . 110
Total Subject Count 110

User Activity Monitoring. . . . 110
Enterprise Data Catalog. . . . 111
Amazon S3 Resource. . . . 111
Business Term Propagation. . . . 111
Custom Resources. . . . 111
Default Lineage and Impact Diagram View. . . . 111
Enterprise Data Catalog Authentication. . . . 111
Informatica Similarity Discovery Resource 111
Reference Resources. . . . 112
Search Tabs Replaced with Search Prefilters. . . . 112
Terminology Changes in Documentation. . . . 112
Advanced Scanners uses System Models. . . . 112
Informatica Transformations. . . . 113
Intelligent Structure Models 113
PowerCenter. . . . 113
PowerExchange Adapters for Informatica. . . . 113
Security 113
SAML Authentication. . . . 114

Part IV: Versions 10.4.1 - 10.4.1.3. . . . 115

Chapter 13: Version 10.4.1.3. . . . 116

Support Changes. . . . 116
Technical Preview. . . . 116
What's New. . . . 116
Cloudera CDP. . . . 116

Chapter 14: Version 10.4.1.2. . . . 117

Important Notices. . . . 117
Technical Preview Initiated. . . . 117
What's New. . . . 117
Data Engineering Integration. . . . 117
Data Privacy Management. . . . 118
Enterprise Data Catalog. . . . 118
Security. . . . 120
What's Changed. . . . 120
Enterprise Data Catalog. . . . 120

Chapter 15: Version 10.4.1.1. . . . 123

Important Notices. . . . 123
Support Changes. . . . 123
What's New. . . . 124
Data Engineering Integration 124

Data Privacy Management. . . . 124
Enterprise Data Catalog. . . . 125
What's Changed. . . . 126
Data Privacy Management . . . 126
Enterprise Data Catalog. . . . 127

Chapter 16: 10.4.1 Important Notices. . . . 129

10.4.1 Support Changes. . . . 129
Technical Preview Support. . . . 129
Deferment. . . . 131
IPv6 Support. . . . 131
Distribution Support. . . . 131
Enterprise Data Catalog. . . . 132
New Products. . . . 132
PowerExchange for MongoDB JDBC . . . 132
Installer Changes. . . . 132
Release Tasks . . . 133
PowerExchange Adapters for Informatica. . . . 133
PowerExchange Adapters for PowerCenter. . . . 133

Chapter 17: 10.4.1 What's New. . . . 134

Connectivity. . . . 134
DataDirect ODBC and JDBC Drivers. . . . 134
Oracle Connection Manager. . . . 134
Data Engineering Integration. . . . 134
Hierarchical Data on Hive Sources and Targets. . . . 135
SAML Authentication for Mass Ingestion. . . . 135
Data Engineering Streaming. . . . 135
FileName Port for ADLS Gen2. . . . 135
Ingest CDC Data from Multiple Kafka Topics. . . . 135
JDBC V2 Lookup Transformation. . . . 135
Parquet Data Format for Complex Targets. . . . 136
Rollover Parameters in Amazon S3 and ADLS Gen2 Targets. . . . 136
Sources and Targets in Databricks. . . . 136
Streaming Mappings in AWS Databricks. . . . 136
Data Privacy Management. . . . 137
Data Stores. . . . 137
Jobs. . . . 137
Privacy Dashboard. . . . 138
Subject Registry Data Encryption. . . . 138
Unstructured Data Stores. . . . 138
Enterprise Data Catalog. . . . 138
Business Term Association . . . 138

Configure Conflict Resolution for Data Rule and Column Name Rule. . . . 139
Context Lineage Information . . . 139
Data Asset Analytics . . . 139
Data Curation through Export and Import. . . . 139
Data Discovery. . . . 139
Enterprise Data Catalog Walkthroughs. . . . 140
Extracting Transformation Endpoints. . . . 140
Field-Level Lineage for Flat Files. . . . 140
File Lineage for Cloud Storage. . . . 141
Hive Resource. . . . 141
Informatica MDM Resource. . . . 141
Microsoft SQL Server Resource. . . . 142
MicroStrategy Resource. . . . 142
Partitioned File Discovery. . . . 142
Reference Resources and Reference Assets. . . . 142
Resource Level Permissions. . . . 143
Extract the Database Name. . . . 143
REST APIs. . . . 143
SAP BW and SAP BW/4HANA Resources. . . . 143
SSIS Resource. . . . 143
Informatica Transformations. . . . 144
Macro Transformation. . . . 144
Intelligent Structure Model. . . . 144
Change the Root Node in an Array Structure. . . . 144
Metadata Manager. . . . 144
Configure Java Location and Heap Size for Business Object Resources. . . . 145
PowerExchange Adapters. . . . 145
PowerExchange Adapters for Informatica. . . . 145
PowerExchange Adapters for PowerCenter. . . . 147

Chapter 18: 10.4.1 What's Changed. . . . 149

Command Line Programs. . . . 149
Data Engineering Integration. . . . 150
Intelligent Structure Discovery. . . . 150
Data Privacy Management. . . . 150
Data Privacy Management Installer. . . . 150
Overview Workspace. . . . 150
Scan Job Reports. . . . 151
Sensitive Files Page Views. . . . 151
Documentation. . . . 151
Data Privacy Management. . . . 151
Test Data Management. . . . 152
Domain Security. . . . 152

Enabling Access to a Kerberos-Enabled Domain. . . . 152
Enterprise Data Catalog. . . . 152
Change Summary Page. . . . 152
Export Asset Data to a Tableau Data Extract File. . . . 152
Export Path of the Assets. . . . 153
Show Active Business Terms. . . . 153
Sort Column and Field Assets. . . . 153
Start Tab Name in Catalog Administrator. . . . 153
Informatica Developer . . . 153
Relational Object Search . . . 153
Viewing the Execution Plan. . . . 154
Metadata Manager. . . . 154
View Data Lineage in HTML 5. . . . 154
Test Data Management. . . . 154
Test Data Management Installer. . . . 154

Part V: Versions 10.4 - 10.4.0.2. . . . 155

Chapter 19: Version 10.4.0.2. . . . 156

What's New. . . . 156
Update Cluster Distribution Version. . . . 156

Chapter 20: Version 10.4.0.1. . . . 157

Important Notices. . . . 157
Technical Preview Lifted. . . . 157
What's New . . . 158
Data Engineering Integration. . . . 158
Enterprise Data Catalog. . . . 158
PowerExchange Adapters for Informatica. . . . 158
What's Changed. . . . 158
Enterprise Data Catalog. . . . 158
Data Engineering Integration. . . . 159
PowerExchange Adapters for Informatica. . . . 159

Chapter 21: 10.4 Important Notices. . . . 161

New Products. . . . 161
PowerExchange Adapters for Informatica. . . . 161
PowerExchange Adapters for PowerCenter. . . . 162
Product Name Changes. . . . 162
Upgrading to Version 10.4.0. . . . 163
Support Changes. . . . 163
Technical Preview Support. . . . 164
Deferment. . . . 165

Dropped Support. . . . 165
PowerCenter. . . . 165
PowerExchange Adapters for PowerCenter. . . . 165
PowerExchange Adapters for Informatica. . . . 166
Release Tasks. . . . 166
Data Engineering Integration. . . . 167

Chapter 22: 10.4 What's New. . . . 168

CI/CD. . . . 168
Data Integration Service REST API. . . . 168
infacmd dis Commands. . . . 169
Reverse Proxy Server. . . . 170
infacmd roh Commands. . . . 170
Application Patch Deployment. . . . 171
Connect to a Run-time Application. . . . 172
Object Explorer View. . . . 172
Tags. . . . 173
Command Line Programs. . . . 173
infacmd isp Commands. . . . 173
Data Engineering Integration. . . . 173
New Data Types Support . . . 174
AWS Databricks Integration. . . . 174
Cluster Workflows for HDInsight Access to ALDS Gen2 Resources. . . . 175
Databricks Delta Lake Storage Access. . . . 175
Display Nodes Used in Mapping. . . . 175
Log Aggregation . . . 175
Parsing Hierarchical Data on the Spark Engine. . . . 175
Profiles and Sampling Options on the Spark Engine. . . . 176
Python Transformation. . . . 176
Sqoop. . . . 176
Data Engineering Streaming . . . 177
Confluent Schema Registry in Streaming Mappings . . . 177
Data Quality Transformations in Streaming Mappings. . . . 177
Ephemeral Cluster in Streaming Mappings. . . . 177
FileName Port in Amazon S3 . . . 178
Microsoft Azure Data Lake Storage Gen2. . . . 178
Streaming Mappings in Azure Databricks. . . . 178
Dynamic Mappings in Data Engineering Streaming . . . 179
Enterprise Data Catalog. . . . 179
Assigning Custom Attributes to Resources and Classes. . . . 179
New Resources. . . . 179
Reference Resources and Reference Assets. . . . 180
Export Assets from Search Results. . . . 180

Lineage and Impact Filters. . . . 180
Asset Control Summary. . . . 181
Rules and Scorecards. . . . 181
Unique Key Inference. . . . 181
Data Domain Discovery on the CLOB File Type. . . . 181
Data Discovery and Sampling Options on the Spark Engine. . . . 181
Tracking Technical Preview. . . . 182
Data Preview and Provisioning. . . . 182
Supported Resource Types for Standalone Scanner Utility. . . . 182
REST APIs. . . . 183
Enterprise Data Preparation. . . . 183
Data Lake Access Management. . . . 184
Microsoft Azure Data Lake Storage as a Data Source. . . . 184
Publish Files to the Data Lake. . . . 184
Upload Files to the Data Lake. . . . 184
Informatica Mappings 185
Binding Mapping Outputs to Mapping Parameters. . . . 185
CLAIRE Recommendations and Insights. . . . 185
Update Mapping Optimizer Level. . . . 186
Informatica Transformations. . . . 186
Address Validator Transformation. . . . 186
Informatica Workflows. . . . 186
Amazon EMR Create Cluster Task Advanced Properties. . . . 186
Informatica Installation. . . . 187
PostgreSQL. . . . 187
Pre-installation (i10Pi) System Check Tool in Silent Mode. . . . 187
Encrypt Passwords in the Silent Installation Properties File. . . . 187
Intelligent Structure Model. . . . 187
Additional Input Types. . . . 187
Create Model from Sample at Design Time. . . . 187
Handling of Unidentified Data. . . . 188
PowerCenter. . . . 188
Connectivity. . . . 188
PowerExchange Adapters. . . . 189
PowerExchange Adapters for Informatica. . . . 189
PowerExchange Adapters for PowerCenter. . . . 191
Security. . . . 193
Configure Web Applications to Use Different SAML Identity Providers. . . . 193

Chapter 23: 10.4 What's Changed. . . . 194

Data Engineering Integration. . . . 194
Data Preview. . . . 194
Date/Time Format on Databricks. . . . 195

Null Values in Target. . . . 195
Python Transformation. . . . 196
Enterprise Data Preparation. . . . 196
NULL Value Handling. . . . 196
Solr Port Property. . . . 196
Enterprise Data Catalog. . . . 196
Lineage Enhancement for SAP HANA Resource. . . . 196
Profiling and Data Domain Discovery. . . . 197
Informatica Data Engineering Streaming. . . . 197
Resource Name Changes. . . . 197
Search Suggestions. . . . 197
Configuring Custom Attributes 197
Informatica Developer. . . . 197
Importing Relational Data Objects. . . . 197
Informatica Transformations. . . . 198
Address Validator Transformation. . . . 198
PowerCenter. . . . 198
Refresh Metadata in Designer and in the Workflow Manager. . . . 199
Import and Export 199
PowerExchange Adapters for Informatica. . . . 199
PowerExchange for Amazon Redshift. . . . 200
PowerExchange for Amazon S3. . . . 200
PowerExchange for Microsoft Azure Blob Storage. . . . 200
PowerExchange for Microsoft Azure Data Lake Storage Gen1. . . . 200
Security. . . . 200
infacmd isp Commands. . . . 201
LDAP Directory Services. . . . 201
LDAP Configurations. . . . 201
SAML Authentication. . . . 201

Part VI: Versions 10.2.2 - 10.2.2 HotFix 1 202

Chapter 24: Version 10.2.2 HotFix 1 203

Important Notices. . . . 203
Support Changes. . . . 203
What's New. . . . 204
Command Line Programs 204
Big Data Management. . . . 204
Enterprise Data Catalog. . . . 204
PowerExchange Adapters for Informatica. . . . 207
What's Changed. . . . 208
Data Transformation. . . . 208
Enterprise Data Preparation. . . . 208

Enterprise Data Catalog. . . . 209
PowerExchange Adapters for Informatica. . . . 210

Chapter 25: Version 10.2.2 Service Pack 1. . . . 211
Important Notices. . . . 211
Support Changes. . . . 211
Product and Service Name Changes. . . . 212
Release Tasks. . . . 212
What's New. . . . 212
Big Data Management. . . . 212
Big Data Streaming. . . . 213
Enterprise Data Catalog. . . . 214
Enterprise Data Preparation. . . . 214
PowerExchange Adapters for Informatica. . . . 214
What's Changed. . . . 215
Big Data Management 215
Big Data Streaming 215
Informatica Analyst. . . . 216
PowerExchange Adapters for Informatica. . . . 216

Chapter 26: 10.2.2 Important Notices. . . . 218
Installation and Configuration. . . . 218
OpenJDK. . . . 218
Installer Code Signing. . . . 218
Resume the Installer. . . . 219
Informatica Docker Utility. . . . 219
Installer. . . . 219
Support Changes. . . . 219
Hive Engine. . . . 219
Distribution Support. . . . 220
Support Changes for Big Data Management. . . . 220
Support Changes for Big Data Streaming. . . . 220
Universal Connectivity Framework in Enterprise Data Catalog. . . . 221
Release Tasks. . . . 221
Big Data Management. . . . 221
Big Data Streaming. . . . 223
PowerExchange Adapters for Informatica. . . . 223
New Products. . . . 224
PowerExchange Adapters for Informatica. . . . 224

Chapter 27: 10.2.2 What's New. . . . 225
Application Services. . . . 225
Mass Ingestion Service. . . . 225

Metadata Access Service. . . . 225
REST Operations Hub Service. . . . 226
Big Data Management. . . . 226
Azure Databricks Integration. . . . 226
Data Preview on the Spark Engine. . . . 227
Hierarchical Data. . . . 228
High Availability. . . . 228
Intelligent Structure Model. . . . 228
Mass Ingestion. . . . 229
Monitoring. . . . 229
Security. . . . 230
Targets. . . . 230
Big Data Streaming 230
Azure Event Hubs Data Objects. . . . 230
Cross-account IAM Role in Amazon Kinesis Connection. . . . 231
Intelligent Structure Model. . . . 231
Header Ports for Big Data Streaming Data Objects. . . . 231
AWS Credential Profile in Amazon Kinesis Connection. . . . 231
Spark Structured Streaming. . . . 231
Window Transformation. . . . 231
Command Line Programs. . . . 232
infacmd dis Commands. . . . 232
infacmd ihs Commands. . . . 233
infacmd ipc Commands. . . . 233
infacmd ldm Commands. . . . 233
infacmd mi Commands. . . . 234
infacmd ms Commands. . . . 235
infacmd oie Commands. . . . 235
infacmd tools Commands. . . . 235
infasetup Commands. . . . 236
Enterprise Data Catalog. . . . 236
Automatically Assign Business Title to a Column. . . . 236
User Collaboration on Assets. . . . 236
Create Enterprise Data Catalog Application Services Using the Installer. . . . 237
Custom Metadata Validation Utility. . . . 237
Change Notifications. . . . 237
Business Glossary Assignment Report. . . . 237
Operating System Profiles. . . . 238
REST APIs. . . . 238
Source Metadata and Data Profile Filter. . . . 238
Scanner Utility. . . . 238
Resource Types. . . . 238

Enterprise Data Lake. . . . 239
Apply Active Rules. . . . 239
Delete Duplicate Rows. . . . 239
Cluster and Categorize Column Data 239
CLAIRE-based Recommendations. . . . 240
Conditional Aggregation. . . . 240
Data Masking. . . . 240
Localization. . . . 240
Partitioned Sources and Targets. . . . 241
Add Comments to Recipe Steps. . . . 241
Save a Recipe as a Mapping. . . . 241
Amazon S3, ADLS, WASB, MapR-FS as Data Sources. . . . 241
Statistical Functions. . . . 241
Date and Time Functions. . . . 242
Math Functions. . . . 242
Text Functions. . . . 243
Window Functions. . . . 243
Purge Audit Events. . . . 244
Spark Execution Engine. . . . 244
Informatica Developer. . . . 244
Applications. . . . 244
Informatica Mappings. . . . 244
Data Types. . . . 244
Mapping Outputs. . . . 244
Mapping Parameters. . . . 245
Optimizer Levels. . . . 245
Sqoop. . . . 245
Informatica Transformations. . . . 246
Address Validator Transformation. . . . 246
Update Strategy Transformation. . . . 249
PowerExchange Adapters for Informatica. . . . 249
PowerExchange for Amazon Redshift. . . . 249
PowerExchange for Amazon S3. . . . 250
PowerExchange for Google BigQuery. . . . 250
PowerExchange for HBase. . . . 251
PowerExchange for HDFS. . . . 251
PowerExchange for Hive. . . . 252
PowerExchange for MapR-DB. . . . 252
PowerExchange for Microsoft Azure Blob Storage. . . . 252
PowerExchange for Microsoft Azure Cosmos DB SQL API. . . . 253
PowerExchange for Microsoft Azure Data Lake Store. . . . 253
PowerExchange for Microsoft Azure SQL Data Warehouse. . . . 254

PowerExchange for Salesforce. . . . 254
PowerExchange for Snowflake. . . . 255
PowerExchange for Teradata Parallel Transporter API. . . . 255

Chapter 28: 10.2.2 What's Changed. . . . 256

Application Services. . . . 256
Metadata Access Service. . . . 256
Big Data Management. . . . 256
Hive Connection. . . . 256
Mass Ingestion. . . . 257
Spark Monitoring. . . . 257
Sqoop. . . . 258
Transformations in the Hadoop Environment. . . . 258
Big Data Streaming. . . . 259
Big Data Streaming and Big Data Management Integration. . . . 259
Kafka Connection. . . . 259
Transformations. . . . 259
Enterprise Data Catalog. . . . 260
Java Development Kit Change. . . . 260
Enterprise Data Lake. . . . 260
MAX and MIN Functions. . . . 260
Informatica Developer. . . . 261
Informatica Developer Name Change. . . . 261
Informatica Transformations. . . . 261
Address Validator Transformation. . . . 261
Write Transformation. . . . 262
PowerExchange Adapters for Informatica. . . . 262
PowerExchange for Amazon Redshift. . . . 263
PowerExchange for Amazon S3. . . . 263
PowerExchange for Google Analytics. . . . 264
PowerExchange for Google Cloud Storage. . . . 264
PowerExchange for HBase. . . . 264
PowerExchange for HDFS. . . . 264
PowerExchange for Hive. . . . 264
PowerExchange for Microsoft Azure Blob Storage. . . . 264

Part VII: Version 10.2.1. . . . 265

Chapter 29: 10.2.1 What's New. . . . 266

Application Services. . . . 266
Content Management Service. . . . 266
Data Integration Service. . . . 267
Mass Ingestion Service. . . . 267

Metadata Access Service. . . . 267
Model Repository Service. . . . 267
Big Data Management. . . . 268
Blaze Engine Resource Conservation. . . . 268
Cluster Workflows. . . . 268
Cloud Provisioning Configuration. . . . 269
High Availability. . . . 269
Hive Functionality in the Hadoop Environment. . . . 269
Importing from PowerCenter. . . . 270
Intelligent Structure Model. . . . 270
Mass Ingestion. . . . 271
Monitoring. . . . 271
Processing Hierarchical Data on the Spark Engine. . . . 272
Rule Specification Support on the Spark Engine. . . . 273
Security. . . . 273
Sqoop. . . . 273
Transformation Support in the Hadoop Environment. . . . 274
Big Data Streaming. . . . 276
Sources and Targets. . . . 276
Stateful Computing in Streaming Mappings. . . . 277
Transformation Support. . . . 277
Truncate Partitioned Hive Target Tables. . . . 277
Command Line Programs. . . . 277
infacmd autotune Commands. . . . 277
infacmd ccps Commands 278
infacmd cluster Commands. . . . 278
infacmd cms Commands. . . . 279
infacmd dis Commands. . . . 279
infacmd ihs Commands 279
infacmd isp Commands. . . . 280
infacmd ldm Commands 280
infacmd mi Commands. . . . 280
infacmd mrs Commands. . . . 281
infacmd wfs Commands. . . . 281
infasetup Commands. . . . 281
Enterprise Data Catalog. . . . 282
Adding a Business Title to an Asset. . . . 282
Cluster Validation Utility in Installer. . . . 282
Data Domain Discovery Types. . . . 282
Filter Settings. . . . 282
Missing Links Report. . . . 283
New Resource Types. . . . 283

REST APIs. . . . 283
SAML Authentication for Enterprise Data Catalog Applications. . . . 283
SAP Resource. . . . 284
Import from ServiceNow. . . . 284
Similar Columns. . . . 284
Specify Load Types for Catalog Service. . . . 284
Supported Resource Types for Data Discovery. . . . 284
Enterprise Data Lake. . . . 285
Column Data. . . . 285
Manage Data Lake Resources. . . . 285
Data Preparation Operations. . . . 285
Prepare JSON Files. . . . 286
Recipe Steps. . . . 286
Schedule Export, Import, and Publish Activities. . . . 286
Security Assertion Markup Language Authentication. . . . 286
View Project Flows and Project History. . . . 287
Informatica Developer. . . . 287
Default Layout. . . . 287
Editor Search. . . . 288
Import Session Properties from PowerCenter 288
Views. . . . 289
Informatica Mappings. . . . 289
Dynamic Mappings. . . . 289
Mapping Parameters. . . . 290
Running Mappings. . . . 291
Truncate Partitioned Hive Target Tables. . . . 292
Informatica Transformation Language. . . . 292
Complex Functions for Map Data Type. . . . 292
Complex Operator for Map Data Type. . . . 293
Informatica Transformations. . . . 293
Address Validator Transformation. . . . 293
Informatica Workflows. . . . 295
Import a Command Task from PowerCenter 296
PowerExchange Adapters for Informatica. . . . 296
PowerExchange for Amazon Redshift. . . . 296
PowerExchange for Amazon S3. . . . 296
PowerExchange for Cassandra. . . . 297
PowerExchange for HBase. . . . 297
PowerExchange for HDFS. . . . 297
PowerExchange for Hive. . . . 297
PowerExchange for Microsoft Azure Blob Storage. . . . 298
PowerExchange for Microsoft Azure SQL Data Warehouse. . . . 298

PowerExchange for Salesforce. . . . 298
PowerExchange for SAP NetWeaver. . . . 299
PowerExchange for Snowflake. . . . 299
Security. . . . 299
Password Complexity. . . . 299

Chapter 30: 10.2.1 What's Changed. . . . 300

Support Changes. . . . 300
Upgrade Support Changes. . . . 300
Big Data Hadoop Distribution Support. . . . 301
Hive Run-Time Engine . . . 303
Installer Changes. . . . 303
Product Name Changes. . . . 305
Application Services. . . . 305
Model Repository Service. . . . 305
Big Data Management. . . . 305
Azure Storage Access. . . . 306
Configuring the Hadoop Distribution. . . . 306
Developer Tool Configuration. . . . 306
Hadoop Connection Changes . . . 307
Hive Connection Properties. . . . 308
Monitoring. . . . 309
Precision and Scale on the Hive Engine. . . . 309
Sqoop. . . . 310
Transformation Support on the Hive Engine. . . . 310
Big Data Streaming. . . . 310
Configuring the Hadoop Distribution. . . . 310
Developer Tool Configuration. . . . 311
Kafka Connection Properties. . . . 311
Command Line Programs . . . 311
Content Installer. . . . 312
Enterprise Data Catalog . . . 312
Additional Properties Section in the General Tab. . . . 313
Connection Assignment. . . . 313
Column Similarity. . . . 313
Create a Catalog Service. . . . 313
HDFS Resource Type Enhancemets. . . . 313
Hive Resources. . . . 313
Informatica Platform Scanner. . . . 314
Overview Tab. . . . 314
Product Name Changes . . . 314
Proximity Data Domains. . . . 314
Search Results. . . . 314

Universal Connectivity Framework. . . . 315
Informatica Analyst . . . 315
Scorecards. . . . 315
Informatica Developer. . . . 315
Importing and Exporting Objects from and to PowerCenter. . . . 315
Informatica Transformations. . . . 316
Address Validator Transformation. . . . 316
Data Transformation . . . 316
Sequence Generator Transformation. . . . 316
Sorter Transformation. . . . 317
PowerExchange Adapters for Informatica. . . . 317
PowerExchange for Amazon Redshift . . . 317
PowerExchange for Cassandra. . . . 317
PowerExchange for Snowflake. . . . 318

Chapter 31: 10.2.1 Release Tasks. . . . 319

PowerExchange Adapters for Amazon S3. . . . 319

Part VIII: Version 10.2 - 10.2 HotFix 2. . . . 320

Chapter 32: Version 10.2 HotFix 2. . . . 321

Important Notices. . . . 321
Support Changes. . . . 321
New Products . . . 322
Release Tasks. . . . 322
What's New. . . . 323
Command Line Programs. . . . 323
Informatica Analyst. . . . 324
Informatica Transformations. . . . 324
Metadata Manager. . . . 326
PowerCenter. . . . 326
PowerExchange Adapters for PowerCenter. . . . 327
Security. . . . 329
What's Changed. . . . 329
Analyst Tool. . . . 329
Command Line Programs. . . . 330
Informatica Transformations. . . . 330
Metadata Manager. . . . 330
PowerCenter. . . . 331
PowerExchange Adapters for Informatica. . . . 331
PowerExchange Adapters for PowerCenter. . . . 331

Chapter 33: Version 10.2 HotFix 1. 333

What's New. 333

Application Services. 333

Business Glossary 333

Command Line Programs. 333

Connectivity. 334

Data Types. 334

Installer. 335

Informatica Transformations. 335

Metadata Manager. 339

PowerCenter. 339

PowerExchange Adapters for Informatica. 340

PowerExchange Adapters for PowerCenter. 341

Security. 344

What's Changed. 344

Support Changes. 344

Application Services. 345

Big Data Management. 346

Business Glossary. 346

Documentation 346

Informatica Development Platform. 347

Informatica Transformations. 347

PowerCenter. 347

PowerExchange Adapters for PowerCenter. 348

Reference Data. 349

Release Tasks. 350

PowerExchange Adapters for Informatica. 350

PowerExchange Adapters for PowerCenter. 351

Chapter 34: 10.2 Important Notices. 352

New Products. 352

PowerExchange Adapters for Informatica. 352

Release Tasks. 352

PowerExchange Adapters for PowerCenter. 352

Chapter 35: 10.2 What's New. 355

Application Services. 355

Model Repository Service. 355

Big Data 356

Big Data Management Installation. 356

Cluster Configuration. 356

Processing Hierarchical Data. 356

Stateful Computing on the Spark Engine. . . . 357
Data Integration Service Queuing. . . . 357
Blaze Job Monitor. . . . 357
Data Integration Service Properties for Hadoop Integration. . . . 358
Sqoop. . . . 358
Autoscaling in an Amazon EMR Cluster. . . . 358
Transformation Support on the Blaze Engine. . . . 359
Hive Functionality for the Blaze Engine. . . . 359
Transformation Support on the Spark Engine. . . . 359
Hive Functionality for the Spark Engine. . . . 359
Command Line Programs. . . . 359
infacmd cluster Commands. . . . 359
infacmd dis Options. . . . 360
infacmd ipc Commands. . . . 361
infacmd isp Commands. . . . 361
infacmd mrs Commands. . . . 365
infacmd ms Commands. . . . 366
infacmd wfs Commands. . . . 366
infasetup Commands. . . . 366
pmrep Commands. . . . 367
Data Types. . . . 367
Informatica Data Types. . . . 368
Documentation. . . . 368
Enterprise Information Catalog. . . . 369
New Data Sources. . . . 369
Custom Scanner Framework. . . . 370
REST APIs. . . . 370
Composite Data Domains. . . . 370
Data Domains. . . . 370
Export and Import of Custom Attributes. . . . 371
Rich Text as Custom Attribute Value. . . . 371
Transformation Logic . . . 371
Unstructured File Types. . . . 372
Value Frequency. . . . 372
Deployment Support for Azure HDInsight. . . . 372
Informatica Analyst. . . . 372
Profiles. . . . 373
Intelligent Data Lake. . . . 373
Validate and Assess Data Using Visualization with Apache Zeppelin. . . . 373
Assess Data Using Filters During Data Preview. . . . 373
Enhanced Layout of Recipe Panel. . . . 373
Apply Data Quality Rules. . . . 374

View Business Terms for Data Assets in Data Preview and Worksheet View. . . . 374
Prepare Data for Delimited Files. . . . 374
Edit Joins in a Joined Worksheet. . . . 374
Edit Sampling Settings for Data Preparation. . . . 374
Support for Multiple Enterprise Information Catalog Resources in the Data Lake. . . . 374
Use Oracle for the Data Preparation Service Repository. . . . 374
Improved Scalability for the Data Preparation Service. . . . 375
Informatica Developer. . . . 375
Nonrelational Data Objects. . . . 375
Profiles. . . . 375
Informatica Installation. . . . 375
Informatica Upgrade Advisor. . . . 375
Intelligent Streaming. . . . 375
CSV Format. . . . 376
Data Types. . . . 376
Connections. . . . 376
Pass-Through Mappings. . . . 376
Sources and Targets. . . . 376
Transformation Support. . . . 376
Metadata Manager. . . . 377
Cloudera Navigator. . . . 377
PowerCenter. . . . 377
PowerExchange Adapters. . . . 378
PowerExchange Adapters for Informatica. . . . 378
PowerExchange Adapters for PowerCenter. . . . 380
Rule Specifications. . . . 382
Security. . . . 382
User Activity Logs. . . . 383
Transformation Language. . . . 383
Informatica Transformation Language. . . . 383
Transformations. . . . 384
Informatica Transformations. . . . 384
PowerCenter Transformations. . . . 388
Workflows. . . . 388
Informatica Workflows. . . . 388

Chapter 36: 10.2 What's Changed. . . . 390

Support Changes. . . . 390
Big Data Hadoop Distribution Support. . . . 391
Metadata Manager. . . . 393
Application Services. . . . 393
Content Management Service. . . . 394
Data Integration Service. . . . 394

Big Data. 394
Hadoop Connection. 395
HBase Connection Properties 397
Hive Connection Properties. 397
HBase Connection Properties for MapR-DB 398
Mapping Run-time Properties. 399
Monitoring. 399
S3 Access and Secret Key Properties. 399
Sqoop. 400
Command Line Programs. 400
Enterprise Information Catalog. 401
Product Name Changes. 401
Informatica Analyst. 401
Parameters. 401
Intelligent Streaming. 401
Kafka Data Object Changes. 401
PowerExchange Adapters. 402
PowerExchange Adapters for Informatica. 402
PowerExchange Adapters for PowerCenter. 402
Security. 404
SAML Authentication. 404
Transformations. 404
Informatica Transformations. 404
Workflows. 405
Informatica Workflows. 406

Preface

See *Informatica® What's New and Changed* to learn about new functionality and changes for current and recent product releases. You can also read the important notices for information related to support changes, new products, and release tasks that you might need to perform after you upgrade from a previous version. *What's New and Changed* includes content for all products integrated in the Informatica platform installation.

Informatica Resources

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

Informatica Network

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

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

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

Informatica Knowledge Base

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

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

Informatica Documentation

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

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

Informatica Product Availability Matrices

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

Informatica Velocity

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

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

Informatica Marketplace

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

Informatica Global Customer Support

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

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link:
https://www.informatica.com/services-and-training/customer-success-services/contact-us.html.

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

Part I: Version 10.5.2 - 10.5.2.1.x

This part contains the following chapters:

- Version 10.5.2.1 - 10.5.2.1.x, 31
- 10.5.2 Important Notices, 35
- 10.5.2 What's New, 38
- 10.5.2 What's Changed, 45

CHAPTER 1

Version 10.5.2.1 - 10.5.2.1.x

This chapter includes the following topics:

- Version 10.5.2.1.2, 31
- Version 10.5.2.1.1, 32
- Version 10.5.2.1, 32

Version 10.5.2.1.2

Read this section to learn what's new and changed for Enterprise Data Catalog in version 10.5.2.1.2.

What's New

MetaDex repository property

Effective in version 10.5.2.1.2, you can configure the scanners.respository.askBeforeUpdate property to prompt before updating the MetaDex repository database if there are new repository changes required.

For more information, see the *MetaDex Scanners Administrator Guide*.

Asset statistics

Effective in version 10.5.2.1.2, you can view statistics in the form of asset names and numbers for MetaDex scanners such as Snowflake that don't generate lineage from assets.

For more information, see the *MetaDex Scanners Administrator Guide*.

Informatica Intelligent Cloud Services MetaDex scanner extracts from Business 360

Effective in version 10.5.2.1.2, you can use the Informatica Intelligent Cloud Services MetaDex scanner to perform the following types of Business 360 extract:

- Extract metadata from business entities.
- Extract lineage when you use a business entity as a target in a Cloud Data Integration mapping.

For more information, see the *MetaDex Scanners Configuration Guide*.

What's Changed

Microsoft SQL Server Reporting Services web service for UNIX

Effective in version 10.5.2.1.2, if you use UNIX, you can connect to the Microsoft SQL Server Reporting Services (SSRS) server directly from its web service URL. You no longer have to use the export utility code to extract SSRS reports from the SSRS server.

For more information, see the *MetaDex Scanners Configuration Guide*.

Version 10.5.2.1.1

Read this section to learn what's new for Enterprise Data Catalog in version 10.5.2.1.1.

IBM InfoSphere DataStage MetaDex scanner supports HiveConnectorPX Stage

Effective in version 10.5.2.1.1, you can use the HiveConnectorPX Stage for the IBM InfoSphere DataStage MetaDex scanner.

For more information, see the *MetaDex Scanner Reference*.

JCL MetaDex scanner supports new JCL programs

Effective in version 10.5.2.1.1, the JCL MetaDex scanner supports the following JCL programs: BTEQ and MLOAD.

For more information, see the *MetaDex Scanner Reference*.

Enhanced Microsoft SQL Server Integration Service MetaDex scanner lineage

Effective in version 10.5.2.1.1, the Microsoft SQL Server Integration Service MetaDex scanner lineage is enhanced for generic conversion of components that use synchronous outputs.

MetaDex repository on Amazon Relational Database Service

Effective in 10.5.2.1.1, you can create a MetaDex repository on Amazon Relational Database Service for SQL Server and Amazon Relational Database Service for Oracle.

Version 10.5.2.1

Important Notices (10.5.2.1)

Read this section to learn about important notices in version 10.5.2.1.

Technical Preview Initiated

Effective in version 10.5.2.1, Informatica includes the following functionality for technical preview:

Launch Data Preparation from the Enterprise Data Catalog tool

You can seamlessly launch Data Preparation of Informatica Intelligent Cloud Services for relational tables and views from the Enterprise Data Catalog tool. You can cleanse, transform, and structure the datasets present in the catalog using Data Preparation. For more information, see the *Informatica Enterprise Data Catalog User Guide*.

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

What's New (10.5.2.1)

Read this section to learn what's new in version 10.5.2.1.

Data Privacy Management

Read this section to learn what's new for Data Privacy Management in version 10.5.2.1.

Snowflake Advanced Scanner

Effective in version 10.5.2.1, you can use the advanced scanner for Snowflake with Data Privacy Management. You can use the scanner to import and view domain discovery scan results and to run subject registry scans in Data Privacy Management.

For information about how to use the advanced scanner with Data Privacy Management, see the How-To article: *Using the Snowflake Advanced Scanner in Data Privacy Management*.

User behavior activity logs

Effective in version 10.5.2.1, the log messages for system jobs related to UBA applications such as Augmentor, Percolator, and UBA Manager, include the URL of the Spark user interface. Use the URL to view the logs for all UBA Spark applications directly from Data Privacy Management.

Enterprise Data Catalog

Read this section to learn what's new for Enterprise Data Catalog in version 10.5.2.1.

Relational resource enhancements

Effective in version 10.5.2.1, you can configure the data preparation settings for relational resources in the Catalog Administrator tool. For more information, see the *Informatica Catalog Administration Guide*.

Direct read for data profiling

Effective in version 10.5.2.1, you can configure direct read for an Amazon Redshift resource for data profiling. You can enable direct read using custom options during resource configuration. For more information, see the *Enterprise Data Catalog Scanner Configuration Guide*.

Profiling and data domain discovery on external tables for Amazon Redshift resources

Effective in version 10.5.2.1, you can run a profile and perform data domain discovery on external tables for an Amazon Redshift resource.

Schedule enhancements

Effective in version 10.5.2.1, you can view the server time zone when you configure a resource in the Catalog Administrator tool. You can view the server time zone for both custom and reusable schedules.

For more information, see the *Informatica Catalog Administration Guide*.

Data domains in search results

Effective in version 10.5.2.1, the **Search Results** page displays data domains in the asset details. For more information, see the *Informatica Enterprise Data Catalog User Guide*.

Opening primary keys and foreign keys of related tables from the Relationships view

Effective in version 10.5.2.1, you can open the primary keys and foreign keys of related tables from the Relationships view of a tabular asset.

For more information on foreign keys, primary keys, and the Relationships view, see the *Informatica Enterprise Data Catalog User Guide*.

Lineage enhancements

Effective in version 10.5.2.1, the diagram or compact view in the **Lineage and Impact** tab includes the following enhancements:

- You can view the control flow between the assets in the lineage diagram.

- You can search for specific assets or objects in the lineage diagram using the find text box.
- You can configure the direction of the data flow for an asset in the lineage diagram.

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

Business term association

Effective in version 10.5.2.1, you can associate business terms to assets from a Microsoft Power BI V2 resource in the catalog. For more information on business term association, see the *Informatica Enterprise Data Catalog User Guide*.

ADO NET

Effective in version 10.5.2.1, you can extract metadata from the ADO NET destination component with the Microsoft SQL Server Integration Services MetaDex scanner. For more information, see the *MetaDex Scanner Configuration Guide*.

Amazon S3 Cloud Storage

Effective in version 10.5.2.1, you can process file locations of the Amazon S3 Cloud Storage type with the SAP BODS MetaDex scanner.

What's Changed (10.5.2.1)

Read this section to learn what's changed in version 10.5.2.1.

Data Privacy Management

Read this section to learn what's changed for Data Privacy Management in version 10.5.2.1.

Include views in Snowflake data store scans

Effective in version 10.5.2.1, you can configure a Snowflake data store to include views in scan jobs. For information about how to configure a Snowflake data store, see the *Informatica Data Privacy Management User Guide*.

Enterprise Data Catalog

Read this section to learn what's changed for Enterprise Data Catalog in version 10.5.2.1.

Reference assets in the Missing Links report

Effective in version 10.5.2.1, you can generate the **Missing Links** report and identify the missing connection links that are resolved to reference assets.

Previously, you could generate the **Missing Links** report and identify only the connection links that are missing after you assign schemas from a resource to connections.

For more information on the **Missing Links** report, see the *Informatica Catalog Administration Guide*.

Data Provisioning tab renamed

Effective in version 10.5.2.1, the Data Provisioning tab in the Catalog Administrator tool is renamed to Data Access. For more information on the Data Access tab, see the *Informatica Catalog Administration Guide*.

CHAPTER 2

10.5.2 Important Notices

This chapter includes the following topics:

- Support Changes, 35
- Release Tasks, 36
- Informatica Installer , 36
- Enterprise Data Catalog, 37

Support Changes

Read this section to learn about support changes in version 10.5.2.

Technical Preview

Read this section to learn about technical preview functionality that is initiated or lifted in version 10.5.2.

Technical Preview Initiated

Effective in version 10.5.2, Informatica includes the following functionality for technical preview:

V2 MetaDex scanners and Enterprise Data Catalog resources

You can extract metadata from the following V2 metadata sources and resources:

- Informatica Intelligent Cloud Services - Cloud Data Integration V2 MetaDex scanner
  For more information, see the *MetaDex Scanner Configuration Guide*.
- Salesforce V2 MetaDex scanner
  For more information, see the *MetaDex Scanner Configuration Guide*.
- Amazon S3 V2 resource
  The resource includes implementation and performance enhancements such as custom partition detection and partition pruning.

  For more information, see the *Enterprise Data Catalog Scanner Configuration Guide.*
- Azure Data Lake Storage Gen2 V2 resource.
  The resource includes implementation and performance enhancements such as custom partition detection and partition pruning.

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

Technical preview functionality is supported for evaluation purposes but is unwarranted and is not supported in production environments or any environment that you plan to push to production. Informatica intends to include the preview functionality in an upcoming release for production use, but might choose not to in

accordance with changing market or technical circumstances. For more information, contact Informatica Global Customer Support.

Technical Preview Lifted

Effective in version 10.5.2, Informatica lifted the following functionality from technical preview:

PowerExchange for Microsoft Azure Data Lake Storage Gen2

Effective in version 10.5.2, Informatica lifted the following functionality from technical preview:

- You can read objects stored in subdirectories in Microsoft Azure Data Lake Storage Gen2 in the native environment or on the Spark engine.
- You can use wildcard characters to specify the source directory name or the source file name when you read data from a flat file or complex file in the native environment or on the Spark engine.

Dropped Support

Read this section to learn what support is dropped for Enterprise Data Catalog.

Resources

Effective in version 10.5.2, Enterprise Data Catalog dropped support for the following resource versions that were lower than 10.2:

- Informatica Platform
- PowerCenter
- Informatica Data Quality

Release Tasks

Read this section to learn about tasks that you need to perform after you upgrade to version 10.5.2.

PowerExchange for Microsoft Dynamics CRM

Microsoft deprecated the WS Trust protocol for Microsoft Dynamics CRM and MSD CRM Online. Effective in 10.5.2, you must use the OAuth authentication for the passport authentication type to access Microsoft Dynamics CRM Online.

After you upgrade to 10.5.2, mappings fail when you use the password authentication type without the client ID and client secret to access Microsoft Dynamics CRM Online. You must update the Microsoft Dynamics CRM connection to OAuth and use the CRM Organization web service URL.

Note: You can continue to use the Claims-based and Active Directory authentication types to connect to the on-premise Microsoft Dynamics CRM.

Informatica Installer

Read this section to learn what's new for the Informatica installer in version 10.5.2.

Informatica Deployment Manager

Effective in version 10.5.2, you can use Informatica Deployment Manager to install Informatica Data Quality on Azure Kubernetes Service (AKS) clusters. Additionally, you can use Informatica Deployment Manager to manage existing domains and nodes that are running on Kubernetes for Informatica Data Quality.

For more information, see *Install Data Quality on Kubernetes With Informatica Deployment Manager 10.5.x* How-to Library (H2L) article.

Azure PostgreSQL

Effective in version 10.5.2, you can use the Azure PostgreSQL database for the domain configuration repository, the Model repository, and the PowerCenter repository.

For more information about Azure PostgreSQL, see the Informatica installation guides and the *Informatica Application Service Guide*.

Enterprise Data Catalog

Read this section to learn more about the name change in the 10.5.2 version of Enterprise Data Catalog.

MetaDex Component Names Updated

Effective in version 10.5.2, Informatica updated the component names within the MetaDex framework to better reflect the usage.

The following table lists updates to component names:

Previous name	Updated name
Advanced Scanners tool	MetaDex tool
advanced scanner	MetaDex scanner
Advanced Scanners server	MetaDex server
Advanced Scanners repository	MetaDex repository

For more information about the update and usage, see the FAQs on the name update and usage of MetaDex scanners Knowledge Base (KB) article.

CHAPTER 3

10.5.2 What's New

This chapter includes the following topics:

- Data Engineering, 38
- Data Privacy Management, 39
- Data Quality, 39
- Data Transformation, 39
- Enterprise Data Catalog, 40
- Informatica Installer, 43
- PowerExchange Adapters, 43

Data Engineering

Read this section to learn what's new for Data Engineering in version 10.5.2.

Databricks

Read this section to learn about new features that support Databricks functionality.

Delta Lake Schema Evolution

Effective in version 10.5.2, you can configure dynamic mappings to apply Databricks source schema changes to Delta Lake targets. You can choose from among three different strategies to update Delta Lake target tables when a source schema changes. Mappings succeed with merge enabled in the mapping and schema enforcement disabled in the Databricks workspace.

For more information, see the following documentation:

- Databricks documentation.
- *Data Engineering Integration User Guide*

Use Databricks and Custom Parameters to Configure Ephemeral Clusters

Effective in version 10.5.2, you can use custom parameters and Databricks parameters in a JSON file to configure ephemeral clusters when you use a cluster creation workflow. You can use the JSON file, which is easily edited, instead of specifying parameter values in the Create Cluster task properties.

For more information, see cluster workflow documentation in the *Data Engineering Integration User Guide*.

Unmanaged Databricks Table Targets

Effective in version 10.5.2, you can configure mappings to write to unmanaged Databricks table targets.

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

Data Transformation

Effective in version 10.5.2, you can configure the ExcelToDataXml document processor to add elements to the output when a source table contains empty cells in the middle or at the end of table rows.

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

Kudu Dynamic Mapping

Effective in version 10.5.2, you can enter a comma-separated list of column names that you want to be primary keys when you create a Kudu object using the dynamic mapping flow option.

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

Data Privacy Management

Read this section to learn what's new for Data Privacy Management in version 10.5.2.

Subject Registry

Effective in version 10.5.2, you can use a Snowflake data store as a golden data store for a Subject Registry.

For information about how to create a Subject Registry, see the *Informatica Data Privacy Management Administrator Guide*.

Data Quality

Read this section to learn what's new for Data Quality in version 10.5.2.

Human Tasks and Exception Management

Effective in version 10.5.2, you can read and write Human task data in a PostgreSQL database, and you can create an exception management audit database on PostgreSQL. Use a JDBC connection to connect to a Human task database and an exception management audit database on PostgreSQL.

For more information, see the *Informatica 10.5.2 Application Service Guide*.

Data Transformation

Read this section to learn what's new for Data Transformation in version 10.5.2.

Data Transformation ExcelToDataXml document processor

When the ExcelToDataXml document processor reads a Microsoft Excel document that contains empty cells, the document processor includes the empty cells in the XML output.

For more information about the ExcelToDataXml document processor, see the *Data Transformation User Guide*.

Enterprise Data Catalog

Read this section to learn what's new for Enterprise Data Catalog in version 10.5.2.

Bulk certification of assets

Effective in version 10.5.2, resource stakeholders such as data owners, data stewards, and subject matter experts can certify or decertify assets from the resource in bulk.

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

Enrichment Migration Utility

Effective in version 10.5.2, you can use the Enrichment Migration Utility to migrate enrichments between resources of the same type. For example, from the Salesforce scanner V1 to the Salesforce MetaDex scanner V2. This utility is available in the MetaDex tool.

For more information, see the *MetaDex Scanner Configuration Guide*.

MetaDex Job Reports

Effective in version 10.5.2, you can generate Microsoft Excel reports that contain logs and objects for jobs that you run in the MetaDex tool.

For more information, see the *Enterprise Data Catalog 10.5.2 MetaDex Scanner Configuration Guide*.

Scanners

Effective in version 10.5.2, you can extract metadata from Microsoft Azure Synapse Analytics. You configure the Microsoft Azure Synapse Analytics MetaDex scanner in the MetaDex tool.

Effective in version 10.5.2, you can also extract metadata from data sources with the following standard scanners that you configure in the MetaDex tool:

- Informatica Intelligent Cloud Services - Cloud Data Integration V2
- Salesforce V2

Note: This functionality is available as technical preview. Technical preview functionality is supported for evaluation purposes but is unwarranted and is not supported in production environments or any environment that you plan to push to production. Informatica intends to include the preview functionality in an upcoming release for production use, but might choose not to in accordance with changing market or technical circumstances. For more information, contact Informatica Global Customer Support.

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

Profiling Enhancements

Read this section to learn what's new for profiling in Enterprise Data Catalog.

Profile Avro files

You can extract metadata, discover Avro partitions, and run profiles on Avro files with multiple-level hierarchy. You can run profiles on Avro files for the following types of resources:

- HDFS
- Microsoft Azure Data Lake Storage Gen2
- Amazon S3

You can run profiles on Avro files using HDFS, Microsoft Azure Data Lake Storage Gen2, and Amazon S3 resources on the Spark engine.

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

Profile Google BigQuery resource with reserved keywords

You can run profiles on the Google BigQuery resource with reserved keywords in the Google BigQuery table.

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

Use Amazon S3 and AWS Databricks Delta tables to run profiles

You can use Amazon S3 and AWS Databricks Delta tables to run column profiles and discover data domains on both native and AWS Databricks run-time environments.

External PostgreSQL database to improve performance of Similarity Discovery resources

You can configure an external PostgreSQL database to improve the performance of Informatica Similarity Discovery resources. You can choose to either install the PostgreSQL database server bundled with the Enterprise Data Catalog installer or configure an external PostgreSQL database. You can configure the external PostgreSQL database after installation of Enterprise Data Catalog.

For more information, see the KB article *How To: Configure an external PostgreSQL database for similarity profiling*.

Resource Enhancements

Read this section to learn what's new for resources in Enterprise Data Catalog.

SAP S/4HANA Resource

You can enable the streaming mode for data discovery when accessing data from the SAP S/4HANA resource.

Axon and Business Glossary Resources

Effective in version 10.5.2, you can view the hierarchy of a business term associated with a technical asset for Axon and Business Glossary resources.

Amazon S3 V2 and Azure Data Lake Storage Gen2 V2 Resources

The Amazon S3 V2 and Azure Data Lake Storage Gen2 V2 resources include the following enhancements:

Partitioned file configuration

You can configure the resources to eliminate non relevant partitions, and detect custom partitions in the data source.

Exclusion filter

You can configure the resources to exclude folders or directories recursively from the data source or the selected first level directories.

Note: This functionality is available as technical preview. Technical preview functionality is supported for evaluation purposes but is unwarranted and is not supported in production environments or any environment that you plan to push to production. Informatica intends to include the preview functionality in an upcoming release for production use, but might choose not to in accordance with changing market or technical circumstances. For more information, contact Informatica Global Customer Support.

Proxy server connection

You can configure the following resources to extract metadata using a proxy server:

- Azure Data Lake Storage Gen2
- Amazon S3
- Microsoft Sharepoint
- Azure Microsoft SQL Server
- Azure Microsoft SQL Data Warehouse
- Google BigQuery
- Workday
- Google Cloud Storage
- IICS-CDI

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

REST APIs

Effective in version 10.5.2, you can use the following Informatica Enterprise Data Catalog REST APIs:

- Security Info REST API. You can list the access permissions on resources for specific users and groups.
- DAA Dashboard APIs. You can view the chart data and hero stats data in Data Asset Analytics.

For more information about the REST APIs, see the *Informatica 10.5.2 Enterprise Data Catalog REST API Reference*.

Search Results

Effective in version 10.5.2, the **Search Results** page now displays composite data domains in the asset details.

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

Reverse Proxy Support

Informatica Enterprise Data Catalog supports reverse proxy configuration.

View foreign key relationships

Effective in version 10.5.2, you can view details of foreign keys for assets such as tables, views, and files in the Enterprise Data Catalog tool.

You can also view the relationship between the asset and the foreign keys. For more information, see the *Informatica Enterprise Data Catalog User Guide*.

Informatica Installer

This section describes new installation and upgrade features in 10.5.2.

Azure PostgreSQL

Effective in version 10.5.2, you can use the Azure PostgreSQL database for the domain configuration repository, the Model repository, and the PowerCenter repository.

For more information about Azure PostgreSQL, see the Informatica installation guides and the *Informatica Application Service Guide*.

PowerExchange Adapters

Read this section to learn what's new for PowerExchange adapters in version 10.5.2.

PowerExchange Adapters for Informatica

Read this section to learn what's new for Informatica PowerExchange adapters in version 10.5.2.

PowerExchange for Amazon S3

Effective in version 10.5.2, you can configure a proxy server to connect to Amazon S3 to run a mapping on the Databricks Spark engine.

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

PowerExchange for Amazon Redshift

Effective in version 10.5.2, you can use Amazon Redshift Spectrum to directly run queries and read large Amazon S3 data files without the need to load or transform the data.

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

PowerExchange for Kudu

Effective in version 10.5.2, you can configure dynamic mappings to change Kudu targets at run time based on parameters and rules that you define.

When you configure dynamic mappings, you can also create or replace the target at run time. You can select the **Create** or **replace** table at run time option in the advanced properties of the Kudu data object write operation.

For more information, see the *Informatica 10.5.2 PowerExchange for Kudu User Guide*.

PowerExchange for Microsoft Azure Data Lake Storage Gen2

Effective in version 10.5.2, PowerExchange for Microsoft Azure Data Lake Storage Gen2 includes the following features:

- You can use the shared key authentication to connect to Microsoft Azure Data Lake Storage Gen2 using the account name and account key.
- You can use access control lists to grant different levels of permissions to access directories and files to each user and service. If you do not want to use role-based access control to grant access to all of the data in a storage account, you can use access control lists to grant read, write, and execute permissions to a specific directory or file.

For more information, see the *Informatica 10.5.2 PowerExchange for Microsoft Azure Data Lake Storage Gen2 User Guide*.

PowerExchange for Microsoft Azure SQL Data Warehouse

Effective in version 10.5.2, you can use a Microsoft Azure Data Warehouse connection to access a Microsoft Azure Data Warehouse endpoint in the China region.

For more information, see the *Informatica 10.5.2 PowerExchange for Microsoft Azure SQL Data Warehouse User Guide*.

PowerExchange Adapters for PowerCenter

Read this section to learn what's new for PowerCenter adapters in version 10.5.2.

PowerExchange for MS Dynamics CRM

Effective in version 10.5.2, you must use OAuth authentication for the passport authentication type to connect to the MS Dynamics CRM connector.

For more information, see the *Informatica 10.5.2 PowerExchange for MS Dynamics CRM User Guide for PowerCenter*.

PowerExchange for PostgreSQL

Effective in version 10.5.2, you can use the pre-SQL and post-SQL queries to call a stored procedure in a PostgreSQL database.

For more information, see the Informatica 10.5.2 PowerExchange for PostgreSQL User Guide for PowerCenter.

CHAPTER 4

10.5.2 What's Changed

This chapter includes the following topics:

- Data Engineering, 45
- Data Privacy Management, 46
- Enterprise Data Catalog, 47
- Informatica Developer, 48
- PowerCenter, 48

Data Engineering

Read this section to learn what's changed for Data Engineering in version 10.5.2.

Time zone information in logs

Effective in version 10.5.2, the log event time stamp of the log entries consists of date and time followed by time zone information in all the logs.

Previously, the log event time stamp consists of date and time only.

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

User Authentication Error Logging

Effective in version 10.5.2, authentication error messages and the location of detailed information about user authentication failures have changed in the following ways:

The User Activity tab in the Administrator tool contains a log of user actions including logins from clients. Effective in version 10.5.2, the log contains enhanced information about failed attempts to log into web applications, including the Administrator tool, the Developer tool, and the Analyst tool.

Previously, information about user authentication failures appeared in error messages.

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

Data Privacy Management

Read this section to learn what's changed for Data Privacy Management in version 10.5.2.

Data Store Connectivity

Data Privacy Management version 10.5.2 includes changes in connectivity to the following data stores:

OneDrive

Effective in version 10.5.2, you can create a data store to run scans on multiple user accounts in a tenant. Configure the data store with tenant details to include multiple user accounts in the data store and scans. Enter user account details to include a single account. To scan multiple user accounts, the data store must use a remote agent.

Previously, you could create a data store to scan only single user accounts.

SharePoint

Effective in version 10.5.2, you can create a data store to run scans on multiple user accounts in a tenant. Configure the data store with tenant details to include a tenant in the data store and scans. To scan multiple user accounts, the data store must use a remote agent. Use filters to further include or exclude specific sites. Enter user account details to include a single account.

Previously, you could create a data store to scan only single user accounts.

For information about how to configure data store properties, see the *Informatica Data Privacy Management User Guide*.

Risk Score Information

Effective in version 10.5.2, risk score information includes the following changes:

View the maximum risk score

You can configure risk score settings to display the maximum risk score in addition to the average risk score. Previously, you could not change the default setting that includes only the average risk score.

Access the Risk Score Details page from risk score column values

You can click the risk score value that appears in the Risk Score, Average Risk Score, and Max Risk Score columns, to open the Risk Score Details page. Previously, you could not open the Risk Score Details page from the column values.

View policy names on the Risk Score Details page

The Risk Score Details page includes the name of each classification policy with the risk score associated with it. Previously, you could not view the policy name with the risk score value.

For information about risk scores and how to view the maximum risk score, see the *Informatica Data Privacy Management User Guide*.

Default Dashboard Settings

Effective in version 10.5.2, Data Privacy Management changed the minimum requirements to change the default dashboard.

To change the default dashboard, a user must meet at least one of the following requirements:

- Belongs to the domain Administrator group.

- Is assigned the Administrator role.
- Belongs to a group that is assigned the Administrator role.

Previously, only users who belong to the domain Administrator group could change the default dashboard.

For information about how to change the default dashboard, see the *Informatica Data Privacy Management User Guide*.

Enterprise Data Catalog

Read this section to learn what's changed for Enterprise Data Catalog in version 10.5.2.

MetaDex Server Configuration File Change

Effective in version 10.5.2, the server.properties configuration file is merged with the scanners.properties file. Use the scanners.properties file for MetaDex scanners configuration.

For more information, see the *Enterprise Data Catalog 10.5.2 MetaDex Scanner Administrator Guide*.

MetaDex Server Memory

Effective in version 10.5.2, the default MetaDex server memory is 4 GB.

Previously, the default server memory was 2 GB.

Tableau Extension Download

Effective in version 10.5.2, you can download and install the Tableau extension from the **Help** menu in the Enterprise Data Catalog tool.

Previously you downloaded the extension from the **User** menu in the catalog header.

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

Enterprise Data Catalog Plug-in Download

Effective in version 10.5.2, you can download and install the Enterprise Data Catalog Plug-in extension from the **Help** menu in the Enterprise Data Catalog tool.

Previously you downloaded the plug-in from the Chrome web store.

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

MITI Windows Agent

Effective in version 10.5.2, the `Miti_Windows_Agent_Azul.zip` file that you downloaded from the Akamai link does not include the JDBC driver JAR file for the following resources:

- IBM DB2
- Oracle
- Microsoft SQL Server

Previously, the JDBC driver JAR files were included in the `Miti_Windows_Agent_Azul.zip` file.

Scanner Utility

Effective in version 10.5.2, the `eic-scannercli-utility.zip` file does not include the scanner binary files for offline resources.

Previously, the scanner utility .zip file included the scanner binary files for offline resources.

Informatica Developer

Read this section to learn what's changed for Informatica Developer in version 10.5.2.

Enabling Nullable Values

Effective in version 10.5.2, you can enable or disable all nullable columns within a relational or logical data object at one time.

Previously, you could only enable or disable the nullable columns one at a time.

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

PowerCenter

Read this section to learn what's changed for PowerCenter in version 10.5.2.

Time zone information in the logs

Effective in version 10.5.2, the log event time stamp of the log entries consists of date and time followed by time zone information in all the logs.

Previously, the log event time stamp consists of date and time only.

For more information, see the *Workflow Basics Guide*.

Part II: Version 10.5.1 - 10.5.1.1

This part contains the following chapters:

- Version 10.5.1.1, 50
- 10.5.1 Important Notices, 52
- 10.5.1 What's New, 56
- 10.5.1 What's Changed, 62

CHAPTER 5

Version 10.5.1.1

This chapter includes the following topics:

- What's New (10.5.1.1), 50
- What's Changed (10.5.1.1), 51

What's New (10.5.1.1)

Read this section to learn what's new in version 10.5.1.1.

Data Engineering Streaming

Read this section to learn what's new for Data Engineering Streaming in version 10.5.1.1.

Data Preview

Effective in version 10.5.1.1, you can preview data within a streaming mapping that runs on the Spark engine in the Developer tool. Previewing data helps to design and debug streaming mappings.

For information about data preview, see the *Data Engineering Streaming User Guide*.

Auto-scaling

Effective in version 10.5.1.1, you can enable auto-scale and run a streaming mapping on Databricks cluster. When you enable auto-scaling, the Databricks cluster automatically adds or subtracts cluster core and task nodes when Spark mappings activate performance thresholds.

For information about auto-scaling, see the *Data Engineering Integration User Guide*.

Data Privacy Management

Read this section to learn what's new for Data Privacy Management in version 10.5.1.1.

Data Privacy Management Repository Migration

Effective in version 10.5.1.1, you can use the DPMRepoMigration utility to migrate a Data Privacy Management repository on Oracle to PostgreSQL.

For information about how to run the utility, see the *Informatica Data Privacy Management Administrator Guide*.

Sources

Effective in version 10.5.1.1, you can connect to a Workday source. Use CDATA connectors for Workday to create a JDBC data store for Workday. You can perform domain discovery and Subject Registry scans on data stores for Workday applications.

For information about configuring connectivity to sources, see the *Informatica Data Privacy Management Administrator Guide*.

Enterprise Data Catalog

Read this section to learn what's new for Enterprise Data Catalog in version 10.5.1.1.

New Resources

Effective in version 10.5.1.1, the TIBCO Spotfire resource is added in Enterprise Data Catalog. You can use the resource to extract metadata, relationship, and lineage information from the TIBCO Spotfire data source.

Note: Enterprise Data Catalog supports OAuth authentication to connect to the TIBCO Spotfire Server.

For more information, see the *Informatica 10.5.1 Enterprise Data Catalog Scanner Configuration Guide*.

Upper Limit on Random Percentage Sampling Option in Hive Resource

Effective in version 10.5.1.1, you can specify an upper limit when you profile a Hive resource with random percentage as a sampling option using the Spark engine. For optimal performance, the limit is applied on both large and small tables.

What's Changed (10.5.1.1)

Read this section to learn what's changed in version 10.5.1.1.

Enterprise Data Catalog

Read this section to learn what's changed for Enterprise Data Catalog in version 10.5.1.1.

Sampling Options in Google BigQuery Resource

Effective in version 10.5.1.1, you can push the processing for the First N rows and Random Percentage sampling options to Google BigQuery. You can perform sampling on native tables, clustered tables, partitioned tables with require_partition_filter option set to false, and external tables created on top of Google Cloud Storage.

For optimal performance, consider running the profile using the First N rows sampling option.

Sudo Requirements for the Gateway User

Effective in version 10.5.1.1, the commands `tee` and `sed` do not require sudo privileges. Sudo privileges are required for `sshd` to run the cluster validation utility before you upgrade Enterprise Data Catalog.

Cluster Pre-validation

Effective in version 10.5.1.1, the Informatica Cluster Service validates the MaxStartups parameter in the SSH server configuration file. The maximum number of unauthenticated concurrent SSH connections must be set to a value greater than or equal to 30.

CHAPTER 6

10.5.1 Important Notices

This chapter includes the following topics:

- Support Changes, 52
- Release Tasks, 53
- Informatica Installer, 54

Support Changes

Read this section to learn about support changes in version 10.5.1.

Technical Preview

Read this section to learn about technical preview functionality that is initiated or lifted in version 10.5.1.

Technical Preview Lifted

Effective in version 10.5.1, the following functionality is lifted from technical preview:

Data Flow Analytics

You can use Data Flow Analytics to view analytical data about PowerCenter mappings in the catalog. You can use the data to improve the mapping quality and efficiency in your enterprise. You can also view mappings as diagrams for ease of understanding.

The analytical data includes details about the following mapping metrics:

- Similar mappings
- Duplicate mappings
- Reusable mapplets, transformations, and user functions for complex expressions

In addition, the data also includes the mapping group details and the mapping discovery details to provide additional insights into the mapping data. You can download the analytical data as reports for further analysis or enhance the reports using third-party reporting solutions.

For more information, see the *Enterprise Data Catalog 10.5.1 User Guide*.

Data preview on Azure HDInsight and MapR clusters

You can preview data using the Spark engine for mappings configured to run on Azure HDInsight and MapR cluster distributions.

For more information, see the *Data Engineering Integration 10.5.1 User Guide*.

Dropped Support

Effective in version 10.5.1, Data Privacy Management dropped support for Cloudera Sentry and Hortonworks Ranger protection extensions. You cannot create or use the extensions to protect sensitive data.

For information about how to use protection extensions to protect sensitive data, see the Informatica Data Privacy Management User Guide.

For more information about the dropped support, contact Informatica Global Customer Support.

Deprecation

Effective in version 10.5.1, the following functionality is deprecated:

Deprecated functionality is supported, but Informatica intends to drop support in a future release. Informatica requests that you to transition to different functionality before the functionality is dropped.

Enterprise Data Catalog

Effective in 10.5.1, the following resources are deprecated:

- IBM InfoSphere DataStage
- Oracle Data Integrator

Deprecated functionality is supported, but Informatica intends to drop support in a future release. Informatica requests that you to transition to different functionality before the functionality is dropped.

Informatica suggests configuring advanced scanners for these sources in the Advanced Scanners tool before we drop support for the resources. When you create an advanced scanner, you can configure additional functionality such as embedded procedural calls and additional lineage detail.

Release Tasks

Read this section to learn about tasks that you need to perform after you upgrade to version 10.5.1.

PowerExchange for Microsoft Dynamics CRM

Microsoft deprecated the Discovery Service for Microsoft Dynamics CRM Online. Effective in 10.5.1, you must use the Organization Service for the password authentication type to access Microsoft Dynamics CRM Online.

After you upgrade to 10.5.1, mappings fail when you use the Discovery Service URL for the password authentication type to access Microsoft Dynamics CRM Online. You must update the Microsoft Dynamics CRM connection to use the CRM Organization Service URL.

Note: You can continue to use the Discovery Service web service URL for the Claims-based and Active Directory authentication types to connect to the on-premise Microsoft Dynamics CRM.

For more information, see the *Informatica PowerExchange for Microsoft Dynamics CRM 10.5.1 User Guide for PowerCenter*.

Informatica Installer

Read this section to learn what's new for the Informatica installer in version 10.5.1.

Documentation

Some instructions to apply the hotfix for 10.5.1 are moved.

The following table lists the documents where you can find hotfix instructions for each product:

Product	Document
Data Engineering, PowerCenter, Data Quality	Release Notes
Enterprise Data Catalog	Enterprise Data Catalog Upgrading from Versions 10.4.0 and Later
Enterprise Data Preparation	Enterprise Data Preparation Upgrade Guide
Data Privacy Management	Data Privacy Management Upgrading from Versions 10.4.0 and Later
Test Data Management	Test Data Management Upgrade Guide

Upgrade Paths

The upgrade paths that you take depend on the products that you upgrade. If you are on a version that does not support a direct upgrade, you must first upgrade to a supported version.

The following table lists the upgrade paths for each product that the installer supports.

Product	Versions
PowerCenter	You can apply the 10.5.2 hotfix to version 10.5 and 10.5.1 including any service pack or cumulative patch. You can upgrade to version 10.5.2 from the following previous versions: - 10.2 including any hotfix, service pack, or cumulative patch - 10.4 including any service pack or cumulative patch - 10.4.1 including any service pack or cumulative patch
Informatica Data Quality	You can upgrade to version 10.5.2 from the following previous versions: - 10.2 including any hotfix, service pack, or cumulative patch - 10.4 including any service pack or cumulative patch - 10.4.1 including any service pack or cumulative patch
Data Engineering Integration Data Engineering Quality	You can apply the 10.5.2 hotfix to version 10.5 and 10.5.1 including any service pack or cumulative patch. You can upgrade to version 10.5.2 from the following previous versions: - 10.2 including any hotfix, service pack, or cumulative patch - 10.2.1 including any hotfix, service pack, or cumulative patch - 10.2.2 including any hotfix, service pack, or cumulative patch - 10.4 including any service pack or cumulative patch - 10.4.1 including any service pack or cumulative patch **Important:** If Data Engineering is in the same domain as any other product supported by the Informatica installer, verify that all products are upgraded to the same supported upgrade version before you upgrade to 10.5.2.

Product	Versions
Data Engineering Streaming	You can apply the 10.5.2 hotfix to version 10.5 and 10.5.1 including any service pack or cumulative patch. You can upgrade to version 10.5.2 from the following previous versions: - 10.2.2 including any hotfix, service pack, or cumulative patch - 10.4 including any service pack or cumulative patch - 10.4.1 including any service pack or cumulative patch
Data Privacy Management	You can apply the 10.5.2 hotfix to version 10.5 and 10.5.1 including any service pack or cumulative patch. You can upgrade to version 10.5.2 from version 10.4.1.x including any cumulative patch. **Important:** If Data Engineering, Enterprise Data Catalog, and Data Privacy Management are in the same domain of a version earlier than 10.4.1.x, upgrade them all to version 10.4.1.x before you upgrade to 10.5.2.
Enterprise Data Catalog	You can apply the 10.5.2 hotfix to version 10.5 and version 10.5.1 including any service pack or cumulative patch. You can upgrade to version 10.5.2 from the following previous versions: - 10.4 including any service pack or cumulative patch - 10.4.1 including any service pack or cumulative patch **Important:** If Enterprise Data Catalog and Data Engineering are in the same domain of a version earlier than 10.4, upgrade them both to version 10.4 or 10.4.1 before you upgrade to 10.5.2.
Enterprise Data Preparation	You can apply the 10.5.2 hotfix to version 10.5 and 10.5.1 including any service pack or cumulative patch. You can upgrade to version 10.5.2 from the following previous versions: - 10.4 including any service pack or cumulative patch - 10.4.1 including any service pack or cumulative patch **Important:** If Data Engineering, Enterprise Data Catalog, and Enterprise Data Preparation are in the same domain of a version earlier than 10.4, upgrade them all to version 10.4 or 10.4.1 before you upgrade to 10.5.2.
Test Data Management	You can apply the 10.5.2 hotfix to upgrade from version 10.5 or 10.5.1 including any service pack or cumulative patch.

Informatica Deployment Manager

Effective in version 10.5.1, you can use Informatica Deployment Manager to install Enterprise Data Catalog and Informatica Data Quality on Docker.

For more information, see *Install Enterprise Data Catalog on Docker With Informatica Deployment Manager 10.5.1* and *Install Data Quality on Docker With Informatica Deployment Manager 10.5.1* How-to Library (H2L) articles.

Enterprise Data Catalog

Effective in version 10.5.1, when you upgrade Enterprise Data Catalog from a supported version or apply the hotfix on 10.5, the installer validates if you have backed up the catalog. When you roll back the hotfix, the installer validates if a catalog backup exists.

CHAPTER 7

10.5.1 What's New

This chapter includes the following topics:

- Data Engineering Streaming, 56
- Data Privacy Management, 57
- Data Transformation, 58
- Enterprise Data Catalog, 58
- Informatica Data Quality, 60
- Informatica Transformations, 60
- PowerExchange Adapters, 60
- Security (10.5.1), 61
- Test Data Management, 61

Data Engineering Streaming

Read this section to learn what's new for Data Engineering Streaming in version 10.5.1.

Transformations in Databricks

Effective in version 10.5.1, you can add the following transformations to streaming mappings that run in Databricks environment on Azure and AWS platforms:

- Address Validator
- Classifier
- Parser
- Standardizer

For more information, see the *Data Engineering Streaming 10.5.1 User Guide*.

Intelligent Structure Models on Databricks

Effective in version 10.5.1, you can incorporate an intelligent structure model in a Confluent Kafka, Kafka,Azure Event Hubs, and Amazon Kinesis Stream data object for streaming mappings that run on the Databricks Spark engine.

The Databricks Spark engine can process any input type that the model can parse.

For more information about intelligent structure models, see the *Data Engineering Streaming 10.5.1 User Guide*.

Data Privacy Management

Read this section to learn what's new for Data Privacy Management in version 10.5.1.

Security Policy Violation Tags

Effective in version 10.5.1, you can apply tags to security policy violations.

For information about security policy violations, see the *Informatica Data Privacy Management User Guide*.

Security Policy Task Assignment

Effective in version 10.5.1, you can configure task assignment in a Send an Email action in a security policy. You can assign the task to yourself or to another user when you create the security policy. You can view and update task assignment when you edit the security policy.

For information about security policies, see the *Informatica Data Privacy Management User Guide*.

DSAR Report Template

Effective in version 10.5.1, you can customize the template of a DSAR PDF report.

You can customize the logos that appear in the header and footer of the report. You can also change background colors and move or delete field values that appear in the default report.

For information about how to customize a DSAR PDF report template, see the *Informatica Data Privacy Management User Guide*.

Microsoft Azure SQL Server Support

Effective in version 10.5.1, for new installations, you can use a Microsoft Azure SQL Server database to create the Data Privacy Management repository.

For information, see the *Informatica Data Privacy Management Installation and Configuration Guide*.

Sources

Effective in version 10.5.1, you can connect to the following sources:

Google BigQuery

Create a data store to connect to a Google BigQuery database. You can perform domain discovery and Subject Registry scans on Google BigQuery data stores.

SAP

Use CDATA connectors for SAP to create a data store to connect to an SAP application. You can perform Subject Registry scans on SAP data stores that you create with CDATA connectors.

MongoDB

Use CDATA connectors for MongoDB to create a JDBC data store for MongoDB. You can perform domain discovery and Subject Registry scans on MongoDB data stores.

REST Services

Use CDATA connectors for REST Services to create a JDBC data store for REST Services. You can perform domain discovery and Subject Registry scans on REST Services data stores.

For information about configuring connectivity to sources, see the *Informatica Data Privacy Management Administrator Guide*.

Data Transformation

Read this section to learn what's new for Data Transformation in version 10.5.1.

Data Transformation HIX Library

Informatica enhances the performance of message processing within the Data Transformation HIX Library.

For more information about the HIX Library, see the *Data Transformation 10.5 Libraries Guide*.

Enterprise Data Catalog

Read this section to learn what's new for Enterprise Data Catalog in version 10.5.1.

Advanced Scanners

New Advanced Scanners

Effective in version 10.5.1, you can extract metadata using the following advanced scanners:

- Azure Data Factory
- Databricks Notebooks
- SAP Business Object Data Services (BODS)
- Stored procedures support for Amazon Redshift

For more information, see the *Enterprise Data Catalog Advanced Scanners Configuration Guide*.

Global Scopes in SSIS

Effective in version 10.5.1, you can configure global scopes and package level scopes in the Microsoft SQL Server Integration Service (SSIS) scanner default values file.

For more information, see the "HOW TO: Configure global scopes and package level scopes in the default values file of the SQL Server Integration Service scanner" Knowledge Base article.

Sudo Requirements for the Gateway User

Effective in version 10.5.1, after you enable Informatica Cluster Service for the first time, you can choose to disable the sudo permissions validation for the gateway user. After you disable the validation, sudo permissions are not required to restart the Informatica Cluster Service.

SELinux in Enforcing Mode

Effective in version 10.5.1, Enterprise Data Catalog supports SELinux in enforcing mode. You no longer need to set SELinux to disabled or permissive modes.

Support for RHEL 7.9 in FIPS mode

Effective in version 10.5.1, Enterprise Data Catalog supports Red Hat Enterprise Linux 7.9 running in FIPS mode.

Data Asset Analytics

Read this section to learn what's new for Data Asset Analytics in version 10.5.1.

Improved Performance of Key Metrics

Effective in version 10.5.1, the performance of Data Asset Analytics is improved. The key metrics and charts appear faster on the Data Asset Analytics dashboard.

In version 10.5.1, Data Asset Analytics displays key metrics and charts based on analytical data stored in pre-computed tables in the Data Asset Analytics repository

Before you upgrade to version 10.5.1, you must run the Data Asset Analytics utility to populate the pre-computed tables.

Additional Filters for Data Asset Analytics Reports

Effective in version 10.5.1, you can customize the Asset Usage Report, Business Term Association Report, and the User Collaboration Report in Data Asset Analytics using the following additional filters available in the **Filter by** section:

Report Type	Additional Filter
Asset Usage Report	User Login Name
Business Term Association Report	Business Term Glossary Source
User Collaboration Report	User Name Asset Type

Sort Search Results

Effective in version 10.5.1, you can sort the assets based on the name of the listed assets, the relevance, or the system attributes or the custom attributes selected.

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

Informatica Data Quality

Read this section to learn what's new for Informatica Data Quality in version 10.5.1.

Docker Deployment

Effective in version 10.5.1, you can deploy an Informatica domain and the application services for Informatica Data Quality and Data Engineering Quality to a docker container.

Reference Data Database

Effective in version 10.5.1, you can create a reference data database on PostgreSQL.

For more information, see the *Informatica 10.5.1 Installation for PowerCenter and Data Quality Guide*.

Salesforce Connection Support in Informatica Analyst

Effective in version 10.5.1, you can create a connection to a Salesforce system and import Salesforce objects with the Analyst tool.

For more information, see the *Informatica 10.5.1 Analyst Tool Guide*.

Informatica Transformations

Read this section to learn what's new for Informatica transformations in version 10.5.1.

Address Validator Transformation

Effective in version 10.5.1, Informatica is an accredited eircode encoder in Ireland. The Address Validator transformation meets the formal requirements that the eircode accreditation programme defines for the addition of eircodes to addresses.

For comprehensive information about the features and operations of the address verification software engine in version 10.5.1, see the *Informatica Address Verification 5.16.1 Developer Guide*.

PowerExchange Adapters

Read this section to learn what's new for PowerExchange adapters in version 10.5.1.

PowerExchange Adapters for Informatica

Read this section to learn what's new for Informatica PowerExchange adapters in version 10.5.1.

PowerExchange for Amazon S3

Effective in version 10.5.1, PowerExchange for Amazon S3 includes the following functionalities:

- You can configure a proxy server to connect to Amazon S3 to run a mapping on the Spark engine.
- You can run a mapping on the Databricks Spark engine to read data from and write data to Avro, ORC, and Parquet files that are partitioned based on directories.

For more information, see the *Informatica 10.5.1 PowerExchange for Amazon S3 User Guide*.

Security (10.5.1)

Read this section to learn what's new for domain security in version 10.5.1.

Kerberos Constrained Delegation

Effective in version 10.5.1, you can configure an Informatica domain that uses Kerberos authentication to use constrained delegation.

Configure constrained delegation when you want to restrict access to services using delegated authentication. You can use this method to restrict the additional services to which a Kerberos-authenticated user or service can request access.

For more information, see the *Informatica 10.5.1 Security Guide*.

Test Data Management

Read this section to learn what's new for Test Data Management in version 10.5.1.

Hive Data Types

Effective in version 10.5.1, you can use additional Hive data types for Hive and HDFS target connections.

You can use the following data types:

- Char
- Date
- Timestamp
- Varchar

For information about data types, see the *Test Data Management User Guide*.

CHAPTER 8

10.5.1 What's Changed

This chapter includes the following topics:

- Data Privacy Management , 62
- Informatica Transformations, 62
- Intelligent Structure Models, 63
- Enterprise Data Catalog, 64

Data Privacy Management

Read this section to learn what's changed for Data Privacy Management in version 10.5.1.

Sensitive Field Information Export

Effective in version 10.5.1, exported sensitive field information for unstructured sources includes an additional file. The Data.zip file includes a FileSystemClassificationPolicyDetails.csv file. The file lists the impression count for each classification policy that you use in the scan.

Previously, exported data did not include information on the impression count for classification policies.

For information about exporting sensitive field information, see the Informatica Data Privacy Management User Guide

Informatica Transformations

Read this section to learn what's changed for Informatica transformations in version 10.5.1.

Address Validator Transformation

Read this section to learn what's changed for the Address Validator transformation in version 10.5.1.

All Countries

Effective in version 10.5.1, the Address Validator transformation uses version 5.16.1 of the Informatica Address Verification software engine.

Previously, the transformation used version 5.15.0 of the Informatica Address Verification software engine.

Australia

Informatica introduces the following enhancements for Australia:

AMAS Standard

Effective in version 10.5.1, Informatica supports the 2021 Address Matching Approval System (AMAS) standard for address verification software.

Previously, Informatica supported the 2019 AMAS standard for address verification software.

Optimization Level Updates

Effective in version 10.5.1, the Address Validator transformation enables WIDE optimization when you verify Australia data in certified mode.

Previously, the Address Validator transformation disabled WIDE optimization in this scenario. When you selected WIDE optimization in an earlier release and you verified Australia data in certified mode, the Address Validator transformation applied the criteria defined by STANDARD optimization.

Note: The AMAS certification standard does not support addresses that you verify with the WIDE optimization level. You can set the optimization level in certified mode to evaluate the effect of the level on your data. If you plan to submit an address set to the Australian postal service as certified to the AMAS standard, use a STANDARD optimization level.

Canada

Effective in version 10.5.1, Informatica supports the 2020 Software Evaluation and Recognition Program (SERP) standard for address verification software.

Previously, Informatica supported the 2019 SERP standard for address verification software.

New Zealand

Effective in version 10.5.1, Informatica supports the 2020 SendRight standard for address verification software.

Previously, Informatica supported the 2019 SendRight standard for address verification software.

United States

Effective in version 10.5.1, the Address Validator transformation can assign different scores to street alias candidates for a given address.

Previously, the Address Validator transformation assigned a single common score to the street alias candidates.

For comprehensive information about updates to the Informatica Address Verification software engine, see the *Informatica Address Verification 5.16.1 Developer Guide*.

Intelligent Structure Models

Read this section to learn what's changed for intelligent structure models in version 10.5.1.

Transformation of Dates and Times

Effective in version 10.5.1, intelligent structure models might transform dates and times to output formats that are different than the output formats in previous versions.

The differences might be caused by changes in the internal Intelligent Structure Discovery libraries.

For more information about intelligent structure models, see the *Data Engineering Integration 10.5.1 User Guide*.

Enterprise Data Catalog

Read this section to learn what's changed for Enterprise Data Catalog in version 10.5.1.

SSL-enabled Informatica Cluster Service and associated services

Effective in version 10.5.1, the services associated with the Informatica Cluster Service such as MongoDB, Nomad, Apache Solr, ZooKeeper, and PostgreSQL are always SSL enabled.

Previously, the services associated with the Informatica Cluster Service were SSL enabled only if you enabled secure communication for the domain.

Custom SSL Certificate Requirements

Effective in version 10.5.1, the following changes are made to the custom SSL certificate requirements:

- The Organizational Unit (OU) parameter is optional. Previously, the OU parameter was a required parameter.
- You can use different Common Name (CN) parameters for the client and cluster certificates.
- You cannot use the domain certificates with signing capability to generate cluster and client certificates.

Scanner Binaries for Informatica Platform Resources

Effective in version 10.5.1, the ScannerBinaries.zip file only includes the Informatica Platform resource binary files for the current release. The binary files for earlier Informatica versions are included in the ExtendedScannerBinaries.zip file.

Previously, the ScannerBinaries.zip file included the Informatica Platform resource binary files from version 10.2.2 through 10.5.

Data Asset Analytics

Read this section to learn what's changed for Data Asset Analytics in version 10.5.1.

Data Asset Analytics Key Metrics

Effective in version 10.5.1, the process to calculate the key metrics displayed in the **Data Asset Inventory**, **Data Asset Enrichment**, and **User Collaboration** tabs has changed.

Data Asset Analytics displays the key metrics based on the total number of resources, assets, enrichments, and collaborations created and available during the specified time range in the catalog.

Previously, Data Asset Analytics displayed the key metrics based on the total number of resources, assets, enrichments, and collaborations added to the catalog during the specified time range.

Part III: Versions 10.5 - 10.5.0.1

This part contains the following chapters:

- Version 10.5.0.1, 67
- 10.5 Important Notices, 71
- 10.5 What's New, 80
- 10.5 What's Changed, 108

CHAPTER 9

Version 10.5.0.1

This chapter includes the following topics:

- Important Notices, 67
- What's New, 67
- What's Changed, 68

Important Notices

Read this section to learn about important notices in version 10.5.0.1.

Support Changes

Read this section to learn about support changes in version 10.5.0.1.

Deferment

Effective in version 10.5.0.1, Informatica lifted deferment for Cloudera CDH 6.3.

Additional Support

Effective in version 10.5.0.1, Informatica supports dual-stack communications for Internet Protocol version 4 (IPv4) and Internet Protocol version 6 (IPv6) address formats in Informatica Data Quality.

The components that support dual-stack communications include the Informatica domain, the application services that Informatica Data Quality uses, and the Developer tool and Analyst tool.

For more information, see the Statement of Support for IPv6 in Informatica Products.

What's New

Read this section to learn what's new in version 10.5.0.1.

Data Privacy Management

Read this section to learn what's new for Data Privacy Management in version 10.5.0.1.

Data Stores

Effective in version 10.5.0.1, you can create an Active Directory data store. Create the data store under the Application data store category. The data store creates a corresponding JDBC resource in Enterprise Data Catalog.

For information about data stores, see the *Informatica Data Privacy Management User Guide*.

Scans

Effective in version 10.5.0.1, you can include the following data stores and file types in scans:

Google Slides, Google Sheets, and Google Docs

You can include Google Slides, Google Sheets, and Google Docs file types in domain discovery scans and Subject Registry scans of a Google Drive data store.

Scality Ring 8

You can include Scality Ring 8 storage systems in an Amazon S3 data store that you create with Amazon S3 supported storage types. You can perform domain discovery scans and Subject Registry scans of transaction data stores.

Active Directory

You can perform domain discovery scans and Subject Registry scans on an Active Directory data store.

For information about scans, see the *Informatica Data Privacy Management User Guide*.

Advanced Scanners

Effective in version 10.5.0.1, you can extract metadata, including stored procedures, using the Greenplum advanced scanner.

For more information, see the *Enterprise Data Catalog 10.5 Advanced Scanners Administrator Guide*.

What's Changed

Read this section to learn what's changed in version 10.5.0.1.

Enterprise Data Catalog

Read this section to learn what's new for Enterprise Data Catalog in version 10.5.0.1.

Advanced Scanners

Enhanced Snowflake Scanner

Effective in version 10.5.0.1, you can extract metadata using an improved version of the Snowflake resource.

The new version enables extracting lineage from the following Snowflake data sources:

- Stored procedures
- Functions
- Pipes

- Stages

The resource that you configured in Catalog Administrator is deprecated.

Informatica suggests configuring advanced scanners for these sources in the Advanced Scanners tool before we drop support for the resources. When you create an advanced scanner, you can configure additional functionality such as embedded procedural calls and additional lineage detail.

For more information, see the *Enterprise Data Catalog 10.5 Advanced Scanners Administrator Guide*.

Enhanced SAP BW Scanner

Effective in version 10.5.0.1, you can extract metadata using the advanced scanner for SAP BW with InfoSource support that includes lineage, impact, and relationships for InfoSource.

Note: You can configure the enhanced SAP BW scanner directly in the Catalog Administrator tool.

For more information, see "SAP Business Warehouse (SAP BW)" and "SAP BW/4HANA" in the *Enterprise Data Catalog 10.5 Scanner Configuration Guide*.

Enterprise Data Catalog

Enhanced Microsoft Power BI Scanner

Effective in version 10.5.0.1, you can extract metadata using an improved version of the Microsoft Power BI resource. The new version provides better lineage and supports proxy connection.

The following features are available for the Microsoft Power BI advanced scanner:

- Enabled proxy support, which allows you to provide a proxy host and a proxy port. Optionally, you can provide authentication for the proxy in the form of a username and a password.
- Detailed lineage for data sets that includes fields and calculations within the data sets.
- Enabled support for viewing lineage from external sources.

The resource that you configured in Catalog Administrator is deprecated.

Informatica suggests configuring advanced scanners for these sources in the Advanced Scanners tool before we drop support for the resources. When you create an advanced scanner, you can configure additional functionality such as embedded procedural calls and additional lineage detail.

For more information, see the *Enterprise Data Catalog 10.5 Advanced Scanners Administrator Guide*.

Enhanced SQL Server Integration Service

Effective in version 10.5.0.1, you can extract metadata using an improved version of the SQL Server Integration Service resource. The new version provides better lineage.

The resource that you configured in Catalog Administrator is deprecated.

Informatica suggests configuring advanced scanners for these sources in the Advanced Scanners tool before we drop support for the resources. When you create an advanced scanner, you can configure additional functionality such as embedded procedural calls and additional lineage detail.

For more information, see the *Enterprise Data Catalog 10.5 Advanced Scanners Administrator Guide*.

View Analytics From Option in Data Asset Analytics

Effective in version 10.5.0.1, the **View Analytics From** option is removed from the **Application Configuration** page in Data Asset Analytics. When you select **From Beginning** from the **Time Range** menu, Data Asset Analytics displays the statistical reports for the catalog, beginning from 01/01/2010.

STDOUT and STDERR log files

Effective in version 10.5.0.1, when you run a resource scan and click the **Log Location** link for a scan job that is in the running status, the STDOUT and STDERR logs no longer appear on the Nomad Web UI. A page with the links to access the STDOUT and STDERR log files appears in a new tab.

PowerExchange Adapters for Informatica

Effective in version 10.5.0.1, when you refresh the source or target schema at run time in a dynamic mapping for a flat file, the Data Integration Service uses the value that you specify for the text qualifier formatting option on the Schema tab.

Previously, Data Integration Service ignored the value of the text qualifier that you specified and used the default double quotes instead.

To retain the previous behavior, edit the existing mapping, set the text qualifier to double quotes, and run the mapping again.

The change is applicable to the following adapters:

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

CHAPTER 10

10.5 Important Notices

This chapter includes the following topics:

- Support Changes , 71
- New Products, 74
- Informatica Installer, 75
- Release Tasks, 78

Support Changes

Read this section to learn about support changes in version 10.5.

Technical Preview

Read this section to learn about technical preview functionality that is initiated or lifted in version 10.5.

Technical Preview Initiated

Effective in version 10.5, the following functionality is available for technical preview:

Data Flow Analytics

You can use Data Flow Analytics to view analytical data about PowerCenter mappings in the catalog. You can use the data to improve the mapping quality and efficiency in your enterprise. You can also view mappings as diagrams for ease of understanding.

The analytical data includes details about the following mapping metrics:

- Similar mappings
- Duplicate mappings
- Reusable mapplets, transformations, and user functions for complex expressions

In addition, the data also includes the mapping group details and the mapping discovery details to provide additional insights into the mapping data. You can download the analytical data as reports for further analysis or enhance the reports using third-party reporting solutions.

For more information, see the *Enterprise Data Catalog 10.5 User Guide*.

Graphical Informatica Installer

You can use a graphical user interface based installer to install PowerCenter and Data Engineering Integration on Linux platforms. The installer provides an easy and interactive interface to run the Pre-

installation (i10Pi) System Check Tool, install the Informatica domain, and configure the domain and application services.

Informatica Domain on Kubernetes Cluster

You can run Informatica domain on a Microsoft Azure Kubernetes cluster. The auto-scaling of the active-active services such as Data Integration Service and PowerCenter Integration Service are supported.

For more information, see the *Install PowerCenter on Kubernetes with Informatica Deployment Manager* or the *Install Data Engineering Integration on Kubernetes with Informatica Deployment Manager*.

Intelligent structure models based on an XSD file

When you create an intelligent structure model you can select an XSD file to base the model on.

Mapping audits on the Blaze engine

The Blaze engine can run mappings with audit rules and conditions.

For more information, see the *Data Engineering Integration 10.5 User Guide*.

PowerExchange for Microsoft Azure Data Lake Storage Gen2

The following functionality is available for technical preview:

- You can read objects stored in subdirectories in Microsoft Azure Data Lake Storage Gen2 in the native environment or on the Spark engine.
- You can use wildcard characters to specify the source directory name or the source file name when you read data from a flat file or complex file in the native environment or on the Spark engine.

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

Technical Preview Lifted

Effective in version 10.5, the following functionality is lifted from technical preview:

Catalog Backup

For Enterprise Data Catalog, you can back up the catalog without disabling the Catalog Service. You can back up the catalog using the REST APIs, the Enterprise Data Catalog interface, or the Catalog Administrator. For more information, see the *Catalog Backup and Log Collection Utilities* section in the *Informatica 10.5 Catalog Administrator Guide* or the *Informatica 10.5 Enterprise Data Catalog User Guide*.

Cloudera CDP integration

For Data Engineering Integration, you can use Cloudera Data Platform (CDP) as a compute cluster to execute data engineering jobs in the Hadoop environment. You can use Cloudera CDP when you run data engineering jobs on the Spark engine. Cloudera CDP uses a base cluster and workload clusters to execute data engineering jobs. This architecture allows you to deploy workloads and share data among components by utilizing a shared catalog, unified security, consistent governance, and data life cycle management.

You can use Cloudera CDP when you run a mapping in the Hadoop environment with the following connections:

- PowerExchange for Google Cloud Storage
- PowerExchange for Google BigQuery
- PowerExchange for Google Analytics

- PowerExchange for Google Cloud Spanner

For more information, see the *Data Engineering 10.5 Integration Guide*.

Compact view

You can choose to display the compact view of the lineage and Impact view. The compact lineage and impact view displays the lineage and impact diagram summarized at the resource level.

For more information, see the "View Lineage and Impact" chapter in the *Informatica 10.5 Enterprise Data Catalog User Guide*.

Rollover parameters for Amazon S3 and Azure Data Lake Storage Gen2 targets

For Data Engineering Streaming, you can use different rollover parameters for Amazon S3 and Azure Data Lake Storage Gen2 targets to decide the rollover time or size for each target. For more information, see *Data Engineering Streaming 10.5 User Guide*.

SAP S/4HANA resource

For Enterprise Data Catalog, you can extract metadata from the SAP S/4HANA data source. For more information, see the *Enterprise Data Catalog 10.5 Scanner Configuration Guide*.

Service and Resource Log Collection

For Enterprise Data Catalog, you can collect logs for resources. Additionally, you can collect logs for the following services:

- Enterprise Data Catalog services such as Catalog Service, Informatica Cluster Service, EBF history, and PhoneHome
- Apache Solr
- MongoDB
- Nomad orchestration service and scan jobs
- Apache Zookeeper
- Postgres

You can collect the logs using the REST APIs, the Enterprise Data Catalog interface, or the Catalog Administrator. For more information, see the *Catalog Backup and Log Collection Utilities* section in the *Informatica 10.5 Catalog Administrator Guide* or the *Informatica 10.5 Enterprise Data Catalog User Guide*.

Unique Key Inference

You can generate unique key candidates from the columns in the source data object. You can view the unique key inference results to understand data quality issues by identifying the columns that have duplicate values. You can accept or reject the inferred unique key inference results. After you accept or reject an inferred unique key inference, you can reset the unique key inference to restore the inferred status.

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

Deprecation

Effective in version 10.5, the following functionality is deprecated:

Enterprise Data Catalog Installation on Hadoop Clusters

Effective in version 10.5, you cannot install Enterprise Data Catalog on a Hadoop cluster.

For more information, see the *Informatica 10.5 Installation for Enterprise Data Catalog*.

Extract Data Lineage at the Column Level

Effective in version 10.5, the functionality to extract data lineage at the column level for stored procedures in Oracle and SQL Server is deprecated and will be removed in a future release of Enterprise Data Catalog.

Informatica recommends that you use Advanced Scanners to extract data lineage at the column level for stored procedures in Oracle and SQL Server.

For more information, see the *Enterprise Data Catalog Advanced Scanners Administrator Guide*.

Extract Metadata for Data Lineage at the Column Level

Effective in version 10.5, the functionality to extract metadata for data lineage at the column level including transformation logic from the following data sources is deprecated and will be removed in a future release of Enterprise Data Catalog.

- Oracle Data Integrator data source
- IBM InfoSphere DataStage data source

Informatica recommends that you use Advanced Scanners to extract metadata for data lineage at the column level for Oracle Data Integrator data source and IBM InfoSphere DataStage data sources.

For more information, see the *Enterprise Data Catalog Advanced Scanners Administrator Guide*.

compareMapping DIS Command

Effective in version 10.5, the infacmd dis compareMapping is deprecated and will be removed in a future release.

If you are using scripts based on infacmd dis compareMapping, Informatica recommends that you update the scripts with the new infacmd dis compareObject command. Deprecated functionality is supported, but Informatica intends to drop support in a future release. Informatica requests that you to transition to different functionality before the functionality is dropped.

Universal Connectivity Framework

Effective in version 10.5, support for Universal Connectivity Framework is deprecated. For additional information, contact Informatica Global Customer Support.

Deferment

Effective in version 10.5, Cloudera CDH 6.3 is deferred.

Deferred functionality is either not available or not supported in the current release. Informatica intends to reinstate it, but might choose not to in accordance with changing market or technical circumstances.

New Products

Read this section to learn about new products in version 10.5.

PowerExchange for Kudu for Informatica

Effective in version 10.5, you can use PowerExchange for Kudu to write data to Kudu. You can create a Kudu connection to connect to a Kudu database. You can validate and run Kudu mappings on the Spark engine in the Hadoop environment.

For more information, see the *Informatica 10.5 PowerExchange for Kudu User Guide*.

Informatica Installer

Read this section to learn what's new for the Informatica installer in version 10.5.

Upgrade Paths

The upgrade paths that you take depend on the products that you upgrade. If you are on a version that does not support a direct upgrade, you must first upgrade to a supported version.

The following table lists the upgrade paths for each product that the installer supports.

Product	Versions
PowerCenter	You can upgrade to version 10.5 from the following versions including any hotfix and service pack: - 10.1.1 - 10.2 - 10.4 - 10.4.1
Informatica Data Quality	You can upgrade to version 10.5 from the following versions including any hotfix and service pack: - 10.1.1 - 10.2 - 10.4 - 10.4.1
Data Engineering Integration Data Engineering Quality	You can upgrade to version 10.5 from the following versions including any hotfix and service pack: - 10.1.1 - 10.2 - 10.2.1 - 10.2.2 - 10.4 - 10.4.1 **Important:** If Data Engineering is in the same domain as any other product supported by the Informatica installer, verify that all products are upgraded to the same supported upgrade version before you upgrade to 10.5.
Data Engineering Streaming	You can upgrade to version 10.5 from the following versions including any hotfix and service pack: - 10.2.2 - 10.4 - 10.4.1
Data Privacy Management	You can upgrade to version 10.5 from the following versions including any service packs: - 10.4 - 10.4.1 **Important:** If Data Engineering, Enterprise Data Catalog, and Data Privacy Management are in the same domain of a version earlier than 10.4, upgrade them all to version 10.4 before you upgrade to 10.5.

Product	Versions
Enterprise Data Catalog	You can upgrade to version 10.5 from the following versions including any service packs: - 10.4 - 10.4.1 **Important:** If Enterprise Data Catalog and Data Engineering are in the same domain of a version earlier than 10.4, upgrade them both to version 10.4 or 10.4.1 before you upgrade to 10.5.
Enterprise Data Preparation	You can upgrade to version 10.5 from the following versions including any service packs: - 10.4 - 10.4.1 **Important:** If Data Engineering, Enterprise Data Catalog, and Enterprise Data Preparation are in the same domain of a version earlier than 10.4, upgrade them all to version 10.4 or 10.4.1 before you upgrade to 10.5.
Test Data Management	You can upgrade to version 10.5 from the following versions: - 10.2.0 Hotfix 2 - 10.4 and all service packs - 10.4.1 and all service packs

Data Engineering

Effective in version 10.5, the Data Engineering installer options include the following update and additional features:

Metadata Access Service

Effective in version 10.5, you can create the Metadata Access Service when you run the installer.

Google DataProc

You can create Google DataProc for cluster configuration when you run the installer.

For more information, see the *Informatica 10.5 Installation for Data Engineering Guide*.

Data Privacy Management

Effective in version 10.5, the Data Privacy Management installer options include the following update and additional features:

Non-Hadoop installation

Effective in version 10.5, you install Data Privacy Management on a non-Hadoop cluster. The installer includes a bundled MongoDB database that is used for the Subject Registry data.

PostgreSQL database as repository

You can create the Data Privacy Management repository on a PostgreSQL database.

User activity monitoring

You can enable user activity monitoring and specify an event file storage directory during installation.

Elasticsearch with X-Pack features

You can use a version of Elasticsearch with X-Pack features if you want to include TLS and authentication for Elasticsearch. You must have the required Elasticsearch version `.tar` file.

Advanced configuration

You can enable advanced configuration for Data Privacy Management. If you enable advanced configuration, you can specify values for Elasticsearch and Spark properties. If not, the installer uses the default values. Data Privacy Management uses Elasticsearch and Spark for tasks related to user activity monitoring.

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

Domain

Effective in version 10.5, the Informatica domain installer options include the following update and additional features:

SAML

You can enable the request signing, signed response, and encrypted assertion SAML authentication features during domain creation.

Site Key

Informatica increased the site key size from 128 bit to 256 bit.

Enterprise Data Catalog

Effective in version 10.5, installation and upgrade of Enterprise Data Catalog includes the following changes:

Common Installation File

You can install Enterprise Data Catalog and Enterprise Data Catalog Advanced Scanners using a single installer. During Enterprise Data Catalog installation, you can choose if you want to install Enterprise Data Catalog Advanced Scanners.

Common License

Enterprise Data Catalog and Enterprise Data Catalog Advanced share a common license file. Enterprise Data Catalog Advanced Scanners reads license details from the Enterprise Data Catalog service.

For more information, see the *Enterprise Data Catalog Advanced Scanners Administrator Guide*.

High Availability

You can deploy Enterprise Data Catalog on one, three, or six nodes included in a cluster.

For more information, see *Informatica 10.5 Installation for Enterprise Data Catalog*.

Application Service Changes

The configuration parameters of Catalog Service and Informatica Cluster Service are updated. The Catalog Service does not include options to specify the size of catalog data as low, medium, and high.

The updates are due to the changes in the deployment, associated services, and authentication protocol.

For more information, see *Informatica 10.5 Installation for Enterprise Data Catalog*.

Installation and Deployment Changes

The options to deploy Enterprise Data Catalog in an embedded or existing cluster no longer exist.

Enterprise Data Catalog uses the following applications and associated services:
PostgreSQL Database

Database to store information related to similarity profiling in place of HBase

Nomad

Service orchestration application in place of Apache Hadoop YARN.

MongoDB

Database used as the metadata store in place of HBase.

For more information, see *Informatica 10.5 Installation for Enterprise Data Catalog*.

Export Utility

You must use the utility to back up the catalog data in addition to using the LDM BackupContents command to back up the catalog data.

For more information, see *Upgrading from Versions 10.4.0 and Later*.

Command to Restore Data Post Upgrade

You must use the `LDM migrateContents` command to restore data after you upgrade Enterprise Data Catalog to version 10.5.

For more information, see *Upgrading from Versions 10.4.0 and Later* and *Informatica Command Reference*.

Informatica Deployment Manager

Effective in version 10.5, the Container Utility is renamed to Informatica Deployment Manager. Informatica Deployment Manager uses a secure REST-based installer, and you can use it to deploy and manage the complete lifecycle of the Informatica domain and its services. Additionally, you can apply Emergency Bug Fixes (EBFs) to the deployed entities using the installer.

For more information, see *Install PowerCenter (10.4.x - 10.5.0) on Docker With Informatica Deployment Manager* and *Install Data Engineering Integration (10.4.x - 10.5.0) on Kubernetes With Informatica Deployment Manager* How-to Library (H2L) articles.

PowerCenter

Effective in version 10.5, when you join a domain, you can create and configure the PowerCenter Integration Service and the PowerCenter Repository Service.

Release Tasks

Read this section to learn about tasks that you need to perform after you upgrade to version 10.5.

Test Data Management

Effective in version 10.5, Test Data Management uses an updated algorithm to optimize dictionary usage in Substitution masking. When you choose the Optimize Dictionary Usage option in a Substitution masking plan the plan uses the updated algorithm. The algorithm ensures better use of dictionary values in masked output.

To ensure backward compatibility with plans created previously, and to generate consistent masked output after upgrade, you can create a project property to choose to use the previous algorithm. Set the property value to true for projects that require backward compatibility or to false if you want to use the updated algorithm.

For information about how to use the option in a masking rule and how to configure a project property, see the *Test Data Management 10.5 User Guide*.

CHAPTER 11

10.5 What's New

This chapter includes the following topics:

- CI/CD, 80
- Command Line Programs, 83
- Data Engineering Integration, 89
- Data Engineering Quality, 91
- Data Engineering Streaming, 92
- Data Privacy Management, 93
- Data Transformation, 94
- Enterprise Data Catalog, 95
- Enterprise Data Preparation, 98
- Informatica Transformations, 98
- Informatica Workflows, 99
- Intelligent Structure Models, 99
- Metadata Manager, 100
- PowerCenter , 100
- PowerExchange Adapters, 101
- Security, 105
- Test Data Management, 106

CI/CD

Read this section to learn what's new for CI/CD in version 10.5.

DIS compareObject Command

Effective in version 10.5, you can compare objects between Data Integration Service and Model Repository Service. You can compare objects within a domain or across multiple domains.

The following table describes infacmd dis compareMapping options and arguments:

Option	Argument	Description
-DomainName -dn	domain_name	Required. Name of the Informatica domain. You can set the domain name with the -dn option or the environment variable INFA_DEFAULT_DOMAIN. If you set a domain name with both methods, the -dn option takes precedence.
-ServiceName -sn	DIS_service_name	Required. Name of the Data Integration Service.
-UserName -un	DIS_user_name	Required if the domain uses Native or LDAP authentication. User name to connect to the domain. You can set the user name with the -un option or the environment variable INFA_DEFAULT_DOMAIN_USER. If you set a user name with both methods, the -un option takes precedence.
-Password -pd	DIS_password	Required if you specify the user name. Password for the user name. The password is case sensitive. You can set a password with the -pd option or the environment variable INFA_DEFAULT_DOMAIN_PASSWORD. If you set a password with both methods, the password set with the -pd option takes precedence.
-SecurityDomain -sdn	DIS_security_domain	Required if the domain uses LDAP authentication. Optional if the domain uses native authentication. Name of the security domain to which the domain user belongs. You can set a security domain with the -sdn option or the environment variable INFA_DEFAULT_SECURITY_DOMAIN. If you set a security domain name with both methods, the -sdn option takes precedence. The security domain name is case sensitive. If the domain uses native or LDAP authentication, the default is Native.
-sourceDomainName -srcdn	source_domain_name	Required. Name of the domain for the source object.
-sourceRepositoryService -srcrs	source_MRS_name	Optional. Name of the Model Repository Service for the source object.
- sourceDataIntegrationService -srcdis	source_DIS_name	Optional. Name of the Data Integration Service for the source object.

Option	Argument	Description
-sourceRepositoryUserName -srcrsun	source_MRS_user_name	Optional. The user name for the Model Repository Service that is used to access the source object. You can set the user name with the -srcrsun option or the environment variable INFA_SOURCE_REPOSITORY_USER. If you set a user name with both methods, the -srcrsun option takes precedence.
-sourceRepositoryPassword -srcrspd	source_MRS_password	Required if you specify the user name. Password for the user name. The password is case sensitive. You can set a password with the -srcrspd option or the environment variable INFA_SOURCE_REPOSITORY_PASSWORD. If you set a password with both methods, the password set with the -srcrspd option takes precedence.
-sourceRepositorySecurityDomain -srcrssdn	source_MRS_security_domain	Required if the domain uses LDAP authentication. Optional if the domain uses native authentication. Name of the security domain to which the domain user belongs. You can set a security domain with the -srcrssdn option or the environment variable INFA_DEFAULT_SOURCE_SECURITY_DOMAIN. If you set a security domain name with both methods, the -srcrssdn option takes precedence. The security domain name is case sensitive. If the domain uses native or LDAP authentication, the default is Native.
-sourceQuery -srcq	source_query	Required. A string that queries the source object.
-targetDomainName -tgtdn	target_domain_name	Required. Name of the domain for the target object.
-targetRepositoryService -tgtrs	target_MRS_name	Optional. Name of the Model Repository Service for the target object.
-targetDataIntegrationService -tgtdis	target_DIS_name	Optional. Name of the Data Integration Service for the target object.
-targetRepositoryUserName -tgtrsun	target_MRS_user_name	Optional. The user name for the Model Repository Service that is used to access the target object. You can set the user name with the -tgtrsun option or the environment variable INFA_TARGET_REPOSITORY_USER. If you set a user name with both methods, the -tgtrsun option takes precedence.

Option	Argument	Description
-targetRepositoryPassword -tgtrspd	target_MRS_password	Required if you specify the user name. Password for the user name. The password is case sensitive. You can set a password with the -tgtrspd option or the environment variable INFA_TARGET_REPOSITORY_PASSWORD. If you set a password with both methods, the password set with the -tgtrspd option takes precedence.
-targetRepositorySecurityDomain -tgtrssdn	target_MRS_security_domain	Required if the domain uses LDAP authentication. Optional if the domain uses native authentication. Name of the security domain to which the domain user belongs. You can set a security domain with the -tgtrssdn option or the environment variable INFA_DEFAULT_TARGET_SECURITY_DOMAIN. If you set a security domain name with both methods, the -tgtrssdn option takes precedence. The security domain name is case sensitive. If the domain uses native or LDAP authentication, the default is Native.
-targetQuery -tgtq	target_query	Required. A string that queries the target object.
-TimeZone -tz	time_zone	Optional. By default, the command uses the time zone on the machine that runs the Data Integration Service process. For a list of valid time zones, refer to the java.time.ZoneID class.

For more information, see the *Informatica 10.5 Command Reference*.

Query Parameters

Effective in version 10.5, you can use the Type and Object parameters in infacmd command queries to retrieve design-time objects. You can use the parameters with infacmd dis compareObject, queryRunTimeObjects, queryDesignTimeObjects, deployObjectsToFile, replaceAllTag, tag, and untag commands. For more information, see the *Informatica 10.5 Command Reference*.

Command Line Programs

Read this section to learn what's new for command line programs in version 10.5.

infacmd cluster Commands

When you use the infacmd cluster createConfiguration command to create a cluster configuration, you can also create the Hadoop, HBase, HDFS, or Hive connections associated with the cluster.

Use the following new option with infacmd cluster createConfiguration:

New Option	Argument	Description
-createConnections -cc	true|false	Optional. Indicates whether to create connections associated with the cluster configuration. Default is false.

For more information, see the *Informatica 10.5 Command Reference*.

infacmd ics Commands

Effective in version 10.5, you can use new shutdownClustercommand options to shut down the Informatica Cluster Service and the corresponding services, such as Nomad, Solr, MongoDB, and Postgres SQL.

infacmd isp Commands

Effective in version 10.5, you can use new command options in infacmd isp switchToGatewayNode to specify the location of and the password to the SAML keystore.

The following table describes the new infacmd isp switchToGatewayNode command options:

Option	Description
-SamlKeyStoreDir -skd	Optional. The directory containing the custom keystore file required to use SAML authentication on the gateway node. Specify the directory only, not the full path to the file.
-SamlKeyStorePassword -skp	Required if you use a custom keystore for SAML authentication. Password to the SAML keystore.

infacmd ldm Commands

Effective in version 10.5, you can use new ldm migrateContents command options in infacmd ldm to migrate content. Provide the input directory from where you want to migrate or verify the content. Run the migrateContents command when the Catalog Service, Informatica Cluster Service, and the required stores are enabled.

infacmd ms Commands

You can use infacmd ms runMapping to run audits with your mapping. After you configure the audit rules and conditions in the Developer tool, use the following new option:

New Option	Description
-EnableAudit -ea	Optional. Indicates whether the mapping runs the audit rules and conditions configured for the mapping. Default is false.

For more information, see the *Informatica 10.5 Command Reference*.

infasetup Commands

Effective in version 10.5, you can use infasetup commands to define or update the domain or domain nodes to set up or enhance SAML authentication for web applications.

Each of the following sections describes new options for infasetup commands:

defineDomain

The following table describes the new options for the infasetup defineDomain command:

Option	Description
-SamlAssertionSigned -sas	Optional. Set to TRUE to enable assertion signing by the identity provider. Default is FALSE.
-AuthnContextComparsion -acc	Specifies the comparison method used to evaluate the requested authorization statement. Set to MINIMUM, MAXIMUM, BETTER, or EXACT. Default is Exact.
-AuthnContextClassRef -accr	The authentication context class. Set to PASSWORD or PASSWORDPROTECTEDTRANSPORT.
-SignSamlRequest -ssr	Set to true to enable signed request. Default is False
-RequestSigningPrivateKeyAlias -rspa	Required if you enable signed request. Alias name of the private key present in the node SAML keystore using which SAML request should be signed
-RequestSigningPrivateKeyPassword -rspp	Required if you enable signed request. Password to access the private key used for signing the SAML request
-RequestSigningAlgorithm -rsa	Required if you enable signed request. Algorithm used to sign the request. You can set to one of several different values, including RSA_SHA256, DSA_SHA1, and DSA_SHA256.
-SamlResponseSigned -srs	Set to true to enable signed response. Default is False.

Option	Description
-ResponseSigningCertificateAlias -rsca	Required if you enable signed response. Alias name of the certificate present in the gateway node SAML truststore using which SAML response signature will be validated.
-SamlAssertionEncrypted -sae	Required if you enable signed response. Set to true to enable encrypted assertion. Default is False.
-EncyptedAssertionPrivateKeyAlias -espa	Required if you enable encrypted assertion. Alias name of the private key present in the gateway node SAML keystore using which key used for encrypting the assertion will be decrypted.
-EncyptedAssertionPrivateKeyPassword -espp	Required if you enable encrypted assertion. Password to access the private key used for decrypting the assertion encryption key

defineGatewayNode

The following table describes the new options for the infasetup defineGatewayNode command:

Option	Description
-SamlKeyStoreDir -skd	Required if you use a custom keystore for SAML authentication. Path to the SAML keystore.
-SamlKeyStorePassword -skp	Required if you use a custom keystore for SAML authentication. Password to the SAML keystore. **Note:** If you currently run scripts that use defineGatewayNode to enable a custom keystore for SAML authentication, you must update them to include this option.

defineWorkerNode

The following table describes the new options for the infasetup defineWorkerNode command:

Option	Description
-EnableSaml -saml	Optional. Enables or disables SAML authentication in the Informatica domain. Set to true or false. Default is false.
-SamlTrustStoreDir -std	Optional. The directory containing the custom truststore file required to use SAML authentication on the gateway node. Specify the directory only, not the full path to the file. The default Informatica truststore is used if no truststore is specified.
-SamlTrustStorePassword -stp	Required if you use a custom truststore for SAML authentication. The password for the custom truststore.

Option	Description
-SamlKeyStoreDir -skd	Optional. The directory containing the custom keystore file required to use SAML authentication on the gateway node. Specify the directory only, not the full path to the file.
-SamlKeyStorePassword -skp	Required if you use a custom keystore for SAML authentication. Password to the SAML keystore. **Note:** If you currently run scripts that use defineWorkerNode to enable a custom keystore for SAML authentication, you must update them to include this option.

updateGatewayNode

The following table describes the new options for the infasetup updateGatewayNode command:

Option	Description
-SamlKeyStoreDir -skd	Optional. The directory containing the custom keystore file required to use SAML authentication on the gateway node. Specify the directory only, not the full path to the file.
-SamlKeyStorePassword -skp	Required if you use a custom keystore for SAML authentication. Password to the SAML keystore.

updateWorkerNode

The following table describes the new options for the infasetup updateWorkerNode command:

Option	Description
-CipherWhiteList -cwl	Optional. Comma-separated list of JSSE cipher suites that you want to add to the effective list. **Note:** The list must contain at least one valid JRE or OpenSSL cipher suite.
-CipherBlackList -cbl	Optional. Comma-separated list of JSSE cipher suites that you want to remove from the effective list. **Note:** The effective list must contain at least one valid JRE or OpenSSL cipher suite.
-CipherWhiteListFile -cwlf	Optional. Absolute file name of the plain text file that contains a comma-separated list of cipher suites that you want to add to the effective list. **Note:** The list must contain at least one valid JRE or OpenSSL cipher suite.
-CipherBlackListFile -cblf	Optional. Absolute file name of the plain text file that contains a comma-separated list of cipher suites that you want to remove from the effective list. **Note:** The effective list must contain at least one valid JRE or OpenSSL cipher suite.
-EnableSaml -saml	Optional. Enables or disables SAML authentication in the Informatica domain. Set to true or false. Default is false.

Option	Description
-SamlKeyStoreDir -skd	Optional. The directory containing the custom keystore file required to use SAML authentication on the gateway node. Specify the directory only, not the full path to the file.
-SamlKeyStorePassword -skp	Required if you use a custom keystore for SAML authentication. Password to the SAML keystore. **Note:** If you currently run scripts that use updateWorkerNode to enable a custom keystore for SAML authentication, you must update them to include this option.

updateDomainSamlConfig

The following table describes the new options for the infasetup updateDomainSamlConfig command:

Option	Description
-SamlAssertionSigned -sas	Optional. Set to TRUE to enable assertion signing by the identity provider. Default is FALSE.
-SamlKeyStoreDir -skd	Optional. The directory containing the custom keystore file required to use SAML authentication on the gateway node. Specify the directory only, not the full path to the file.
-SamlKeyStorePassword -skp	Required if you use a custom keystore for SAML authentication. Password to the SAML keystore. **Note:** If you currently run scripts that use updateDomainSamlConfig to enable a custom keystore for SAML authentication, you must update them to include this option.
-SignSamlRequest -ssr	Set to true to enable request signing Default is False.
-RequestSigningPrivateKeyAlias -rspa	Required if you enable signed request. Alias name of the private key that Informatica uses to sign the request. This private key resides in the keystore in the gateway node. The corresponding public key, usually a certificate, should be imported to the identity provider.
-RequestSigningPrivateKeyPassword -rspp	Plaintext password of the private key that Informatica uses to sign the request. Default is the password of private key present in the keystore file `<Informatica home>\services\shared\security\infa_keystore.jks` with the alias "Informatica LLC".
-RequestSigningAlgorithm -rsa	Required if you enable signed request. Algorithm used to sign the request. You can set to one of several different values, including RSA_SHA256, DSA_SHA1, and DSA_SHA256.
-SamlResponseSigned -srs	Set to true to specify whether the IDP signs the SAML response. **Note:** When set to TRUE, requires the IDP administrator to configure the identify provider to sign the response. Default is False.
-ResponseSigningCertificateAlias -rsca	Required if you enable signed response. Alias name of the certificate in the gateway node SAML truststore to use to verify the signature.

Option	Description
-SamlAssertionEncrypted -sae	Set to true to specify that the IDP encrypts the assertion. **Note:** When set to TRUE, requires the IDP administrator to configure the identify provider to encrypt the assertion. Default is False.
-EncyptedAssertionPrivateKeyAlias -espa	Alias name of the private key present in the gateway node SAML keystore. The private key is used for encrypting the assertion. The IDP administrator must import the corresponding public key (usually a certificate).
-EncyptedAssertionPrivateKeyPassword -espp	Plaintext password. Default is the password of private key present in the keystore file `<Informatica home>\services\shared\security\infa_keystore.jks` with the alias "Informatica LLC".

infasetup defineDomain

Effective in version 10.5, you can use a new infasetup option to require web applications to use HTTPS when they communicate with the domain.

The following table describes the new infasetup defineDomain command option:

Option	Description
EnableHsts -hsts	Optional. Set to TRUE to enable HTTP strict transport security. HTTP strict transport security requires webapps to use HTTPS.

Data Engineering Integration

Read this section to learn what's new for Data Engineering Integration in version 10.5.

EXTRACT_STRUCT Function

Effective in 10.5, you can use the EXTRACT_STRUCT function in dynamic expressions to extract all elements from a dynamic struct port in an Expression transformation.

The EXTRACT_STRUCT function flattens dynamic struct ports. The expression for the output ports uses the dot operator to extract elements in the dynamic struct.

For more information, see the *Informatica 10.5 Transformation Language Reference*.

File Manager for Cloud File Preprocessing

Effective in 10.5, you can perform file preprocessing such as list, copy, rename, move, remove, and watch on cloud ecosystems Microsoft Azure and Amazon AWS.

filemanager Commands

The following table describes the available commands for the filemanager utility:

Commands	Description
copy	Use the copy command to copy files on an Amazon AWS cloud ecosystem.
copyfromlocal	Use the copyfromlocal command to copy files from a local system to a cloud ecosystem.
list	Use the list command to list files on a cloud ecosystem.
move	Use the move command to move files on a cloud ecosystem.
remove	Use the remove command to delete files from a cloud ecosystem.
rename	Use the rename command to rename files on a cloud ecosystem.
watch	Use the watch command to watch files that trigger a file processing event, mapping, or workflow on a cloud ecosystem.

For more information, see the *Informatica 10.5 Command Reference*.

Mapping Audits

You can create an audit to validate the consistency and accuracy of data that is processed in a mapping.

An audit is composed of rules and conditions. Use a rule to compute an aggregated value for a single column of data. Use a condition to make comparisons between multiple rules or between a rule and constant values.

You can configure audits for the following mappings that run on the native environment or the Spark engine:

- Read operations in Amazon S3, JDBC V2, Microsoft Azure SQL Data Warehouse, and Snowflake mappings.
- Read operations for complex files such as Avro, Parquet, and JSON in HDFS mappings.
- Read and write operations in Hive and Oracle mappings.

For more information, see the *Data Engineering Integration 10.5 User Guide*.

Profile on the Databricks Cluster

Effective in version 10.5, you can run profiles on the Databricks cluster.

Profiling on the Databricks cluster

You can create and run profiles on the Databricks cluster in the Informatica Developer and Informatica Analyst tools. You can perform data domain discovery and create scorecards on the Databricks Cluster.

For information about the profiles on the Databricks cluster, see *Informatica 10.5 Data Discovery Guide*.

Sensitive Data Recommendations and Insights by CLAIRE

Effective in version 10.5, CLAIRE artificial intelligence detects sensitive data in mapping sources when Enterprise Data Catalog is configured on the domain.

Recommendations list source columns that contain sensitive data based on data quality rules. You can also add custom types to the sensitive data that CLAIRE detects.

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

Warm Pool Support for Ephemeral Clusters on Databricks

Effective in version 10.5, you can configure ephemeral Databricks clusters with warm pools. A warm pool is a pool of VM instances reserved for ephemeral cluster creation.

When you configure the warm pool instances in the Databricks environment, the instances wait on standby in a running state for ephemeral cluster creation. You can choose to have the instances remain on standby when the ephemeral clusters are terminated.

For more information, see the chapter on cluster workflows in the *Data Engineering Integration User Guide*.

Data Engineering Quality

Read this section to learn what's new for Data Engineering Quality in version 10.5.

Databricks Integration

Effective in version 10.5, you can add the following transformations to a mapping that runs in an Azure Databricks or AWS Databricks environment:

- Address Validator
- Case Converter
- Classifier
- Consolidation
- Decision
- Key Generator
- Labeler
- Match
- Merge
- Parser
- Rule Specification
- Standardizer
- Weight Based Analyzer

For more information, see the *Data Engineering Integration 10.5 User Guide*.

Data Engineering Streaming

Read this section to learn what's new for Data Engineering Streaming in version 10.5.

Cassandra

Effective in version 10.5, you can use Cassandra as a target in streaming mappings.

Cassandra is an open source, NoSQL database that is highly scalable and provides high availability. You can use Cassandra to store large amounts of data spread across data centers or when your applications require high write access speed.

For more information, see the *Data Engineering Streaming 10.5 User Guide*.

DataProc

Effective in version 10.5, Data Engineering Streaming supports DataProc for cluster configuration.

Google Cloud Storage Target on Google Dataproc

Effective in version 10.5, you can use a Google Cloud Storage as a target in streaming mappings to run on a Google Dataproc cluster.

Google Dataproc is a lightweight implementation of Hadoop and Apache Spark on the Google cloud platform. When you integrate Informatica Data Engineering Streaming with Dataproc, you configure an on-premises Informatica domain to run jobs on the Dataproc cloud cluster. You must configure the Dataproc cluster before you integrate the cluster with Data Engineering Streaming.

For more information, see the *Data Engineering Streaming 10.5 User Guide*.

Google PubSub

Effective in version 10.5, you can use Google PubSub as a source in streaming mappings.

Use a Google PubSub source to read messages from the configured Google Cloud PubSub subscription.

Google PubSub is an asynchronous messaging service that decouples services that produce events from services that process events. You can use Google PubSub as a messaging-oriented middleware or for event ingestion and delivery for streaming analytics pipelines. Google PubSub offers durable message storage and real-time message delivery with high availability and consistent performance at scale. You can run Google PubSub servers in all the available Google Cloud regions around the world.

For more information, see the *Data Engineering Streaming 10.5 User Guide*.

High Precision Data Types

Effective in version 10.5, you can enable high-precision mode in streaming mappings. The Spark engine can process decimal values with up to 38 digits of precision.

In high-precision mode, the Spark engine supports decimal data types with precision up to 38 digits and a maximum scale of 38. The scale must be less than the precision.

For more information, see *Data Engineering Streaming 10.5 User Guide*.

Kudu

Effective in version 10.5, you can use Kudu as a target in streaming mappings.

Kudu is a columnar storage manager developed for the Apache Hadoop platform. You can use Kudu to store data in tables. Kudu has a simple data model, where the Kudu table has a primary key made up of one or more columns, each with a defined type. Kudu tables have a columnar structure which helps to vectorize and compress data easily. Use Kudu to perform real-time analytics on fast data. You can use Kudu for fast data search, updates and inserts.

For more information, see the *Data Engineering Streaming 10.5 User Guide*.

Python Transformation in Databricks

Effective in version 10.5, you can add Python transformation to streaming mappings in the Databricks environment on the AWS or Azure platforms.

For more information, see the *Data Engineering Streaming 10.5 User Guide*.

Data Privacy Management

Read this section to learn what's new for Data Privacy Management in version 10.5.

Customize the Brand Logo

Effective in version 10.5, you can customize the logo that appears with the Data Privacy Management product name in multiple locations.

For guidelines and information about how to customize the logo, see the "System Settings" chapter in the *Informatica Data Privacy Management 10.5 User Guide*.

Data Domains

Effective in version 10.5, data domains include an additional rule for conflict handling.

You can configure a data domain to return a match when either the metadata or data condition matches.

You can use the matching rule in data domains that you create in Data Privacy Managementand import from a CSV file.

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

PostgreSQL Support

Effective in version 10.5, you can use a PostgreSQL database to create the Data Privacy Managementrepository.

For information, see the *Informatica Data Privacy Management Installation and Configuration Guide*.

Privacy Dashboard

Effective in version 10.5, the Privacy Dashboard includes the following new features:

Indicator drill-down pages

You can view detailed information on subjects and subject requests from drill-down pages on the following Privacy Dashboard indicators:

- **Top Data Stores by Subject**. View additional information on all Subject Registry data stores and open related pages from select column values.
- **Subject Data by Location**. View detailed information on data stores that contain subjects filtered by location.
- **Subject Requests By Type**. View information on subject requests filtered by the request type, and view corresponding task pages.
- **Subject Requests By Due Date**. View information on subject requests filtered by due date, and view corresponding task pages.

Global and page-specific filters

You can apply global filters on the main page of the Privacy Dashboard. You can apply global and page-specific filters from each of the indicator pages that open.

For information about the Privacy Dashboard and the filters available on the dashboard, see the "Privacy Dashboard" chapter in the *Informatica Data Privacy Management 10.5 User Guide*.

Sources

Effective in version 10.5, you can include the following connection and file types in scans:

Email server

You can create a data store and use the Informatica Discovery Agent to perform domain discovery on an Office 365 Outlook email server.

Optical character recognition for images and scanned PDFs

You can include images and scanned PDFs in domain discovery scans that use the Informatica Discovery Agent. When you enable optical character recognition, you can choose to include an additional Image file type in an unstructured data store. The **Sensitive Files** page includes an additional Image **File Type** filter.

For information about how to create a data store and how to use data sources in an unstructured data store, see the *Informatica Data Privacy Management 10.5 User Guide*.

Data Transformation

Read this section to learn what's new for Data Transformation in version 10.5.

VRL Validations with Data Transformation services

You can use a speed-enhanced Data Transformation engine for VRL validations with Data Transformation services. This enhancement does not affect the existing Data Transformation services that use VRL validations.

For more information, see the *Data Transformation 10.5 User Guide*.

Nillable Attribute in XSD

You can use the `xsd:nillable` attribute to mark XSD elements as nillable. When you mark an element as nillable, the corresponding element in the XML file permits null values.

For more information, see the *Data Transformation 10.5 User Guide*.

Enterprise Data Catalog

Read this section to learn what's new for Enterprise Data Catalog in version 10.5.

Advanced Scanners

Read this section to learn what's new for Advanced Scanners in version 10.5:

Support for new Enterprise Data Catalog Advanced Scanners

You can extract metadata using the following Advanced Scanners:

- Oracle Data Integrator
- Talend Data Integrator
- Snowflake Core Metadata

Streamlined Connections Management

You can extract lineage from ETL Sources and reporting tools without the need to define live connections to the linked systems such as Databases and file systems. In certain cases, those connections are required if metadata of the source system does not fully describe the linked systems.

Note: If the connection to the source or target system is not provided, stored procedure calls lineage is not processed.

Improved Database Editing

You can edit a database in the Connection Registry by selecting from a list of available options, filling in the required fields. You can also use the suggested database URL.

Reference Resources and Reference Assets

You can configure all the Advanced Scanners to extract metadata from data sources or other resources in the catalog that are referenced by the scanners.

For more information, see the *Enterprise Data Catalog Advanced Scanners Administrator Guide*.

Cloning a Resource

Effective in version 10.5, you can clone a resource using the Catalog Administrator. Cloning a resource allows you to create a copy of an existing resource and then modify the configuration as required.

For guidelines and information about how to clone a resource, see the "Cloning a Resource" chapter in the *Informatica 10.5 Catalog Administrator Guide*.

Enhanced Home Page

Effective in version 10.5, the Enterprise Data Catalog home page has been enhanced with an updated background image. Additionally, you can get search results that are filtered based on predefined criteria by applying a search prefilter from the search box before searching for assets.

For more information, see the "Searching for Assets" chapter in the *Informatica 10.5 Enterprise Data Catalog User Guide*.

Enhanced Filter Panel

Effective in version 10.5, in the **Filter** section in the search results page, you can view a collective list of all selections that you have made in each search filter to refine the search results. As required, you can remove selections that you have made from the list.

For more information, see the "Customize Search" chapter in the *Informatica 10.5 Enterprise Data Catalog User Guide*.

Enhanced Search Results

Effective in version 10.5, in the search results page, you can view additional details of an asset in the Additional Information section. The Additional Information section contains the following details:

- The date and time of the previous resource scan
- The name of users associated with the asset. The People subsection may include the names of the data owner, data steward, and subject matter expert.
- Information about whether lineage and impact exists depending on whether data from another asset flows from or to the seed asset.
- The description of the selected asset
- Custom attributes
- Associated Business terms
- Data Domains
- Classifications
- Number of similar columns
- Profiling Information

The assets that appear in the search results page contain the following information:

- Business title
- Technical name
- Asset path
- Certification status
- Asset type
- Ratings

New Walkthroughs

Effective in version 10.5, Enterprise Data Catalog includes new walkthroughs for Enterprise Data Catalog. Walkthroughs are also introduced for the Catalog Administrator.

The following new walkthroughs are available in Catalog Administrator and Enterprise Data Catalog:

Catalog Administrator

- Catalog Administrator Home Page Overview
- Configure Security and Permissions
- Create Resources
- Create Custom Attributes
- Create Data Domains

Enterprise Data Catalog

- Application Configuration
- Business Term Overview
- Data Domain Overview
- Curate Data Domains
- Resource Overview

Note: Before you access the walkthroughs, verify that you have access to the following Walkme domains:

- https://cdn.walkme.com
- https://playerserver.walkme.com
- https://ec.walkme.com
- https://rapi.walkme.com
- https://papi.walkme.com

Notifications Page Enhancements

Effective in version 10.5, you can filter the list of notifications based on the asset type, scan time, and changes followed from a separate panel. You can also sort the list based on the asset name and scan time, and download the list of notifications in a .csv file. If there are multiple changes for an asset, you can view the change history for the asset in the **Asset History** window.

For more information, see the "User Collaboration on Assets" chapter in the *Informatica 10.5 Enterprise Data Catalog User Guide*.

Profile on the Databricks Cluster

Effective in version 10.5, you can perform data discovery on the Databricks cluster.

For more information, see the "Enterprise Data Catalog Concepts" chapter in the *Informatica 10.5. Enterprise Catalog Administrator Guide*.

Resource Enhancements

Effective in version 10.5, Enterprise Data Catalog includes the following resource enhancements:

SAP BW and SAP BW/4HANA Resources

You can configure the SAP BW and SAP BW/4HANA resources to connect to SAP server using Secure Network Communications (SNC) connection.

SAP S/4HANA Resource

The SAP S/4HANA resource includes the following enhancements:

Extract profiling metadata

You can use the SAP S/4HANA resource to extract the following profiling information for a Table asset:

- Domain Discovery
- Column Profiling
- Relationship Discovery

Filter asset types loaded into the catalog

You can configure the SAP S/4HANA resource to extract the asset types that you want to load into the catalog.

SAP server connection type

You can configure the SAP S/4HANA resource to connect to SAP server using a Secure Network Communications (SNC) connection.

Salesforce Resource

You can configure a Salesforce resource to use OAuth authentication to connect to a Salesforce data source.

Axon Resource

You can configure an Axon resource to extract and load specific Axon objects into the Catalog.

Enterprise Data Preparation

Read this section to learn what's new for Enterprise Data Preparation in version 10.5.

Support for New Data Source

Effective in 10.5, you can use the following data sources:

- Cloud data warehouse: Amazon Redshift and Microsoft Azure Synapse
- Cloud data lake: Microsoft Azure Data Lake Storage and Amazon S3
- Cloud database: Amazon RDS and Microsoft Azure SQL
- Data warehouse: Teradata
- Database: Oracle and Microsoft SQL Server

For information about supported data sources, see the *Enterprise Data Preparation 10.5 Administration Guide*.

Import Filtered Data for Preparation

Effective in 10.5, you can import filtered data set for data preparation. For more information, see the *Enterprise Data Preparation 10.5 User Guide*.

Multibyte and Special Characters in Metadata

Effective in 10.5, you can use multibyte or special characters in schema, table, or column names. For more information, see the *Enterprise Data Preparation 10.5 User Guide*.

Informatica Transformations

Read this section to learn what's new with Developer tool transformations in version 10.5.

Data Masking Transformation

Effective in version 10.5, you can perform the following tasks in a Data Masking transformation:

Encryption

You can use the encryption technique to encrypt or decrypt source data. You can configure the technique to preserve format and metadata, preserve metadata, or change the metadata.

Sort order

You can sort the order of entries in a relational dictionary that you use for Substitution masking. Sort the order of entries in the dictionary based on the serial number column to ensure that the order remains consistent and generates consistent masked data.

Optimize dictionary usage

Optimize dictionary usage to increase the number of masked values used from the dictionary in a Substitution masking technique that you configure for repeatable output.

For information, see the "Data Masking Transformation" chapter in the *Informatica 10.5 Developer Transformation Guide*.

Lookup Transformation

Effective in version 10.5, the Spark engine and the Databricks Spark engine can return multiple matches for a Lookup transformation.

Sequence Generator Transformation on Databricks

Effective in version 10.5, the Databricks Spark engine can process the Sequence Generator transformation on Amazon Web Services (AWS) or Microsoft Azure Databricks clusters.

The Sequence Generator transformation is supported on Databricks when the Databricks cluster and the Informatica domain reside in the same virtual network.

Informatica Workflows

Read this section to learn what's new with workflows in version 10.5.

Workflow database

You can create the workflow database on PostgreSQL.

For more information, see the *Informatica 10.5 Application Service Guide*.

Intelligent Structure Models

Read this section to learn what's new for intelligent structure models in version 10.5.

Dynamic Mappings

You can use intelligent structure models in dynamic mappings.

Directory-Level Partitioning in Complex Files

When you create a complex file data object that uses an intelligent structure model, and the input file is Avro, Parquet, or ORC, you can select the directory where the input files reside.

Intelligent Structure Models on Databricks

You can incorporate an intelligent structure model in an Amazon S3 or Microsoft Azure Data Lake Store data object for mappings that run on the Databricks Spark engine. The Databricks Spark engine can process any input type that the model can parse.

For more information about intelligent structure models, see the *Data Engineering Integration 10.5 User Guide*.

Metadata Manager

Read this section to learn what's new for Metadata Manager in version 10.5.

Support for Azure Microsoft SQL Server

Effective in version 10.5, Metadata Manager supports Azure Microsoft SQL Server as a repository.

PowerCenter

Read this section to learn what's new for PowerCenter in version 10.5.

Ciphers Supported for PowerCenter Services

Effective in version 10.5, PowerCenter services supports all the default ciphers supported by the Informatica domain.

Previously, PowerCenter services supported only the following ciphers:

```
TLS_RSA_WITH_3DES_EDE_CBC_SHA,TLS_RSA_WITH_AES_128_CBC_SHA,TLS_RSA_WITH_AES_128_CBC_SHA25
6,TLS_RSA_WITH_AES_256_CBC_SHA,
and TLS_RSA_WITH_AES_256_CBC_SHA256
```

Connectivity

Effective in version 10.5, PowerExchange Adapters for PowerCenter includes the following features:

- You can configure an Oracle target session to update or upsert data in batches of arrays.
  For more information, see the *Informatica 10.5 PowerCenter Workflow Basics Guide*.
- You can parameterize the connection string in the Oracle connection and Oracle external loader connection.
  For more information, see the *Informatica 10.5 PowerCenter Workflow Basics Guide*.
- When you import a source definition for the following connections, you can edit the metadata for the fields in the source qualifier:

  - PowerExchange for Google Cloud Storage

- PowerExchange for Kafka

Transformations

Read this section to learn what's new for PowerCenter transformations in version 10.5.

Data Masking Transformation

Effective in version 10.5, you can use the following options in a Data Masking transformation:

Format preserving encryption

You can use the format preserving encryption technique to encrypt or decrypt source data. You can configure the technique to preserve format and metadata, preserve metadata, or change the metadata.

Sort order

You can sort the order of entries in a relational dictionary that you use for Substitution masking. Sort the order of entries in the dictionary based on the serial number column to ensure that the order remains consistent and generates consistent masked data.

Optimize dictionary usage

Optimize dictionary usage to increase the number of masked values used from the dictionary in a Substitution masking technique that you configure for repeatable output.

For information, see the "Data Masking Transformation" chapter in the*Informatica PowerCenter 10.5 Transformation Guide*.

PowerExchange Adapters

Read this section to learn what's new for PowerExchange adapters in version 10.5.

PowerExchange Adapters for Informatica

Read this section to learn what's new for Informatica PowerExchange adapters in version 10.5.

PowerExchange for Amazon Redshift

Effective in version 10.5, PowerExchange for Amazon Redshift includes the following features:

- You can read and write data of TIMESTAMPZ data type when you run a mapping in the native environment.
- You can configure key range partitioning when you read data from an Amazon Redshift source and run the mapping in the native environment.
- You can configure dynamic partitioning when you write data to an Amazon Redshift target and run the mapping in the native environment.

PowerExchange for Amazon S3

Effective in version 10.5, PowerExchange for Amazon S3 includes the following features:

- When you run a mapping on the Spark engine, you can read data from and write data to Avro, ORC, and Parquet files organized based on directories.
- You can configure an Amazon S3-compatible storage to access and manage the data that is stored over an S3 compliant interface. You can use the Scality RING S3-compatible storage. Use the Amazon S3 connection to connect to Scality RING and perform the read and write operations similar to Amazon S3.
- You can configure Federated Single Sign-On (SSO) authentication to securely access the Amazon S3 resources.
- You can configure a cached lookup operation to cache the lookup data when you run a mapping on the Spark engine.
- You can read and write flat files with and without headers.
- You can configure row delimiter and define qualifier scope when you read and write flat files.
- You can perform audits for read operations in Amazon S3 mappings that run in the native environment or on the Spark engine.

For more information, see the *Informatica 10.5 PowerExchange for Amazon S3 User Guide*.

PowerExchange for Google BigQuery

Effective in version 10.5, PowerExchange for Google BigQuery includes the following features:

- You can write empty strings from a source as null values to a Google BigQuery target.
- You can use the Merge query to perform update, upsert, or delete operations in a single statement when you write to a Google BigQuery target.

For more information, see the *Informatica 10.5 PowerExchange for Google BigQuery User Guide*.

PowerExchange for Google Cloud Storage

Effective in version 10.5, PowerExchange for Google Cloud Storage includes the following features:

- You can read the complete path and names of the Google Cloud Storage source files.
- You can read and write flat files with and without headers.
- You can configure row delimiter and define qualifier scope when you read and write flat files.

For more information, see the *Informatica 10.5 PowerExchange for Google Cloud Storage User Guide*.

PowerExchange for HDFS

Effective in version 10.5, PowerExchange for HDFS includes the following features:

- When you run a mapping on the Spark engine, you can read data from and write data to Avro, ORC, and Parquet files that are partitioned based on directories.
- You can perform audits for read operations for complex files such as Avro, Parquet, and JSON in HDFS mappings that run in the native environment or on the Spark engine.

For more information, see the *Informatica 10.5 PowerExchange for HDFS User Guide*.

PowerExchange for Hive

Effective in version 10.5, you can perform audits for read and write operations in Hive mappings that run in the native environment or on the Spark engine.

For more information, see the *Informatica 10.5 PowerExchange for Hive User Guide*.

PowerExchange for JDBC V2

Effective in version 10.5, PowerExchange for JDBC V2 includes the following features:

- You can use JDBC V2 objects as dynamic sources and targets in mappings.
- You can configure a mapping to create a JDBC V2 target at run time.
- You can use a JDBC V2 connection with the SAP HANA Database subtype in mappings to read from or write to HANA tables. You can also read from HANA data modelling views, such as attribute, analytic, and calculation views. You can validate and run JDBC V2 mappings on the Spark or Databricks Spark engine.
- You can perform audits for read operations in JDBC V2 mappings that run in the native environment or on the Spark engine.

For more information, see the *Informatica 10.5 PowerExchange for JDBC V2 User Guide*.

PowerExchange for Microsoft Azure Blob Storage

Effective in version 10.5, PowerExchange for Microsoft Azure Blob Storage includes the following features:

- You can read and write flat files with and without headers.
- You can configure row delimiter and define qualifier scope when you read and write flat files.

For more information, see the *Informatica 10.5 PowerExchange for Microsoft Azure Blob Storage User Guide*.

PowerExchange for Microsoft Azure Data Lake Storage Gen1

Effective in version 10.5, PowerExchange for Microsoft Azure Data Lake Storage Gen1 includes the following features:

- You can read and write flat files with and without headers.
- You can configure row delimiter and define qualifier scope when you read and write flat files.

For more information, see the *Informatica 10.5 PowerExchange for Microsoft Azure Data Lake Storage Gen1 User Guide*.

PowerExchange for Microsoft Azure Data Lake Storage Gen2

Effective in version 10.5, PowerExchange for Microsoft Azure Data Lake Storage Gen2 includes the following features:

- When you run a mapping on the Spark engine, you can read data from and write data to Avro, ORC, and Parquet files that are partitioned based on directories.
- You can read and write flat files with and without headers.
- You can configure row delimiter and define qualifier scope when you read and write flat files.

For more information, see the *Informatica 10.5 PowerExchange for Microsoft Azure Data Lake Storage Gen2 User Guide*.

PowerExchange for Microsoft Azure SQL Data Warehouse

Effective in version 10.5, PowerExchange for Microsoft Azure SQL Data Warehouse includes the following features:

- You can connect to Microsoft Azure Data Lake Storage Gen2 to stage files at runtime.
- You can stage files in Parquet format when you read data from Microsoft Azure SQL Data Warehouse.
- You can override the schema name and table name when you read data from and write data to Microsoft Azure SQL Data Warehouse.
- You can use Gzip compression to compress data in Gzip format when you write data to Microsoft Azure SQL Data Warehouse in the native environment.
- You can read and write data of Datetimeoffset data type when you run a mapping on the Spark Engine.
- You can perform audits for read operations in Microsoft Azure SQL Data Warehouse mappings that run in the native environment or on the Spark engine.
- You can read data from and write data to Microsoft Azure SQL Data Warehouse case-sensitive databases.

For more information, see the *Informatica 10.5 PowerExchange for Microsoft Azure SQL Data Warehouse User Guide*.

PowerExchange for Salesforce Marketing Cloud

Effective in version 10.5, you can configure dynamic mappings to include the frequent changes to sources, targets, and transformation logic at run time based on parameters and rules that you define.

For more information, see the *Informatica 10.5 PowerExchange for Salesforce Marketing Cloud User Guide*.

PowerExchange for Snowflake

Effective in version 10.5, PowerExchange for Snowflake includes the following features:

- You can read from and write to Snowflake data warehouse that is enabled for staging data in Google Cloud Platform.
- You can perform audits for read operations in Snowflake mappings that run in the native environment or on the Spark engine.

For more information, see the *Informatica 10.5 PowerExchange for Snowflake User Guide*.

PowerExchange Adapters for PowerCenter

Read this section to learn what's new for PowerCenter adapters in version 10.5.

PowerExchange for Google BigQuery

Effective in version 10.5, PowerExchange for Google BigQuery includes the following features:

- When you perform pushdown optimization using the Google BigQuery ODBC connection, you can use the 2.2.5.1012 version of the Informatica ODBC Driver for Google BigQuery to connect to Google BigQuery and push CHR(), DATEDIFF(), DECODE(), LPAD(), and RPAD() functions to the Google BigQuery database. Contact Informatica Global Customer Support to obtain the Informatica ODBC Driver for Google BigQuery.
- You can use the Merge query to perform update, upsert, or delete operations in a single statement when you write to a Google BigQuery target.
- When you configure a session to write data to Google BigQuery, you can write empty strings from a source as null values to a Google BigQuery target.

For more information, see the *Informatica 10.5 PowerExchange for Google BigQuery User Guide for PowerCenter*.

PowerExchange for Google Cloud Storage

Effective in version 10.5, PowerExchange for Google Cloud Storage includes the following features:

- You can read the complete path and names of the Google Cloud Storage source files.
- You can read and write flat files with and without headers.
- You can configure row delimiter and define qualifier scope when you read and write flat files.

For more information, see the *PowerExchange for Google Cloud Storage 10.5 User Guide for PowerCenter*.

PowerExchange for Greenplum

Effective in version 10.5, when the ODBC subtype in an ODBC connection is Greenplum, you can push down functions, such as AVG(), COUNT(), DATE_COMPARE(), DATE_DIFF(), GET_DATE_PART(), IN(), ISNULL(), MAX(), MIN(), MOD(), STDDEV(), SUM(), and VARIANCE(), to the Greenplum database for processing using full pushdown optimization.

For more information, see the *Informatica PowerCenter 10.5 Advanced Workflow Guide*.

PowerExchange for Kafka

Effective in version 10.5, you can configure whether you want to read messages from a Kafka broker as a stream or in batches.

For more information, see the *Informatica PowerExchange for Kafka 10.5 User Guide for PowerCenter*.

PowerExchange for Snowflake

Effective in version 10.5, you can read from and write to Snowflake data warehouse that is enabled for staging data in Google Cloud Platform.

For more information, see the *Informatica 10.5 PowerExchange for Snowflake User Guide for PowerCenter*.

Security

Read this section to learn what's new for domain security in version 10.5.

Encryption Strengthening

Effective in version 10.5, you can encrypt the Informatica domain passwords and secure connection parameters by using the AES 256-bit encryption algorithm.

To strengthen security encryption, the Informatica installer now generates a unique site key when you install the Informatica services. You cannot generate the site key again. Previously you could regenerate the encryption site key using a keyword and domain name in the `GenerateEncryptionKey` command.

After you upgrade to Informatica 10.5, migrate the site key to the new unique site key.

For a multinode domain, copy the unique site key to each node or save the unique site key at a shared location that can be accessed by each node.

You can use transformation language encoding functions, `AES_ENCRYPT` and `AES_DECRYPT` to encrypt or decrypt storage tables with 256-bit encoding. Use the new argument, `keySize` to enable AES 256-bit encryption.

For more information, see the *Informatica 10.5 Security Guide*.

SAML Functionality

Effective in version 10.5, you can use the following new SAML functionality:

- When you enable SAML authentication in the domain, you can enable one or more of the following enhanced authentication modes: request signing, encrypted assertion, and signed response. For more information, see the *Informatica 10.5 Security Guide*.
- You can use Oracle Access manager, Okta SSO, and Azure Active Directory identity providers to manage SAML authentication on the domain for web applications. For more information, see the H2L article about each identity provider on the Informatica Documentation Portal.

Test Data Management

Read this section to learn what's new for Test Data Management in version 10.5.

PostgreSQL Support

Effective in version 10.5, you can use a PostgreSQL database to create the Test Data Management repository and the test data warehouse repository.

You can also use the database to create a staging connection and storage connection.

For information, see the *Test Data Management 10.5 Configuration Guide*.

Sort Entries in a Relational Dictionary

Effective in version 10.5, you can choose to sort the order of entries in a relational dictionary used in a masking plan.

You can sort the order of entries in a dictionary based on the serial number column or any column to ensure that the order remains consistent and generates consistent masked data.

For information about how to configure the option, see the *Test Data Management 10.5 User Guide*.

Test Data Generation for XSD Sources

Effective in version 10.5, Test Data Generation includes perform the following tasks with XSD sources:

Use conditional constraints in data generation rules

You can use conditional constraints in advanced generation rules and ad hoc generation rules with XSD sources.

For information about how to create and apply advanced generation rules and conditional constraints, see the "Data Generation" chapter in the *Test Data Management 10.5 User Guide*

Configure default handling of unbound MaxOccur values

You can enable default handling of MaxOccurs values that are set as unbound or infinite in an XSD. Use the `Default Value for Unbounded MaxOccurs in a Source XSD` global-level property to enter a value. The value is added to the `MinOccurs` value of the element to calculate the `MaxOccurs` value for the element.

For information about how to set the property value and override the value for an element, see the *Test Data Management 10.5 Administrator Guide* and the *Test Data Management 10.5 User Guide*.

CHAPTER 12

10.5 What's Changed

This chapter includes the following topics:

- Command Line Programs, 108
- Data Engineering , 109
- Data Privacy Management , 109
- Enterprise Data Catalog, 111
- Informatica Transformations, 113
- Intelligent Structure Models , 113
- PowerCenter, 113
- PowerExchange Adapters for Informatica, 113
- Security , 113

Command Line Programs

Read this section to learn what's changed for command line programs in version 10.5.

infasetup

Effective in version 10.5, Informatica stores the assertion signing certificate alias in the domain. You can set the assertionSigningCertificateAlias option, which specifies an alias for the certificate, in the following commands:

- infasetup defineDomain
- infasetup updateDomainSamlConfig
- infasetup updateSamlConfig

Previously, the alias was stored in node metadata. As a consequence, the assertionSigningCertificateAlias option is removed from the defineGatewayNode, switchToGatewayNode and updateGatewayNode commands. Update any scripts that use the node-related commands to use the domain-related commands.

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

Data Engineering

Read this section to learn what's changed for Data Engineering products in version 10.5.

Documentation

Effective in version 10.5, instructions for integrating Informatica with a Google Dataproc cluster are in the *Data Engineering Integration Guide*.

Previously, Dataproc integration instructions were published in How-To Library (H2L) articles. You can view these articles, which cover Dataproc integration with Data Engineering 10.4, on the How-to Library tab in the Informatica Documentation Portal.

Performance

Effective in version 10.5, the PERCENTILE function on the Spark engine consumes less memory, especially when multiple columns compute the PERCENTILE function on the same data set at different percentiles.

Processing Invalid Values

Effective in version 10.5, the Spark engine writes NULL values to the target when the mapping contains certain invalid values.

The Spark engine writes NULL values in the following situations:

- The *terms* argument in PV, FV, PMT, and RATE finance functions passes a 0 value. The value of *terms* must be an integer greater than 0.
- The *month* argument in the MAKE_DATE_TIME function passes an invalid value. The value of *month* must be from 1 to 12.

Previously, the Spark engine wrote random values.

Note: The Data Integration Service rejects rows with invalid values. If you want rows with invalid values to be rejected and not written to the target, run the mapping in the native environment.

Data Privacy Management

Read this section to learn what's changed for Data Privacy Management in version 10.5.

Compressed Files in Scans

Effective in version 10.5, scan results list compressed files with specific extensions as folders with individual files.

You view the scan results on the **Sensitive Fields** page.

Previously, the scan results either skipped the files or listed these compressed files as single files and did not list the individual nested files.

For information about scans of compressed files, see the "Scans" chapter in the *Informatica Data Privacy Management 10.5 User Guide*.

Data Store Properties

Effective in version 10.5, the data store properties include the following changes:

IBM DB2 for z/OS data store properties

Properties of the IBM DB2 for z/OS data store type include a new **Database Option** field. You can choose how you want to enter the database from this list. You can include all databases, select from a list, or specify with a regex. Fields to select from a list or specify a regex appear based on what you choose.

The previous **Database** property is renamed to **DDF Location**.

For information, see the Data Store Properties chapter in the *Informatica Data Privacy Management User Guide*

Deletion of Subject Requests

Effective in version 10.5, you cannot delete DSAR and DSR tasks even if they are in Closed or Completed state.

Previously, you could delete DSAR and DSR tasks that were in Closed state.

For information about tasks, see the "Tasks" chapter in the *Informatica Data Privacy Management 10.5 User Guide*.

Subject Registry Configuration File

Effective in version 10.5, the properties in the Subject Registry configuration file are updated.

Because the install is a non-Hadoop install, the properties file does not include Hadoop and HBase properties. The file includes properties for Name3 configuration.

For information about how to create and update the Subject Registry configuration file, see the *Informatica Data Privacy Management Administrator Guide*.

Total Subject Count

Effective in version 10.5, Data Privacy Management has changed the method used to calculate the total subject count that appears on dashboard widgets and in Data Breach reports.

Because of the change, you might see a difference in the total subject count information on the Privacy Dashboard widgets and in Data Breach reports after you upgrade.

For information about how to view Subject Registry information, see the *Informatica Data Privacy Management User Guide*.

User Activity Monitoring

Effective in version 10.5, you can specify the directory where you want to store event files.

Event files are exported from a database that you configure for user activity monitoring. Specify the location during installation or update the property in the Data Privacy Management Service properties.

Previously, messages were streamed to a default HDFS directory.

For information about how to specify the event files directory, see the *Informatica Data Privacy Management 10.5 Installation and Configuration Guide*.

Enterprise Data Catalog

Read this section to learn what's changed for Enterprise Data Catalog in version 10.5.

Amazon S3 Resource

Effective in version 10.5, the Amazon S3 resource includes the following enhancements in the **General** tab:

Metadata extraction

You can configure the Amazon S3 resource to extract metadata from an Amazon S3 compatible storage such as Scality RING.

Support for temporary session token

You can configure the Amazon S3 resource to connect and extract metadata from an Amazon S3 data source by using a temporary session token.

Previously, the enhancements were available as JVM options in the Amazon S3 resource.

For more information, see the *Informatica® 10.5 Enterprise Data Catalog Scanner Configuration Guide*.

Business Term Propagation

Effective in version 10.5, Enterprise Data Catalog infers and associates a business term with an asset based on enhanced name match condition. The name match condition is based on the business term association results to provide improved inference and association of business terms.

Custom Resources

Effective in version 10.5, you can run a custom resource that has the **Run Script** option enabled on the Informatica domain.

To run more than one custom resource on the Informatica domain simultaneously, you must configure the `LdmCustomOptions.orchestration.oop.max.concurrent.jobs` custom option in the Catalog Administrator.

Default Lineage and Impact Diagram View

Effective in version 10.5, the compact view of the lineage and impact diagram is the default view.

Enterprise Data Catalog Authentication

Effective in version 10.5, Enterprise Data Catalog uses mTLS in place of Kerberos for improved security.

Informatica Similarity Discovery Resource

Effective in version 10.5, the Informatica Similarity Discovery resource includes the following changes:

Create Informatica Similarity Discovery resource

You can create the Informatica Similarity Discovery resource from the **Resource** tab under the **Informatica Platform** resource group. When you run the resource, the scanner discovers similar columns based on column name, column data pattern, unique values in the resource. Previously, the Informatica Similarity Discovery was available as a system resource.

Persist computed information in PostgreSQL database

Enterprise Data Catalog persists the computed information about similar columns, column patterns, and unique values in PostgreSQL database. Previously, Enterprise Data Catalog persisted the computed information about similar columns in Apache Hbase.

Discover similar columns

Enterprise Data Catalog discovers similar columns based on column name, column data pattern, and unique values in the resource. Previously, an internal system job discovered similar columns based on column name, column data pattern, unique values, and value frequency in the resource.

Reference Resources

Effective in version 10.5, the reference resources include the following changes:

Reference resource objects

Enterprise Data Catalog extracts reference resource objects such as data sources, data sets, and data elements.

Reference resource name

The reference resource name is changed from `<resource name>_<Provider ID>_<Connection name>` to `<resource name>$$<Provider ID>$$<Connection name>`.

Connection name

Enterprise Data Catalog supports special characters in the reference resource connection name.

Search Tabs Replaced with Search Prefilters

Effective in version 10.5, you can apply search prefilters to quickly find assets that match predefined search criteria.

Previously, search tabs were available in the **Search Results** page that enabled you to find assets matching a set search criteria.

For more information, see the "Customize Search" chapter in the *Informatica 10.5 Enterprise Data Catalog User Guide*.

Note: If you upgrade from an earlier version, the upgrade process converts the customized search tabs into search prefilters.

Terminology Changes in Documentation

Effective in version 10.5, the Enterprise Data Catalog documentation uses the following terms to refer to existing and embedded clusters:

- Internal cluster in place of an embedded cluster.
- External cluster in place of an existing cluster.

Advanced Scanners uses System Models

Effective in version 10.5, Advanced Scanners uses system models instead of custom models.

Previously, Advanced Scanners used custom models to load metadata.

Informatica Transformations

Effective in version 10.5, the Spark engine increased performance of the Router transformation. When the Spark engine runs a mapping that includes a Router transformation, the Spark engine processes the upstream mapping pipeline once and stages the data so that it can be used by each downstream branch.

Previously, the Spark engine processed the upstream mapping pipeline separately for each output group.

For more information, see the "Router Transformation" chapter in the *Informatica 10.5 Developer Transformation Guide*.

Intelligent Structure Models

Effective in version 10.5, the output of nodes with a date data type is a datetime data type, in the datetime format that is defined in the operating system.

Previously, the output of nodes with a date data type was a string data type.

Note: Upgraded mappings continue to output string data types.

For more information about intelligent structure models, see the *Data Engineering Integration 10.5 User Guide*.

PowerCenter

Effective in version 10.5, the Server Datafile Directory name in the external loader connection type for Sybase IQ in PowerCenter is renamed to Datafile Directory.

For more information, see the *PowerCenter 10.5 Advanced Workflow Guide*.

PowerExchange Adapters for Informatica

Read this section to learn what's changed for Informatica PowerExchange adapters in version 10.5.

PowerExchange for Microsoft Azure SQL Data Warehouse

Effective in version 10.5, when you create a new target table and if the target table already exists in Microsoft Azure SQL Data Warehouse, the data is appended to the existing table.

Previously, the mapping failed if the target table already existed in Microsoft Azure SQL Data Warehouse.

Security

Read this section to learn what's changed for domain security in version 10.5.

SAML Authentication

Effective in version 10.5, you enable enhanced SAML authentication modes signed request, signed response, and encrypted assertion through domain configuration properties.

Previously, you enabled these features using environment variables. For more information, see the "SAML Authentication" chapter in the *Informatica 10.5 Security Guide*.

Part IV: Versions 10.4.1 - 10.4.1.3

This part contains the following chapters:

- Version 10.4.1.3, 116
- Version 10.4.1.2, 117
- Version 10.4.1.1, 123
- 10.4.1 Important Notices, 129
- 10.4.1 What's New, 134
- 10.4.1 What's Changed, 149

CHAPTER 13

Version 10.4.1.3

This chapter includes the following topics:

- Support Changes, 116
- What's New, 116

Support Changes

Read this section to learn about support changes in version 10.4.1.3.

Technical Preview

Effective in version 10.4.1.3, Informatica lifted technical preview for Amazon EMR 6.x.

What's New

Read this section to learn what's new in version 10.4.1.3.

Cloudera CDP

Effective in version 10.4.1.3 for Data Engineering, you can use the Blaze engine to run jobs on Cloudera CDP Private Cloud and Public Cloud clusters.

When the Blaze engine runs jobs on a CDP Private Cloud cluster, you can use Kerberos authentication. When the Blaze engine runs jobs on a CDP Public Cloud cluster, you can use Apache Knox or Kerberos authentication. You can use Apache Ranger, HDFS transparent encryption, or SSL/TLS authorization for jobs that the Blaze engine runs on a CDP cluster.

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

CHAPTER 14

Version 10.4.1.2

This chapter includes the following topics:

- Important Notices, 117
- What's New, 117
- What's Changed, 120

Important Notices

Technical Preview Initiated

Effective in version 10.4.1.2, Informatica includes the following functionality for technical preview:

Amazon EMR 6.0

You can use Informatica Data Engineering products with Amazon EMR 6.0 clusters.

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

What's New

Data Engineering Integration

This section describes new Data Engineering Integration features in version 10.4.1.2.

Cloudera CDP Public Cloud

Effective in version 10.4.1.2, you can run data engineering jobs on Cloudera CDP Public Cloud clusters on AWS or Azure.

You can use CDP Public Cloud when you run data engineering jobs on the Spark engine. When you run a mapping in the Hadoop environment, you can use the same connections for CDP Public Cloud as you can for CDP Private Cloud. You can configure Kerberos or Apache Knox authentication for CDP Public Cloud clusters.

For information about integration with Cloudera CDP Public Cloud, see the *Informatica 10.4.1.2 Release Notes* and *Data Engineering 10.4.1 Integration Guide*.

Data Privacy Management

This section describes new Data Privacy Management features in version 10.4.1.2.

Incremental Subject Registry Scans on Unstructured Sources

Effective in version 10.4.1.2, you can configure incremental Subject Registry scans for unstructured sources.

An incremental scan updates changed records from the last scan run time instead of completely scanning all records.

For information about how to configure an incremental Subject Registry scan, see the *Informatica Data Privacy Management User Guide* and *Informatica Data Privacy Management Administrator Guide*.

Enterprise Data Catalog

This section describes new Enterprise Data Catalog features in version 10.4.1.2.

Amazon S3 Resource

Effective in version 10.4.1.2, the Amazon S3 resource includes the following enhancements:

Metadata extraction

You can configure the Amazon S3 resource to extract metadata from an Amazon S3 compatible storage such as Scality RING.

Support for temporary session token

You can configure the Amazon S3 resource to connect and extract metadata from an Amazon S3 data source by using a temporary session token.

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

Informatica Axon Resource

Effective in version 10.4.1.2, the Informatica Axon resource includes the following enhancements:

Metadata extraction

The Informatica Axon resource extracts metadata from the following assets:

- Attribute
- Policy
- System
- Dataset

You can also read the details of the native users, LDAP users, and SSO users associated with the assets in the Catalog.

View relationships

You can view the following relationships in the Catalog:

- Parent-child relationship between Axon artifacts and extracted objects
- Semantic relationship between the glossary assets

Support reconciliation of domain users

You can reconcile the domain users by selecting the **Allow Filtering** option for the Email system attribute in the Catalog Administrator.

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

Bulk Curation of Business Terms through Import and Export

Effective in version 10.4.1.2, you can accept or reject multiple inferred business terms simultaneously in the exported .csv file.

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

Bulk Curation of Data Domains

Effective in version 10.4.1.2, you can accept a data domain and reject all the other assigned, inferred, or accepted data domains for a column or field asset.

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

Data Asset Analytics

Effective in version 10.4.1.2, Data Asset Analytics includes the following features:

Asset Usage Report

You can use the **Asset Usage** report to view the list of assets used in Enterprise Data Catalog. You can filter the report based on the asset type, date of event, resource type and resource name. In addition to the asset usage details, the report also lists the Enterprise Data Catalog features that you used when you accessed the asset. Features listed include the **Overview**, **Lineage and Impact**, and **Relationships** tabs apart from the other features associated with the asset.

Top Assets Viewed Chart

The **Top Assets Viewed** chart on the **User Adoption** tab page displays a list of the most viewed assets in the catalog.

Feature Usage Value Chart

The **Feature Usage Value** chart on the **Data Value** tab page displays the value of Enterprise Data Catalog features based on their usage over a period of time.

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

SAP S/4HANA Resource

Effective in version 10.4.1.2, you can use the SAP S/4HANA resource to view the detailed lineage between the function module and data source assets.

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

Security

This section describes new security features in version 10.4.1.2.

HSTS Response Header

Effective in version 10.4.1.2, you can choose to enable HTTP Strict Transport Security HSTS response header in your domain to stop HTTP redirects to HTTPS. This prevents man-in-the-middle (MITM) security threats.

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

What's Changed

Enterprise Data Catalog

This section describes changes to Enterprise Data Catalog in version 10.4.1.2.

Data Asset Analytics

Effective in version 10.4.1.2, Data Asset Analytics includes the following changes:

Resource Value calculation

On the **Data Value** tab page, the **Resource Value** is calculated by adding the following values:

- Multiply the specified resource type value with the total number of assets selected based on the asset keywords specified.
- Multiply the value specified for all other resource types with the total number of assets selected based on the asset keywords specified.

Note: You can specify the asset keywords in the **Asset Keywords for Resource Value Calculation** text box in **Application Configuration**.

After you apply 10.4.1.2, you lose all resource-related configuration and values that you had configured earlier. A rollback does not retain the values that you had configured earlier. The values and configuration for resource types are retained and you might see an increase in the new values based on the updated calculation method.

Special characters in report name

For new reports, you can use the following special characters in report names: - (hyphen), @ (at sign), and _ (underscore). Reports created with the earlier versions using unsupported special characters continue to be available with the existing issues. If you copy an existing report that includes unsupported special characters to create a new report, Data Asset Analytics displays a validation error.

Asset Keywords for Resource Value Calculation

In **Application Configuration**, you can specify the asset keywords that you want to use to select assets from the catalog to calculate resource value.

Currency

In **Application Configuration**, you can select the currency in which you want to view asset values on the **Data Value** tab page.

Auto Enrichment column

The **Auto Enrichment** column in the **Asset Enrichment**, **Data Domain Association**, and **Business Term Association** reports lists the following details based on how an asset is enriched:

- Inferred. The recommended enrichments for the asset based on profile results.
- Inferred-Accepted. The inferred enrichment is accepted by the user.
- Inferred-Auto-Accepted. The inferred enrichment is automatically accepted based on the conformance scores specified.
- Inferred-Rejected. The inferred enrichment is rejected by the user.
- Manual. The enrichment for the asset is done manually.

Note: Previously, the report listed **Y**, **No**, or **I** for the **Auto Enrichment** column in the reports.

Auto Enrichment filter

The **Auto Enrichment** filter for the **Asset Enrichment** report lists the following options that you can use to filter the report:

- Inferred
- Inferred-Accepted
- Inferred-Auto-Accepted
- Inferred-Rejected
- Manual

Note: Previously, the **Auto Enrichment** filter listed **Yes** or **No** as the options.

Refresh

The **Monitoring** page includes a **Refresh** icon that you can click to refresh the page and view the current status of the reports run.

Time Range

The default option in the **Time Range** drop-down list on the **Analytics** home page is changed from **From Beginning** to **1 Year**.

Asset Group Type filter and column

The **Asset Group Type** filter and report column are removed from the following reports:

Report	Removed from Filter, Report Column, or Both
Asset Enrichment	Both
Asset Count	Report Column
Assets with Lineage and Impact	Report Column
Business Term Association	Report Column
Data Domain Association	Both

Note: For reports created using the earlier versions, the **Asset Group Type** report column is retained if you download the reports.

Resource Value Calculator

The **Resource Type** and **Resource** options are removed from the **Allocate a Value to** section in the **Resource Value Calculator**.

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

Partial Catalog Backup

Effective in version 10.4.1.2, you can back up the catalog partially based on your requirement to troubleshoot specific issues.

You can use the -st option in the backupContents command to back up one or multiple of the following data stores in the catalog:

- Staging details
- Graph database
- Search details
- Similarity details
- Change notifications details
- Scanner staging details

You can also use the backup REST API to back up one or multiple of the following data stores in the catalog:

- Staging details
- Graph database
- Search details

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

SAP S/4HANA, SAP BW and SAP BW/4HANA Resources

Effective in version 10.4.1.2, you cannot view the following tabs in the Catalog for the SAP BW, SAB BW/4HANA and SAP S/4HANA resources if the asset type does not include lineage, hierarchy, category, and relationship information:

- Lineage and Impact
- Relationship
- Contains

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

Configure Conflict Resolution for Auto Acceptance of Data Domains

Effective in version 10.4.1.2, you can choose the **Match data or column name rule** option to auto-accept data domain on the basis of data or column name rule. If you select the **Match data or column name rule** conflict resolution option, Enterprise Data Catalog auto-accepts data domains when data or column name rule results match.

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

CHAPTER 15

Version 10.4.1.1

This chapter includes the following topics:

- Important Notices, 123
- What's New, 124
- What's Changed, 126

Important Notices

Support Changes

This section describes support changes in version 10.4.1.1.

Technical Preview Lifted

Effective in version 10.4.1.1, the following functionalities are lifted from technical preview:

Cloudera Data Platform Integration

For Data Engineering Integration, you can use Cloudera Data Platform (CDP) as a compute cluster to execute data engineering jobs in the Hadoop environment. You can use Cloudera CDP when you run data engineering jobs on the Spark engine. Cloudera CDP is not supported on the Blaze engine.

Cloudera CDP uses a base cluster and workload clusters to execute data engineering jobs. This architecture allows you to deploy workloads and share data among components by utilizing a shared catalog, unified security, consistent governance, and data life cycle management.

You can use Cloudera CDP when you run a mapping in the Hadoop environment with the following connections:

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

- PowerExchange for Microsoft Azure SQL Data Warehouse
- Sqoop

You can also use Cloudera CDP when you run a PowerExchange for HDFS mapping in the native environment.

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

Distribution Support

In each release, Informatica can add, defer, and drop support for the non-native distributions and 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 Customer Portal:

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

What's New

Data Engineering Integration

Effective in version 10.4.1.1, you can access Amazon Glue as the Hive metastore with an Amazon EMR 5.29 cluster.

Amazon Glue is Amazon's ETL service. The Amazon EMR 5.29 cluster can use Glue as the Hive metastore.

To enable integration with an EMR cluster with Glue, see the *Informatica 10.4.1.1 Release Notes*.

Data Privacy Management

This section describes new Data Privacy Management features in version 10.4.1.1.

Added Support for Apache Cassandra NoSQL Database

Effective in version 10.4.1.1, you can perform domain discovery on Apache Cassandra databases.

You can choose the NoSQL category from the category type to create a data store that connects to an Apache Cassandra database. You can use the data store in scans to perform domain discovery.

For information about how to create a data store or run a scan, see the *Informatica Data Privacy Management User Guide*.

Exact Matching in Subject Registry Search, Index, and Match Configuration

Effective in version 10.4.1.1, you can configure search, index, and match rule types to use Exact matching. When you use Exact matching for indexes, in searches, or in finding matches, the task takes less time to complete.

You can use the `Exact` value for *SearchType*, *IndexType*, and *MatchType* fields in a match configuration.

If you configure Exact matching for indexing, configure the `IgnoreNullEmptyValues` property to determine whether to ignore null or empty values during a Link Subjects scan.

For information about how to configure Exact matching, see the *Informatica Data Privacy Management Administrator Guide*.

Incremental Subject Registry Scans

Effective in version 10.4.1.1, you can configure a Subject Registry scan to be incremental on repeat one-time scans and recurrent scans.

An incremental scan considers changed records from the last scan run time, and updates the results instead of completely scanning all records.

For information about how to configure a Subject Registry scan to be incremental on repeat or recurrent scans, see the *Informatica Data Privacy Management User Guide* and *Informatica Data Privacy Management Administrator Guide*.

New Job Type for Privacy Dashboard Data Update

Effective in version 10.4.1.1, a Subject Registry master scan job triggers a new Privacy Dashboard job when it completes.

The Privacy Dashboard job verifies the changes in the Subject Registry HBase database and updates the Privacy dashboard data that you view when you refresh the dashboard.

The separate job ensures the separation of the scan task and the refresh of the data on the Privacy dashboard.

You cannot manually trigger, pause, or stop a Privacy Dashboard job.

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

Subject Registry Search and Scan Fields

Effective in version 10.4.1.1, you can configure fields for use in Subject Registry scans, searches, or both.

You might want to include fields in a scan but not have the fields appear on the Subject Registry search and details pages.

You can include the `UsedIn` property in a MatchConfiguration for an entity to determine whether the fields appear in Subject Registry searches, scans, or both.

For information about how to configure the property, see the *Informatica Data Privacy Management Administrator Guide*.

Support for the Informatica Discovery Agent on Windows

Effective in version 10.4.1.1, you can use the Informatica Discovery Agent to create an agent on Windows machines.

Use the Informatica Discovery Agent to run domain discovery on unstructured sources and Subject Registry scans.

For information about how to install and use the agent in scans, see the *Informatica Data Privacy Management User Guide*.

Enterprise Data Catalog

This section describes new Enterprise Data Catalog features in version 10.4.1.1.

Choose Factors to Propagate Smart Data Domains

Effective in version 10.4.1.1, you can choose one or more factors to propagate smart data domains to other similar columns. The factors include column name, column data pattern, unique values, and value frequency factors.

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

Collect Data Integration Logs for Profiling

Effective in version 10.4.1.1, you can use the log collection utility to collect Data Integration Service logs for profiling in the Collect Resource Logs page.

For more information, see the "Catalog Backup and Log Collection Utilities" chapter in the *Informatica 10.4.1 Enterprise Data Catalog Administrator Guide*.

SAP S/4HANA Resource

Effective in version 10.4.1.1, the SAP S/4HANA resource extracts metadata from the Append View asset. You can view the lineage and the relationship information for the Append View asset in the Catalog. For more information, see the *Informatica 10.4.1 Enterprise Data Catalog Scanner Configuration Guide*.

Supported Resource Types for Standalone Scanner Utility

Effective in version 10.4.1.1, you can extract metadata from the SAP S/4HANA data source when it is inaccessible at runtime or offline.

For more information, see the "Metadata Extraction from Offline and Inaccessible Resources" chapter in the *Informatica 10.4.1 Enterprise Data Catalog Administrator Guide*.

What's Changed

Data Privacy Management

This section describes changes to Data Privacy Management in version 10.4.1.1.

Interface Changes

Effective in version 10.4.1.1, the Data Privacy Management interface includes the following changes:

Subject Scan Options field on the Scan Creation page

An option to specify subject scan options. Includes the **Full Subject Scan** option that you can select to configure incremental scans on repeat and recurrent scans.

NoSQL category and Apache Cassandra Data Store type

The Category list on the **New Data Store** and **New Scan** pages includes an additional **NoSQL** option. The **Data Store Type** list includes Apache Cassandra in the NoSQL category.

For information about data store and scan creation options, see the *Informatica Data Privacy Management User Guide*.

Subject Registry Entity File Changes

Effective in version 10.4.1.1, the Subject Registry entity file includes additional properties to configure options for an entity.

The Subject Registry entity file has the following additional properties:

UsedIn

Specifies whether to use a match configuration in Subject Registry scans, searches, or both.

Previously, you could not configure separate fields for search and scans.

IsIncrementalScan

Add this flag and set the value to true to configure repeat and recurrent Subject Registry scans to be incremental. Add the flag to a field that contains a time value. The scan uses the value in the field to determine the last updated time for an incremental scan. Applicable for structured sources.

Previously, you could not configure incremental Subject Registry scans.

IgnoreNullEmptyValuesForScan

Determines whether to ignore empty values and continue, or fail a job on encountering null or empty values. Applicable for Subject Registry scans on entities that include *Exact* match type, index type, or search type in a match configuration.

Previously, you could not configure Exact match, search, or index types.

For information about entity files, see the *Informatica Data Privacy Management Administrator Guide*.

Enterprise Data Catalog

This section describes changes to Enterprise Data Catalog in version 10.4.1.1.

Clear Cache Interval

Effective in version 10.4.1.1, the **Clear Expire Interval** label is renamed to **Clear Cache Interval**.

For more information, see the "Configure Analytics Application" chapter in the *Informatica 10.4.1 Enterprise Data Catalog User Guide*.

Pattern Detection Using Column Similarity

Effective in version 10.4.1.1, column similarity does not infer numeric data types for pattern match. It also does not infer X, XX, and XXX patterns for all data types.

Previously, column similarity used the pattern match with other factors to infer numeric data and it also inferred X, XX, and XXX patterns for all data types.

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

Resource Scan History Report

Effective in version 10.4.1.1, the **Scan Type** column displays **Manual**, **Offline**, and **Scheduled** as the resource scan type to indicate the resource scans.

Previously, the **Scan Type** column displayed **Online** to indicate all the scan types. The **Scan Type** column displays **Manual** to represent all the scan types that were marked as **Online** previously.

For more information, see the "Analyze Reports" chapter in the *Informatica 10.4.1 Enterprise Data Catalog User Guide*.

CHAPTER 16

10.4.1 Important Notices

This chapter includes the following topics:

- 10.4.1 Support Changes, 129
- New Products, 132
- Installer Changes, 132
- Release Tasks , 133

10.4.1 Support Changes

This section describes the support changes in version 10.4.1.

Technical Preview Support

Technical Preview Initiated

Effective in version 10.4.1, Informatica includes the following functionalities for technical preview:

Cloudera Data Platform Integration

For Data Engineering Integration, you can use Cloudera Data Platform (CDP) as a compute cluster to execute data engineering jobs in the Hadoop environment. You can use Cloudera CDP when you run data engineering jobs on the Spark engine. Cloudera CDP is not supported on the Blaze engine.

Cloudera CDP uses a base cluster and workload clusters to execute data engineering jobs. This architecture allows you to deploy workloads and share data among components by utilizing a shared catalog, unified security, consistent governance, and data life cycle management.

You can use Cloudera CDP clusters when you run a mapping in the Hadoop environment with the following connections:

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

- PowerExchange for Google Cloud Storage
- PowerExchange for Google BigQuery
- PowerExchange for Google Analytics
- PowerExchange for Google Cloud Spanner
- Sqoop

You can also use Cloudera CDP when you run a mapping that uses a PowerExchange for HDFS connection to read from a source in the native environment or the Hadoop environment.

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

Rollover parameters for Amazon S3 and ADLS Gen2 targets

For Data Engineering Streaming, you can use different rollover parameters for Amazon S3 and ADLS Gen2 targets to decide the rollover time or size for each target.

SAP S/4HANA resource

For Enterprise Data Catalog, you can extract metadata from the SAP S/4HANA data source. For more information, see the *Informatica® 10.4.1 Enterprise Data Catalog Scanner Configuration Guide*.

Service and Resource Log Collection

Effective in version 10.4.1, you can collect logs for resources based on the resource name. You can also collect logs for the services such as HBase, Solr, ingestion, and the Catalog Service. You can collect the logs using the infacmd CLI, REST APIs, or the Catalog Administrator. For more information, see the Enterprise Data Catalog Utilities section in the *Informatica® 10.4.1 Catalog Administrator Guide*.

Uninterrupted Catalog Backup

Effective in version 10.4.1, you can back up the catalog without disabling the Catalog Service. You can back up the catalog using the infacmd CLI, REST APIs, or the Catalog Administrator. For more information, see the Enterprise Data Catalog Utilities section in the *Informatica® 10.4.1 Catalog Administrator Guide*.

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

Technical Preview Lifted

Effective in version 10.4.1, the following functionalities are lifted from technical preview:

Python transformation on Databricks

For Data Engineering Integration, you can include the Python transformation in mappings configured to run on the Databricks Spark engine.

SAP BW, SAP BW/4HANA, and IBM InfoSPhere DataStage resources

For Enterprise Data Catalog, you can extract metadata from the SAP BW, SAP BW/4HANA, and IBM InfoSphere DataStage sources. For more information, see the *Informatica® 10.4.1 Enterprise Data Catalog Scanner Configuration Guide*.

Deferment

This section describes the deferment changes in version 10.4.1.

Deferment Lifted

Effective in version 10.4.1, the following functionalities are no longer deferred:

- Multiple targets with JMS source in a streaming mapping.

IPv6 Support

Effective in version 10.4.1, Informatica supports Internet Protocol version 6 (IPv6) address format in addition to the IPv4 address format.

Ensure that the domain or services have the same address format as the corresponding database instance. Both the node and the database host must be on the same address format. You can also use the Internet Protocol version 6 (IPv6) to connect to IPv6-enabled databases from Informatica Developer or PowerCenter hosted on machines configured with IPv6.

The following table lists the supported connections that you can use from the PowerCenter Client or Developer tool:

Connection	Tool
DB2	PowerCenter Client, Developer tool
JDBC	Developer tool
Microsoft SQL Server	PowerCenter Client, Developer tool
ODBC	PowerCenter Client, Developer tool
Oracle	PowerCenter Client, Developer t ool
Sybase ASE	PowerCenter Client
Sybase IQ	PowerCenter Client

In the database-specific connection properties from the PowerCenter Client or the Developer tool, specify the host name or IP address to connect to the database hosted on the IPv6 network.

For more information about IPv6 support, see https://network.informatica.com/docs/DOC-16182 and https://kb.informatica.com/proddocs/PAM%20and%20EOL/1/Informatica%20Support%20Statement%20for%20IPv6.pdf

Distribution Support

In each release, Informatica can add, defer, and drop support for the non-native distributions and 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 Customer Portal:

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

Enterprise Data Catalog

This section describes Enterprise Data Catalog support changes in version 10.4.1.

Resources

This section describes the resource support changes in version 10.4.1.

Licensing for Resources

Effective in version 10.4.1, you need an advanced license to use the following resources:

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

Contact Informatica Global Customer Support to get an advanced license.

New Products

PowerExchange for MongoDB JDBC

Effective in version 10.4.1, you can read data from and write to MongoDB JDBC by using the PowerCenter client. You can import MongoDB objects as sources and targets to create mappings, and run a session to read data from or write data to MongoDB. When you use MongoDB objects in mappings, you can configure properties specific to MongoDB JDBC.

For more information, see the *Informatica PowerExchange for MongoDB JDBC 10.4.1 User Guide for PowerCenter*

Installer Changes

Informatica version 10.4.1 is a hotfix release. You can apply the hotfix to 10.4.0, or you can upgrade from a supported release. You can run the Informatica installer to install 10.4.1 product installation binaries if these are not present in the machine where you run the installer.

Informatica version 10.4.1 also supports Data Privacy Management and Test Data Management products as part of the unified installer. Test Data Management and Data Privacy Management require an upgrade of services in addition to applying the hotfix. The hotfix instructions are the same, but the pre- and post-requisites are different. Refer to those individual upgrade guides for information.

Data Privacy Management

When you install Data Privacy Management, you can install Data Engineering products and Enterprise Data Catalog and create the services. If you have Data Engineering and Enterprise Data Catalog installed, you can run the installer to install Data Privacy Management on top of those products.

Test Data Management

When you create a domain, the Test Data Management binaries are installed and the directory structure is created with both the TDM and TDW folders. To configure Test Data Management, refer to the *Test Data Management Configuration Guide*.

Release Tasks

This section describes release tasks in version 10.4.1. Release tasks are tasks that you must perform after you upgrade to version 10.4.1.

PowerExchange Adapters for Informatica

This section describes the release tasks for PowerExchange adapters in version 10.4.1.

Post-Upgrade Changes for Informatica PowerExchange for Microsoft Azure Data Lake Storage Gen1

If you upgrade from versions earlier than 10.2.2 to version 10.4.1 and the existing mappings were configured to create a header row in the target, after upgrade when you run the existing mappings, headers are not created in the target.

To generate the headers, you can import the target object again, enable the **Import column names from the first line** option, and run the mapping again.

Post-Upgrade Changes for Informatica PowerExchange for Snowflake

After you apply the 10.4.1 hotfix or upgrade from versions earlier than 10.4.0 to version 10.4.1, you must set the **Additional JDBC URL Parameters** connection property as `jdbc_query_result_format=json` when you use an AIX machine.

PowerExchange Adapters for PowerCenter

This section describes the release tasks for PowerExchange adapters in version 10.4.1.

Post-Upgrade Changes for PowerExchange for Snowflake for PowerCenter

After you apply the 10.4.1 hotfix or upgrade from versions earlier than 10.4.0 to version 10.4.1, PowerExchange for Snowflake includes the following updates:

- You must set the **Additional JDBC URL Parameters** connection property as `jdbc_query_result_format=json` when you use an AIX machine.
- You must specify the Java heap space memory in the PowerCenter Integration Service properties to avoid failures of mappings and mapping tasks that read data from Snowflake. The mappings fail because of the JDBC driver upgrade on the 10.4.1 release.
  Configure the heap size value in the **JVMOption1** property as 256 MB or higher.

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

CHAPTER 17

10.4.1 What's New

This chapter includes the following topics:

- Connectivity, 134
- Data Engineering Integration, 134
- Data Engineering Streaming, 135
- Data Privacy Management, 137
- Enterprise Data Catalog, 138
- Informatica Transformations, 144
- Intelligent Structure Model, 144
- Metadata Manager, 144
- PowerExchange Adapters, 145

Connectivity

This section describes new connectivity features in version 10.4.1.

DataDirect ODBC and JDBC Drivers

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

Oracle Connection Manager

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

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

Data Engineering Integration

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

Hierarchical Data on Hive Sources and Targets

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

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

SAML Authentication for Mass Ingestion

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

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

For more information, see the following guides:

Data Engineering Integration 10.4.1 Mass Ingestion Guide

Informatica® 10.4.1 Application Service Guide

Informatica® 10.4.1 Security Guide

Data Engineering Streaming

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

FileName Port for ADLS Gen2

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

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

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

Ingest CDC Data from Multiple Kafka Topics

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

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

JDBC V2 Lookup Transformation

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

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

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

Parquet Data Format for Complex Targets

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

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

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

Rollover Parameters in Amazon S3 and ADLS Gen2 Targets

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

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

Sources and Targets in Databricks

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

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

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

Streaming Mappings in AWS Databricks

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

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

Sources and Targets

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

- Amazon S3
- Kinesis Streams
- Kinesis Firehose

Transformations

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

Aggregator

Expression

Filter

Joiner

Normalizer

Rank

Router

Union

Window

Data Types

AWS Databricks supports the same data types as Azure Databricks.

The following data types are supported:

Array

Bigint

Date/time

Decimal

Double

Integer

Map

Struct

Text

String

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

Data Privacy Management

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

Data Stores

This section describes new data store properties and sources.

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

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

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

Jobs

This section describes new job types.

Version 10.4.1 includes the following job types:

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

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

Privacy Dashboard

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

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

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

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

Subject Registry Data Encryption

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

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

Unstructured Data Stores

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

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

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

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

Enterprise Data Catalog

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

Business Term Association

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

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

Configure Conflict Resolution for Data Rule and Column Name Rule

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

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

Context Lineage Information

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

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

Data Asset Analytics

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

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

Data Curation through Export and Import

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

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

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

Data Discovery

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

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

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

Enterprise Data Catalog Walkthroughs

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

Note: Before you access the walkthroughs, ensure that you have access to the following Walkme domains:

- https://cdn.walkme.com
- https://playerserver.walkme.com
- https://ec.walkme.com
- https://rapi.walkme.com
- https://papi.walkme.com

The following walkthroughs are available in Enterprise Data Catalog:

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

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

Extracting Transformation Endpoints

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

Field-Level Lineage for Flat Files

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

The following table includes the file systems supported by the resources that you can configure to view the field-level lineage:

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

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

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

File Lineage for Cloud Storage

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

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

Hive Resource

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

Extract connection metadata

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

Automatically assign connections

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

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

Informatica MDM Resource

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

The Informatica MDM resource includes the following enhancements:

Metadata extraction

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

Lineage enhancements

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

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

Microsoft SQL Server Resource

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

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

MicroStrategy Resource

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

View lineage at the report level

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

Support for incremental loading

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

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

Partitioned File Discovery

Effective in version 10.4.1, you can use the Amazon S3 and Azure Data Lake Store Gen2 resources to identify and publish horizontally partitioned files under the same directory, and files organized in Hive based hierarchical directory structures as a single partitioned file.

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

Reference Resources and Reference Assets

Effective in version 10.4.1, you can configure the Hive and Microsoft SQL Server resources to extract metadata from data sources or other resources in the catalog that are referenced by the resource.

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

Resource Level Permissions

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

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

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

Extract the Database Name

Effective in version 10.4.1, you can use the Oracle resource to extract the names of Oracle databases that are configured as metadata sources with the **Import Database Name** option. The option is enabled by default.

You cannot use the `-DextractDatabaseName=true` JVM option to extract the database name of the Oracle metadata source.

Note: If you upgrade from an earlier version, the **Import Database Name** option replaces the `-DextractDatabaseName=true` JVM option in the Oracle resource configuration.

REST APIs

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

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

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

SAP BW and SAP BW/4HANA Resources

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

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

SSIS Resource

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

View detailed lineage for transformations

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

Metadata extraction from SSIS database

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

Filed level and control lineage support

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

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

Informatica Transformations

This section describes new Informatica transformation features in version 10.4.1.

Macro Transformation

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

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

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

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

Intelligent Structure Model

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

Change the Root Node in an Array Structure

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

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

Metadata Manager

This section describes new Metadata Manager features in version 10.4.1.

Configure Java Location and Heap Size for Business Object Resources

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

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

PowerExchange Adapters

This section describes new PowerExchange adapter features in version 10.4.1.

PowerExchange Adapters for Informatica

This section describes new PowerExchange adapter features in version 10.4.1.

PowerExchange for Amazon Redshift

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

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

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

PowerExchange for Amazon S3

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

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

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

PowerExchange for Google BigQuery

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

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

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

PowerExchange for Hive

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

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

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

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

PowerExchange for Microsoft Azure Data Lake Storage Gen2

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

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

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

PowerExchange for Salesforce

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

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

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

PowerExchange for SAP NetWeaver

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

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

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

PowerExchange for Snowflake

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

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

PowerExchange Adapters for PowerCenter

This section describes new PowerExchange adapter features in version 10.4.1.

PowerExchange for Db2 Warehouse

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

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

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

PowerExchange for Greenplum

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

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

PowerExchange for Google BigQuery

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

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

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

PowerExchange for Google Cloud Storage

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

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

PowerExchange for Kafka

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

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

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

PowerExchange for Microsoft Azure SQL Data Warehouse V3

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

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

PowerExchange for Salesforce

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

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

PowerExchange for SAP NetWeaver

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

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

PowerExchange for Snowflake

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

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

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

CHAPTER 18

10.4.1 What's Changed

This chapter includes the following topics:

- Command Line Programs, 149
- Data Engineering Integration, 150
- Data Privacy Management, 150
- Documentation, 151
- Domain Security, 152
- Enterprise Data Catalog, 152
- Informatica Developer , 153
- Metadata Manager, 154
- Test Data Management, 154

Command Line Programs

This section describes changes to commands in 10.4.1

infacmd ldm commands

The following table describes changed infacmd ldm commands:

Command	Change Description
BackupContents	Added the **-mode** to back up the contents. You can back up in offline, hdfs, or online modes.
CreateService	The following options are added: - **-ed**. Enable Data Asset Analytics. - **-ddt**. Select a database for Data Asset Analytics. - **-ddu**. Access the database for Data Asset Analytics. - **-ddp**. Password to access Data Asset Analytics. - **-ddl**. Provide the database connection string - **-dds**. Provide the database schema name. - **-dsjdbcp**. Provide the value to secure the JDBC connection parameters.

For more information, see the *Informatica 10.4 .1 Command Reference*.

Data Engineering Integration

This section describes changes to Data Engineering Integration in version 10.4.1.

Intelligent Structure Discovery

Effective in version 10.4.1, when Intelligent Structure Discovery infers data types for Avro, Parquet and ORC files, it uses the same data types that the native connectors use, resulting in unified output across platforms.

Previously, Intelligent Structure Discovery data types inference wasn't aligned with the inference of the native connectors.

For more information, see "Intelligent structure Models" in the *Data Engineering Integration 10.4.1 User Guide*.

Data Privacy Management

This section describes changes to Data Privacy Management in version 10.4.1.

Data Privacy Management Installer

Effective in version 10.4.1, the Data Privacy Management installer is included in the Informatica services installer. You can choose the required install option for Data Privacy Management from the available product installation options.

Previously, the Data Privacy Management installer was a separate installer. You could choose the required install option from the available Data Privacy Management install options.

The remote agent installer files are now available in the following location after install: `<INFA_HOME>/SecureAtSourceService/InformaticaDiscoveryAgent/InformaticaDiscoveryAgent.zip`

You can use the remote agent for Subject Registry scans and domain discovery scans on unstructured sources.

Previously, the remote agent installer files were available in the following location after install: `<INFA_HOME>/SecureAtSourceService/SubjectRegistryAgent/SubjectRegistryAgent.zip`

You could use the agent only for Subject Registry scans on unstructured sources.

For information about how to use the installer and to install the remote agent, see the *Informatica Data Privacy Management Installation and Configuration Guide*

Overview Workspace

This section describes changes to the **Overview** workspace name.

Effective in version 10.4.1, the **Overview** workspace view that includes indicators for sensitive data, users and user behavior, dashboard analytics, sensitive data locations and proliferation, and protection metrics is called the **Security Dashboard**.

When you first log in to Data Privacy Management, the **Overview** workspace displays the **Privacy Dashboard**. To view the **Security Dashboard**, change the default setting on the **Settings** workspace.

Previously, the **Overview** workspace displayed only the information that now appears on the **Security Dashboard**.

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

Scan Job Reports

This section describes changes to data store scan reports that you download on the **Job Details** page.

Effective in version 10.4.1, you can download a report for the Load Catalog job step. The Profiling job step includes a .zip file that contains two .csv file reports.

Previously, the Load Catalog job step did not include reports, and the Profiling job step contained a single .csv file report.

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

Sensitive Files Page Views

This section describes new views and columns on the **Sensitive Files** page.

Version 10.4.1 includes the following views for unstructured data stores:

Flat View

The Flat View lists the files and folders in an unstructured data store that contain sensitive data domains. The files and folders display in alphabetical order.

The Flat View includes the new Confidence column. The column shows the Confidence Level for the classification policy match to files in unstructured data stores that scan with a remote agent.

Tree View

The Tree View lists the top-level directory that contains the unstructured files and folders that the scan identified. You can drill down to folders and files. To navigate to a previous folder in the directory tree, click a folder name in the file path that appears above the list.

Previously, the Sensitive Fields and Imported Fields views displayed sensitive data for all data store types. In version 10.4.1, only the **Sensitive Fields** page for structured data stores includes these views.

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

Documentation

This section describes changes to guides in Informatica documentation in version 10.4.1.

Data Privacy Management

Effective in version 10.4.1, Data Privacy Management Release Guide and Release Notes are merged with the Informatica Release Guide and Release Notes.

You can find information about new and changed Data Privacy Management features in the Informatica Release Guide. Information on fixed issues and known limitations in Data Privacy Management is available in the Informatica Release Notes.

Previously, Data Privacy Management had a separate Release Guide and Release Notes.

Test Data Management

Effective in version 10.4.1, Test Data Management Release Guide and Release Notes are merged with the Informatica Release Guide and Release Notes.

You can find information about new and changed Test Data Management features in the Informatica Release Guide. Information on fixed issues and known limitations in Test Data Management is available in the Informatica Release Notes.

Previously, Test Data Management had a separate Release Guide and Release Notes.

Domain Security

This section describes changes to Informatica domain security in version 10.4.1.

Enabling Access to a Kerberos-Enabled Domain

Effective in version 10.4.1, you must add DNS entries to the krb5.conf file when the Informatica domain is Kerberos-enabled. Entries for dns_lookup_kdc and dns_lookup_realm enable Kerberos clients to locate KDCs and the Kerberos realm.

Previously, these entries were not required.

For more information, see "Configure the Kerberos Configuration File" in the *Informatica Security Guide*.

Enterprise Data Catalog

This section describes changes to Enterprise Data Catalog in version 10.4.1.

Change Summary Page

Effective in version 10.4.1, you can export, sort, and filter the change summary notifications list. You can filter the list based on asset types and scan time. Previously, you could view the change summary notifications list and filter the list by resource scan time.

For more information, see the "User Collaboration on Assets" chapter in the *Informatica 10.4.1 Enterprise Data Catalog*.

Export Asset Data to a Tableau Data Extract File

Effective in version 10.4.1, this feature includes the following changes:

- You can use the **Export** option to export the asset data to a Tableau Data Extract file if you have the **All Permissions** or **Metadata and Data Read** permission for the resources. Previously, a user did not require any permissions and privileges to use the **Export** option to export the asset data to a Tableau Data Extract file.

- You can configure the `LdmCustomOptions.enableTDEExport` custom property for the Catalog Service and set the value as `false` in Informatica Administrator to hide the **Export** option in the catalog for all users irrespective of the permissions assigned. Previously, the **Export** option was visible to all the users in the catalog.

For more information, see the "Perform Asset Tasks" chapter in the *Informatica 10.4.1 Enterprise Data Catalog User Guide*.

Export Path of the Assets

Effective in version 10.4.1, you can view the path to the assets when you export assets from Business Glossary, Axon Glossary, Data Domain, and Domain User resources. Previously, the exported csv file did not include the path to the assets for these resources.

For more information, see the "Perform Asset Tasks" chapter in the *Informatica 10.4.1 Enterprise Data Catalog User Guide*.

Show Active Business Terms

Effective in version 10.4.1, you can choose to display active business terms from the list of glossaries in the **Add Business Title** dialog box. Previously, the glossaries list displayed inactive business terms along with active business terms.

For more information, see the "Perform Asset Tasks" chapter in the *Informatica 10.4.1 Enterprise Data Catalog User Guide*.

Sort Column and Field Assets

Effective in 10.4.1, you can sort the column and delimited field assets by source position on the **Columns** and **Fields** tabs. Previously, the columns and delimited field assets appeared in ascending or descending order irrespective of their position in the data source.

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

Start Tab Name in Catalog Administrator

Effective in version 10.4.1, the **Start** tab is renamed to the **Overview** tab in Catalog Administrator.

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

Informatica Developer

This section describes changes to Informatica Developer in version 10.4.1.

Relational Object Search

Effective in version 10.4.1, relational searches in the Developer tool are enhanced.

When you create a relational data object, you can search the schemas available in the connection. The Developer tool displays the number of available schemas and the number of tables that match the search.

The list of available resources is organized by resource type. Types that have no members will not appear in the list. For example, if the resource contains no views, the Views node will not appear in the selection list.

For more information, see the section on importing a relational data object in the *Developer Tool Guide*.

Viewing the Execution Plan

Effective in version 10.4.1, you can view the execution plan for mappings that run in the non-native environment from the mapping.

Previously, you viewed the execution plan from the Data Viewer view.

For more information, see "Mapping Execution Plans" in the *Data Engineering Integration 10.4.1 User Guide*.

Metadata Manager

This section describes changes to Metadata Manager in version 10.4.1.

View Data Lineage in HTML 5

Effective in version 10.4.1, you can view the data lineage diagram in HTML 5.

Previously, you could view the data lineage diagram in Adobe Flash player.

For more information, see the "Working with Data Lineage" chapter in the *Informatica 10.4.1 Metadata Manager User Guide*.

Test Data Management

This section describes changes to Test Data Management in version 10.4.1.

Test Data Management Installer

Effective in version 10.4.1, the Test Data Management installer is included in the Informatica services installer. Test Data Management binaries are installed with both traditional and Data Engineering products.

To install Test Data Management, choose to install Informatica services from the available product installation options.

Previously, the Test Data Management installer was a separate installer. You could choose the required installation option from the available Test Data Management install options.

For information about how to install Test Data Management, see *Informatica Installation for PowerCenter and Data Quality* and *Informatica Installation for Data Engineering*.

For information about how to configure Test Data Management, see the *Test Data Management Configuration Guide*.

Part V: Versions 10.4 - 10.4.0.2

This part contains the following chapters:

- Version 10.4.0.2, 156
- Version 10.4.0.1, 157
- 10.4 Important Notices, 161
- 10.4 What's New, 168
- 10.4 What's Changed, 194

CHAPTER 19

Version 10.4.0.2

This chapter includes the following topic:

- What's New, 156

What's New

This section describes what's new in version 10.4.0.2.

Update Cluster Distribution Version

Effective in version 10.4.0.2, you can use the infacmd dis command updateClusterDistributionVersion to update the distribution version for Create Cluster tasks in workflows.

When you apply service pack 10.4.0.2, the cluster version property of Create Cluster tasks does not update to the latest supported version. Use the command to update the cluster distribution version for all workflows in an application.

For more information, see the KB article
"HOW TO: Update cluster distribution version after you apply service pack 10.4.0.2."

CHAPTER 20

Version 10.4.0.1

This chapter includes the following topics:

- Important Notices, 157
- What's New , 158
- What's Changed, 158

Important Notices

Technical Preview Lifted

Effective in version 10.4.0.1, the following functionalities are lifted from technical preview:

Databricks delta table as streaming mapping target

For Data Engineering Streaming, you can use Databricks delta table as a target of streaming mapping for the ingestion of streaming data.

Dynamic streaming mapping

You can configure dynamic streaming mappings to change Kafka sources and targets at run time based on the parameters and rules that you define in a Confluent Schema Registry.

Hive Warehouse Connector and Hive LLAP

For Data Engineering products, use the Hive Warehouse Connector and Hive LLAP with Azure HDInsight 4.x and Hortonworks HDP 3.x clusters to enable Spark code to interact with Hive tables and allow ACID-enabled Hive tables on the Spark engine.

For more information, see the "Mapping Optimizations" chapter in the *Data Engineering Integration 10.4.1 User Guide*.

Snowflake as a streaming mapping target

For Data Engineering Streaming, you can configure Snowflake as a target in a streaming mapping to write data to Snowflake.

What's New

Data Engineering Integration

This section describes new Data Engineering Integration features in version 10.4.0.1.

CLAIRE Recommendations and Insights

Effective in version 10.4.0.1, you can run CLAIRE analysis on individual mappings you multi-select.

You can multi-select more than one mapping, then right-click and choose **Analyze** to read CLAIRE insights about similarities between the selected mappings.

Previously, you could run analysis only on a project or folder.

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

Enterprise Data Catalog

This section describes the new Enterprise Data Catalog features in version 10.4.0.1.

New Resources

Effective in version 10.4.0.1, the following new resources are added in Enterprise Data Catalog:

- Apache Kafka
- Qlik Sense
- MongoDB

You can extract metadata, relationship and lineage information from all the above resources. For more information, see the *Informatica 10.4.0.1 Enterprise Data Catalog Scanner Configuration Guide*.

PowerExchange Adapters for Informatica

This section describes new Informatica adapter features in version 10.4.0.1.

PowerExchange for Microsoft Azure Data Lake Storage Gen2

Effective in version 10.4.0.1, when you create a data object read or write operation, you can read data present in the **FileName** port that contains the endpoint name and source path of the file.

For more information, see the *Informatica 10.4.0 PowerExchange for Microsoft Azure Data Lake Storage Gen2 User Guide*.

What's Changed

Enterprise Data Catalog

This section describes the changes to Enterprise Data Catalog in version 10.4.0.1.

Compact View

Effective in version 10.4.0.1, you can view transformation logic in the Compact view. Previously, you could not view transformation logic in the Compact view.

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

Data Engineering Integration

This section describes the changes to Data Engineering Integration in version 10.4.0.1.

Data Preview on the Spark engine

Effective in version 10.4.0.1, you can use HTTPS when you preview data on the Spark engine using the Spark Jobserver.

If you enable HTTPS protocol on the Data Integration Service, the Spark Jobserver also uses HTTPS protocol. The Spark Jobserver uses the same HTTPS keystore configuration that you set in the Data Integration Service process properties in the Administrator tool.

For more information, see the *Informatica® 10.4.0 Application Service Guide*.

Hive Warehouse Connector and Hive LLAP

Effective in version 10.4.0.1, you can enable the Hive Warehouse Connector and Hive LLAP when you run mappings on the Spark engine. Previously, these were available for technical preview.

Use the Hive Warehouse Connector and Hive LLAP for faster execution of Hive queries when you read and write to Hive tables with the Spark engine. You can use the Hive Warehouse Connector and Hive LLAP with Hortonworks HDP 3.x and Microsoft Azure HDInsight 4.x clusters on the Spark engine.

To enable the Hive Warehouse Connector and Hive LLAP, configure the following properties in the Spark advanced properties for the Hadoop connection:

- `infaspark.useHiveWarehouseAPI=true`
- `spark.datasource.hive.warehouse.load.staging.dir=<HDFS temporary directory for writing to Hive tables>`
- `spark.datasource.hive.warehouse.metastoreUri=<URI for the Hive metastore>`
  Use the value for *hive.metastore.uris* from the hive_site_xml cluster configuration properties.
- `spark.hadoop.hive.llap.daemon.service.hosts=<Application name for the LLAP service>`
  Use the value for *hive.llap.daemon.service.hosts* from the hive_site_xml cluster configuration properties.
- `spark.hadoop.hive.zookeeper.quorum=<Zookeeper hosts used by Hive LLAP>`
  Use the value for *hive.zookeeper.quorum* from the hive_site_xml cluster configuration properties.
- `spark.sql.hive.hiversever2.jdbc.url=<URL for HiveServer2 Interactive.`
  Use the value in Ambari for HiveServer2 JDBC URL.

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

PowerExchange Adapters for Informatica

This section describes changes to Informatica adapters in version 10.4.0.1.

PowerExchange for Microsoft Azure Data Lake Storage Gen2

Effective in version 10.4.0.1, the FileName port appears for the sources and targets used in the existing 10.4.0 dynamic mappings. To run the existing dynamic mappings successfully, add transformations such as Expression to eliminate the FileName port from the source and target objects.

For more information, see the *Informatica 10.4.0.PowerExchange for Microsoft Azure Data Lake Storage Gen2 User Guide*.

CHAPTER 21

10.4 Important Notices

This chapter includes the following topics:

- New Products, 161
- Product Name Changes, 162
- Upgrading to Version 10.4.0, 163
- Support Changes, 163
- Release Tasks, 166

New Products

PowerExchange Adapters for Informatica

This section describes new Informatica adapters in version 10.4.0.

PowerExchange for JDBC V2

Effective in version 10.4.0, you can create a JDBC V2 connection to connect to Aurora PostgreSQL, Azure SQL Database, or any database that supports the Type 4 JDBC driver. You must select the required database subtype in the JDBC V2 connection properties. You can use the JDBC V2 connection in mappings to read from or write data to databases with the Type 4 JDBC driver. You can validate and run JDBC V2 mappings in the native environment or in the non-native environment. In the non-native environment, you can run mappings on the Spark engine in the Hadoop environment, or on Databricks.

For more information, see the *Informatica 10.4.0 PowerExchange for JDBC V2 User Guide*.

PowerExchange for Microsoft Azure Data Lake Storage Gen2

Effective in version 10.4.0, you can create a Microsoft Azure Data Lake Storage Gen2 connection to connect to Microsoft Azure Data Lake Storage Gen2. You can use the Microsoft Azure Data Lake Storage Gen2 connection in mappings to read from or write to Microsoft Azure Data Lake Storage Gen2 objects. You can validate and run Microsoft Azure Data Lake Storage Gen2 mappings in the native environment or in the non-native environment. In the non-native environment, you can run mappings on the Spark engine in the Hadoop environment, or on Databricks.

For more information, see the *Informatica 10.4.0 PowerExchange for Microsoft Azure Data Lake Storage Gen2 User Guide*.

PowerExchange for Salesforce Marketing Cloud

Effective in version 10.4.0, you can create a Salesforce Marketing Cloud connection to connect to Salesforce Marketing Cloud. You can use the Salesforce Marketing Cloud connection in mappings to read from or write contact data to Salesforce Marketing Cloud. Use Salesforce Marketing Cloud data extensions as sources or targets in a mapping. You can validate and run Salesforce Marketing Cloud mappings in the native environment.

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

PowerExchange Adapters for PowerCenter

This section describes new PowerCenter adapters in version 10.4.0.

PowerExchange for Db2 Warehouse

Effective in version 10.4.0, you can create a Db2 Warehouse connection to connect to IBM Db2 Warehouse from PowerCenter. You can import Db2 Warehouse objects as sources and targets to create mappings, and run a session to read data from or write data to Db2 Warehouse.

For more information, see the *Informatica 10.4.0 PowerExchange for Db2 Warehouse User Guide*.

PowerExchange for Microsoft Dynamics 365 for Sales

Effective in version 10.4.0, you can create a Microsoft Dynamics 365 for Sales connection to connect to Microsoft Dynamics 365 for Sales from PowerCenter. You can import Microsoft Dynamics 365 for Sales objects as sources and targets to create mappings, and run a session to extract read data from or write data to Microsoft Dynamics 365 for Sales. When you use Microsoft Dynamics 365 for Sales objects in mappings, you can configure properties specific to Microsoft Dynamics 365 for Sales.

For more information, see the *Informatica 10.4.0 PowerExchange for Microsoft Dynamics 365 for Sales User Guide*.

PowerExchange for PostgreSQL

Effective in version 10.4.0, you can create a PostgreSQL connection to connect to Aurora PostgreSQL from PowerCenter. You can import PostgreSQL objects as sources and targets to create mappings, and run a session to extract read data from or write data to PostgreSQL. When you use PostgreSQL objects in mappings, you can configure properties specific to PostgreSQL.

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

Product Name Changes

Effective in version 10.4.0, Informatica implemented some name changes.

The Big Data product family is renamed to Data Engineering. The following product names changed:

- Big Data Management changed to Data Engineering Integration.
- Big Data Quality changed to Data Engineering Quality.
- Big Data Streaming changed to Data Engineering Streaming.
- Big Data Masking changed to Data Engineering Masking.

Enterprise Data Catalog and Enterprise Data Preparation are aligned within the Data Catalog product family.

Upgrading to Version 10.4.0

Installer Changes

Effective in version 10.4.0, the Informatica installer has the following changes:

- You can run the 10.4.0 installer to install Data Engineering, Data Catalog, and traditional products. While you can install traditional products in the same domain with Data Engineering and Data Catalog products, Informatica recommends that you install the traditional products in separate domains.
- You can run the 10.4.0 installer to upgrade Data Engineering, Data Catalog, and traditional products.
- When you create a domain, you can choose to create the PowerCenter Repository Service and the PowerCenter Integration Service.

Effective in version 10.4.0, the Informatica upgrade has the following change:

- The Search Service creates a new index folder and re-indexes search objects. You do not need to perform the re-index after you upgrade.

Running Mappings with High or Low Precision

Effective in version 10.2.2, mappings that use the Spark engine run in high precision mode by default.

You can disable the high precision setting in the Developer tool in the mapping run-time configurations advanced settings:

1. In the Developer tool, select **Window** > **Preferences**.
2. Select **Informatica** > **Run Configurations** > **Mapping**.
3. Select the **Advanced** tab.
4. Deselect **High precision**.

You might want to do this to run mappings on the Spark engine that were developed in an earlier release and that you prefer to run with low precision. In some cases, such as a mapping in which scale is not specified, mappings fail at the default high precision setting and require low precision to run.

Effective in version 10.4.0, the following additional changes affect mappings that run on the Spark engine:

- Decimal to string conversion no longer appends a decimal to integers. For example, the value of integer 1 in decimal format remains "1" in string format.
- In decimal to string conversion, trailing zeroes after the decimal point are trimmed. For example, the value of 1.000 is decimal format is rendered as "1" in string format.
- When you run a mapping in low precision mode, decimal to string conversion for a precision setting of more than 15 results in exponential notation.

Support Changes

This section describes the support changes in version 10.4.0.

Technical Preview Support

Technical Preview Initiated

Effective in version 10.4.0, Informatica includes the following functionalities for technical preview:

Connect to a blockchain

For Data Engineering Integration, you can connect to a blockchain to use blockchain sources and targets in mappings that run on the Spark engine.

Data preview with Data Integration Service running on a grid

For Data Engineering Integration, you can preview hierarchical data on the Spark engine when the Data Integration Service runs on a grid.

Databricks delta table as streaming mapping target

For Data Engineering Streaming, you can use Databricks delta table as a target of streaming mapping for the ingestion of streaming data.

Dynamic streaming mapping

You can configure dynamic streaming mappings to change Kafka sources and targets at run time based on the parameters and rules that you define in a Confluent Schema Registry.

HL7 input in intelligent structure models

Intelligent Structure Discovery can process HL7 inputs.

Python transformation on Databricks

For Data Engineering Integration, you can include the Python transformation in mappings configured to run on the Databricks Spark engine.

Snowflake as a streaming mapping target

For Data Engineering Streaming, you can configure Snowflake as a target in a streaming mapping to write data to Snowflake.

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

Technical Preview Lifted

Effective in version 10.4.0, the following functionalities are lifted from technical preview:

Data preview on the Spark engine

For Data Engineering Integration, you can preview data within a mapping that runs on the Spark engine from the Developer tool for mappings configured to run with Amazon EMR, Cloudera CDH, and Hortonworks HDP. Previewing data in mappings configured to run with Azure HDInsight and MapR is still available for technical preview.

PowerExchange for Amazon S3

For Data Engineering Integration, you can use intelligent structure models when importing a data object.

PowerExchange for Microsoft Azure Cosmos DB SQL API

For Data Engineering Integration, you can develop and run mappings in the Azure Databricks environment.

PowerExchange for Microsoft Azure SQL Data Warehouse

For Data Engineering Integration, you can use the following functionalities:

- Create and run dynamic mappings.
- Use full pushdown optimization when an ODBC connection is used to connect to the Microsoft Azure SQL Data Warehouse database.

SSL-enabled Kafka connections

For Data Engineering Streaming, you can use SSL-enabled Kafka connections for streaming mappings.

Deferment

This section describes the deferment changes in version 10.4.0.

Deferment Lifted

Effective in version 10.4.0, the following functionalities are no longer deferred:

- Data Masking transformation in streaming mappings.
- Kerberos cross realm authentication.
- Monitoring statistics for streaming jobs.

Dropped Support

Effective in version 10.4.0, Informatica dropped support for Solaris. If you are using Solaris, Informatica requires you to upgrade to use a supported operating system.

For more information about how to upgrade to a supported operating system, see the Informatica 10.4.0 upgrade guides. For information about supported operating systems, see the Product Availability Matrix on the Informatica Network:

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

PowerCenter

This section describes PowerCenter support changes in version 10.4.0.

Connectivity

This section describes the connectivity support changes in version 10.4.0.

Licensing for SAP HANA

Effective in version 10.4.0, you need an SAP HANA license to read data from SAP HANA sources and write data to SAP HANA targets.

In the ODBC connection, if the **ODBC subtype** is not set to **SAP HANA** and the SAP HANA license is not available, then sessions fail at run time.

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

PowerExchange Adapters for PowerCenter

This section describes support changes to PowerCenter adapters in 10.4.0.

PowerExchange for SAP NetWeaver

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

- Informatica dropped support for non-Unicode transports.

  Previously, Informatica supported non-Unicode transports.
- Informatica ships transports for SAP Unicode versions 5.0 and later in the following folders:

  - Unicode cofiles: `Informatica installer zip file/saptrans/mySAP/cofiles`
  - Unicode data files: `Informatica installer zip file/saptrans/mySAP/data`

  Previously, Informatica packaged the transports for SAP Unicode versions 5.0 and later 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`
- Informatica dropped support for reading data from SAP tables through HTTP/HTTPS streaming with PowerExchange for SAP NetWeaver. Use PowerExchange for SAP Dynamic ABAP Table Extractor to read data from SAP tables through HTTP/HTTPS streaming.
  Previously, Informatica supported reading data from SAP tables through HTTP/HTTPS streaming with PowerExchange for SAP NetWeaver.

For more information, see the *Informatica 10.4.0 PowerExchange for SAP NetWeaver User Guide* and *PowerExchange for SAP NetWeaver 10.4.0 Transport Versions Installation Notice*.

PowerExchange Adapters for Informatica

This section describes support changes to Informatica adapters in 10.4.0.

PowerExchange for SAP NetWeaver

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

- Informatica ships transports for SAP Unicode versions 5.0 and later in the following folders:

  - Unicode cofiles: `Informatica installer zip file/saptrans/mySAP/cofiles`
  - Unicode data files: `Informatica installer zip file/saptrans/mySAP/data`

  Previously, Informatica packaged the transports for SAP Unicode versions 5.0 and later 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`

For more information, see the *PowerExchange for SAP NetWeaver 10.4.0 Transport Versions Installation Notice*.

Release Tasks

This section describes release tasks in version 10.4.0. Release tasks are tasks that you must perform after you upgrade to version 10.4.0.

Data Engineering Integration

This section describes release tasks for Data Engineering Integration in version 10.4.0.

Python Transformation

Effective in version 10.4.0, the Python code component in the Python transformation is divided into the following tabs:

- Pre-Input. Defines code that can be interpreted once and shared between all rows of data.
- On Input. Defines how a Python transformation behaves when it receives an input row while processing a partition.
- At End. Defines how a Python transformation behaves after it processes all input data in a partition.

In upgraded mappings, the code that you entered in the Python code component appears on the **On Input** tab.

Review the code to verify that the code operates as expected. If necessary, refactor the code using the **Pre-Input**, **On Input**, and **At End** tabs.

For information about how to configure code on each tab, see the "Python Transformation" chapter in the *Informatica Data Engineering Integration 10.4.0 User Guide*.

CHAPTER 22

10.4 What's New

This chapter includes the following topics:

- CI/CD, 168
- Command Line Programs, 173
- Data Engineering Integration, 173
- Data Engineering Streaming , 177
- Enterprise Data Catalog, 179
- Enterprise Data Preparation, 183
- Informatica Mappings , 185
- Informatica Transformations, 186
- Informatica Workflows, 186
- Informatica Installation, 187
- Intelligent Structure Model, 187
- PowerCenter, 188
- PowerExchange Adapters, 189
- Security, 193

CI/CD

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

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

Data Integration Service REST API

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

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

Query objects.

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

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

Perform version control operations.

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

Manage tags.

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

Update applications.

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

Manage applications.

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

Run deployed mappings.

Run a deployed mapping to test application output.

Compare mappings.

Compare two mappings in the same domain.

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

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

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

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

infacmd dis Commands

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

The following table describes the new infacmd dis commands:

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

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

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

Reverse Proxy Server

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

The REST Operations Hub is grid enabled.

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

infacmd roh Commands

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

The following table describes the new infacmd roh commands:

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

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

- listROHProperties to listProcessProperties.
- updateROHService to updateServiceProcessOptions.

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

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

Application Patch Deployment

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

Incremental Deployment Wizard

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

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

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

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

Patch History

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

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

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

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

Application Patch Archive Files

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

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

infacmd tools Commands

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

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

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

Connect to a Run-time Application

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

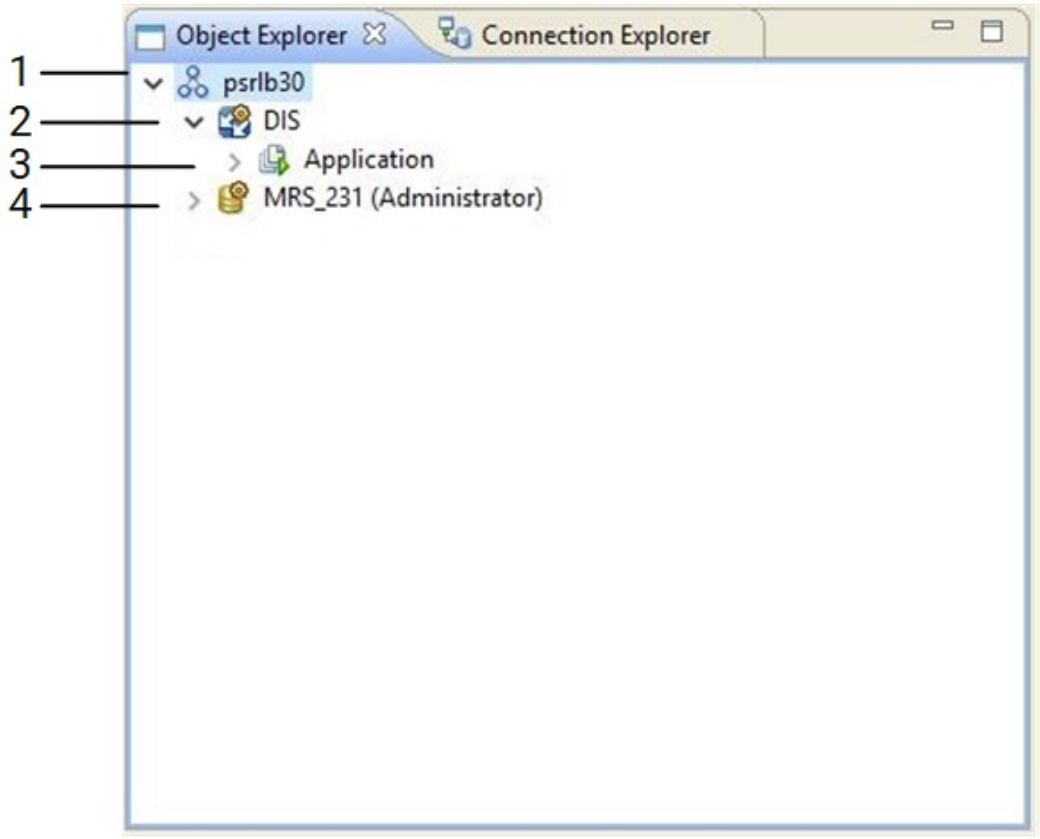
After you connect to a run-time application, the searches that you perform in the Developer tool can locate run-time objects in the application.

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

Object Explorer View

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

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



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

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

Tags

Effective in version 10.4.0, tags have the following functionality:

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

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

Command Line Programs

This section describes new commands in version 10.4.0.

infacmd isp Commands

The following table describes new infacmd isp commands:

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

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

Data Engineering Integration

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

New Data Types Support

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

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

The new data types are applicable to the following adapters:

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

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

AWS Databricks Integration

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

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

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

AWS Databricks supports the same data types as Azure Databricks.

For more information, refer to the following guides:

Data Engineering 10.4.0 Integration Guide

Data Engineering 10.4.0 Administrator Guide

Data Engineering Integration10.4.0 User Guide

Informatica 10.4.0 Developer Workflow Guide

Cluster Workflows for HDInsight Access to ALDS Gen2 Resources

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

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

Databricks Delta Lake Storage Access

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

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

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

Display Nodes Used in Mapping

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

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

For more information about the REST API, see the "Monitoring REST API Reference" chapter in the *Data Engineering10.4.0 Administrator Guide*

Log Aggregation

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

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

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

Parsing Hierarchical Data on the Spark Engine

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

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

- PARSE_JSON
- PARSE_XML

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

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

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

Profiles and Sampling Options on the Spark Engine

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

Profiling on the Spark engine

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

Sampling options on the Spark engine

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

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

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

Python Transformation

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

Active Mode

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

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

Partitioned Data

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

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

Sqoop

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

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

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

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

Data Engineering Streaming

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

Confluent Schema Registry in Streaming Mappings

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

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

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

Data Quality Transformations in Streaming Mappings

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

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

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

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

Ephemeral Cluster in Streaming Mappings

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

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

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

FileName Port in Amazon S3

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

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

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

Microsoft Azure Data Lake Storage Gen2

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

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

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

Streaming Mappings in Azure Databricks

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

Sources and Targets

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

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

Transformations

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

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

Data Types

The following data types are supported:

- Array
- Bigint
- Date/time

Decimal

Double

Integer

Map

Struct

Text

String

Workflows

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

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

Dynamic Mappings in Data Engineering Streaming

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

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

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

Enterprise Data Catalog

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

Assigning Custom Attributes to Resources and Classes

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

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

New Resources

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

- AWS Glue
- Microsoft Power BI
- Apache Cassandra

- Snowflake
- Google Cloud Storage

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

Reference Resources and Reference Assets

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

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

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

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

Export Assets from Search Results

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

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

Lineage and Impact Filters

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

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

Asset Control Summary

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

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

Rules and Scorecards

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

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

Unique Key Inference

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

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

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

Data Domain Discovery on the CLOB File Type

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

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

Data Discovery and Sampling Options on the Spark Engine

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

Data Domain Discovery on the Spark Engine

You can perform data domain discovery on the Spark engine.

Sampling Options on the Spark Engine

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

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

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

Tracking Technical Preview

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

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

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

Data Preview and Provisioning

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

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

Supported Resource Types for Standalone Scanner Utility

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

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

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

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

REST APIs

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

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

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

Enterprise Data Preparation

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

Data Lake Access Management

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

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

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

Microsoft Azure Data Lake Storage as a Data Source

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

Publish Files to the Data Lake

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

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

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

Upload Files to the Data Lake

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

You can choose from the following upload options:

Upload a file directly to the data lake.

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

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

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

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

Note: Uploading an Excel spreadsheet is a preview feature.

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

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

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

Informatica Mappings

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

Binding Mapping Outputs to Mapping Parameters

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

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

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

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

infacmd ms Commands

The following table describes new and updated infacmd ms commands:

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

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

CLAIRE Recommendations and Insights

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

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

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

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

Update Mapping Optimizer Level

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

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

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

Informatica Transformations

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

Address Validator Transformation

This section describes new Address Validator transformation features.

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

United States

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

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

Informatica Workflows

This section describes new Informatica workflow features in version 10.4.0.

Amazon EMR Create Cluster Task Advanced Properties

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

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

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

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

Informatica Installation

This section describes new installation features in 10.4.0.

PostgreSQL

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

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

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

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

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

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

Encrypt Passwords in the Silent Installation Properties File

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

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

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

Intelligent Structure Model

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

Additional Input Types

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

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

Create Model from Sample at Design Time

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

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

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

Handling of Unidentified Data

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

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

PowerCenter

This section describes new PowerCenter features in version 10.4.0.

HTTP Transformation

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

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

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

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

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

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

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

Connectivity

This section describes new connectivity features in version 10.4.0.

User Impersonation for Oracle Connections

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

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

Support for SAP HANA Database Modelling Views

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

- Analytics Views
- Attribute Views
- Calculated Views

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

PowerExchange Adapters

This section describes new PowerExchange adapter features in version 10.4.0.

PowerExchange Adapters for Informatica

This section describes new Informatica adapter features in version 10.4.0.

PowerExchange for Amazon Redshift

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

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

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

PowerExchange for Amazon S3

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

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

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

PowerExchange for Google BigQuery

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

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

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

PowerExchange for Google Cloud Storage

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

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

  Overrides the Google Cloud Storage path to the file that you selected in the Google Cloud Storage data object.
  Use the following format:

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

  Source File Name

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

  Is Directory

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

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

PowerExchange for Microsoft Azure Blob Storage

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

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

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

PowerExchange for Microsoft Azure SQL Data Warehouse

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

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

PowerExchange for Salesforce

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

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

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

PowerExchange for SAP NetWeaver

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

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

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

PowerExchange for Snowflake

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

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

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

PowerExchange for HDFS

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

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

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

PowerExchange Adapters for PowerCenter

This section describes new PowerCenter adapter features in version 10.4.0.

PowerExchange for Google BigQuery

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

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

PowerExchange for Google Cloud Storage

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

Google Cloud Storage Path

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

Use the following format:

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

Source File Name

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

Is Directory

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

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

PowerExchange for Greenplum

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

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

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

PowerExchange for JD Edwards EnterpriseOne

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

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

PowerExchange for Kafka

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

- SSL Mode

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

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

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

PowerExchange for Salesforce

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

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

PowerExchange for SAP NetWeaver

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

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

Security

This section describes new security features in version 10.4.0.

Configure Web Applications to Use Different SAML Identity Providers

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

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

CHAPTER 23

10.4 What's Changed

This chapter includes the following topics:

- Data Engineering Integration, 194
- Enterprise Data Preparation, 196
- Enterprise Data Catalog, 196
- Informatica Developer, 197
- Informatica Transformations, 198
- PowerCenter, 198
- PowerExchange Adapters for Informatica, 199
- Security, 200

Data Engineering Integration

This section describes the changes to Data Engineering Integration in version 10.4.0.

Data Preview

Effective in version 10.4.0, the Data Integration Service uses Spark Jobserver to preview data on the Spark engine. Spark Jobserver allows for faster data preview jobs because it maintains a running Spark context instead of refreshing the context for each job. Mappings configured to run with Amazon EMR, Cloudera CDH, and Hortonworks HDP use Spark Jobserver to preview data.

Previously, the Data Integration Service used spark-submit scripts for all data preview jobs on the Spark engine. Mappings configured to run with Azure HDInsight and MapR use spark-submit scripts to preview data on the Spark engine. Previewing data on mappings configured to run with Azure HDInsight and MapR is available for technical preview.

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

Union Transformation

Effective in version 10.4.0, you can choose a Union transformation as the preview point when you preview data. Previously, the Union transformation was not supported as a preview point.

infacmd dp Commands

You can use the infacmd dp plugin to perform data preview operations. Use infacmd dp commands to manually start and stop the Spark Jobserver.

The following table describes infacmd dp commands:

Command	Description
startSparkJobServer	Starts Spark Jobserver on the Integration Service machine. By default, the Spark Jobserver starts when you preview hierarchical data.
stopSparkJobServer	Stops the Spark Jobserver running on specified Integration Service. By default, the Spark Jobserver stops if it is idle for 60 minutes or when the Data Integration Service is stopped or recycled.

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

Date/Time Format on Databricks

Effective in version 10.4.0, when the Databricks Spark engine reads and writes date/time values, it uses the format YYYY-MM-DD HH24:MM:SS.US.

Previously, you set the format in the mapping properties for the run-time preferences of the Developer tool.

You might need to perform additional tasks to continue using date/time data on the Databricks engine. For more information, see the "Databricks Integration" chapter in the *Data Engineering 10.4.0 Integration Guide*.

Null Values in Target

Effective in version 10.4.0, the following changes are applicable when you write data to a complex file:

- If the mapping source contains null values and you use the **Create Target** option to create a Parquet target file, the default schema contains optional fields and you can insert null values to the target.

  Previously, all the fields were created as REQUIRED in the default schema and you had to manually update the data type in the target schema from "required" to "Optional" to write the columns with null values to the target.
- If the mapping source contains null values and you use the **Create Target** option to create an Avro target file, null values are defined in the default schema and you can insert null values into the target file.

  Previously, the null values were not defined in the default schema and you had to manually update the default target schema to add "null" data type to the schema.

Note: You can manually edit the schema if you do not want to allow null values to the target. You cannot edit the schema to prevent the null values in the target with mapping flow enabled.

These changes are applicable to the following adapters:

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

Python Transformation

Effective in version 10.4.0, you access resource files in the Python code by referencing an index in the array `resourceFilesArray`. Use `resourceFilesArray` in new mappings that you create in version 10.4.0.

Previously, the array was named `resourceJepFile`. Upgraded mappings that use `resourceJepFile` will continue to run successfully.

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

Enterprise Data Preparation

This section describes changes to Enterprise Data Preparation in version 10.4.0.

NULL Value Handling

Effective in version 10.4.0, Enterprise Data Preparation handles NULL values following Informatica Data Engineering processing behavior. After you upgrade the application, formulas in recipes that operate on worksheet columns containing a NULL value or an error value return NULL.

When you open a project created in a previous release, a dialog opens prompting you to upgrade the worksheets in the project. If you choose to upgrade the worksheets, the application recalculates the formulas in each worksheet in the project, and then updates the worksheet with the new formula results.

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

Solr Port Property

Effective in version 10.4.0, the Solr port property is removed from the Interactive Data Preparation Service.

Previously, the Enterprise Data Preparation application used Apache Solr to recommend steps to add to a recipe during data preparation. The application now uses an internal algorithm to recommend steps to add to a recipe.

Enterprise Data Catalog

This section describes changes to Enterprise Data Catalog in version 10.4.0.

Lineage Enhancement for SAP HANA Resource

Effective in version 10.4.0, you can use the SAP HANA resource to view lineage for the following views:

- Tables and Normal Views
- Calculation Views
- Analytical Views

Effective in version 10.4.0, Enterprise Data Catalog now supports profiling the SAP HANA database tables and views to extract column profiling and domain discovery statistics.

Profiling and Data Domain Discovery

Effective in version 10.4.0, you can enable column profiling and data domain discovery for the Google BigQuery resource.

Informatica Data Engineering Streaming

Effective in version 10.4.0, the Informatica Platform resource supports metadata extraction from streaming mapping including streaming sources. The streaming data sources are created as reference objects.

The supported streaming source is Apache Kafka. Strong type reference objects are supported for the Apache Kafka and Hive data sources.

Resource Name Changes

Effective in version 10.4.0, the Profiling Warehouse resource is renamed to Informatica Data Quality.

Search Suggestions

Effective in version 10.4.0, Enterprise Data Catalog now displays both business titles and asset names as probable matches in the search suggestions. Previously, Enterprise Data Catalog displayed only asset names as probable matches in the search suggestions.

Configuring Custom Attributes

Effective in version 10.4.0, you can now prioritize the search rank for custom attributes using the **High** and **Low** options. Previously, you could prioritize the search rank for custom attributes using **High**, **Medium**, and **Low** options.

Informatica Developer

This section describes changes to Informatica Developer in version 10.4.0.

Importing Relational Data Objects

Effective in version 10.4.0, when you import a resource with the same name as an existing physical data object, the Developer tool allows you to choose how to resolve the conflict. You can choose to create an object with a different name, reuse the existing object, or replace the resource for the existing object.

Previously, the Developer tool failed the table import and did not attempt to import any subsequent tables.

For more information, see the *Informatica® 10.4.0 Developer Tool Guide*.

Informatica Transformations

This section describes changes to Informatica transformations in version 10.4.0.

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 10.4, the Address Validator transformation uses version 5.15.0 of the Informatica Address Verification software engine.

Previously, the transformation used version 5.14.0 of the Informatica Address Verification software engine.

Effective in version 10.4, the Address Validator transformation retains province information in an output address when the reference data does not contain province information for the country. If the output address is valid without the province data, Address Verification returns a V2 match score to indicate that the input address is correct but that the reference database does not contain every element in the address.

Previously, if the address reference data did not contain province information for the country, Address Verification moved the province information to the Residue field and returned a Cx score.

Spain

Effective in version 10.4, the Address Validator transformation returns an Ix status for a Spain address that requires significant correction to yield a match in the reference data.

Previously, Address Verification might correct an address that required significant changes and return an overly optimistic match score for the address.

United States

Effective in version 10.4, the Address Validator transformation can validate a United States address when organization information precedes street information on a delivery address line. The types of organization that the transformation recognizes include universities, hospitals, and corporate offices. Address Verification recognizes the organization information when the parsing operation also finds a house number and street type in the street information on the delivery address line.

Previously, Address Verification returned an Ix match score when organization information preceded street information on a delivery line.

For comprehensive information about updates to the Informatica Address Verification software engine, see the *Informatica Address Verification 5.15.0 Release Guide*.

PowerCenter

This section describes the changes to PowerCenter in 10.4.0.

Refresh Metadata in Designer and in the Workflow Manager

Effective in version 10.4.0, you can refresh the repository and the folder in the Workflow Manager and the Designer without disturbing the connection state. Repository and folder updates occur when you create, delete, modify folder, or when you import objects in to the PowerCenter client.

To refresh the folder list in a repository, right-click the repository and select Refresh Folder List. To refresh the folder and all the contents within the folder, right-click the folder and select Refresh.

Previously, you had to disconnect and reconnect the PowerCenter clients to view the updates to the repository or at the folder.

For more information, see the *PowerCenter 10.4.0 Repository Guide*, *PowerCenter 10.4.0 Designer Guide*, and *PowerCenter 10.4.0 Workflow Basics Guide*.

Import and Export

Effective in version 10.4.0, you can import and export mappings, sessions, workflows, and worklets.

To import data from PowerCenter into the Model Repository, complete the following tasks:

1. Export PowerCenter objects to a file using the PowerCenter Client or with the following command:
   `pmrep ExportObject`
2. Convert the export file to a Model repository file using the following command:
   `infacmd ipc importFromPC`
3. Import the objects using the Developer tool or with the following command:
   `infacmd tools importObjects`

To export data from the Model repository into the PowerCenter repository, complete the following tasks:

1. Export Model Repository objects to a file using the Developer tool or with the following command:
   `infacmd tools ExportObjects`

   Or, you can directly run `infacmd ipc ExportToPC` to export.
2. Convert the export file to a PowerCenter file using the following command:
   `infacmd ipc ExporttoPC`
3. Import the objects using the PowerCenter or with the following command:
   `pmrep importObjects`

In versions 10.2.2 and 10.2.1, you could import from PowerCenter with the installer patches but you could not export to PowerCenter. Prior to version 10.2.1, you could import from and export to PowerCenter.

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

PowerExchange Adapters for Informatica

This section describes changes to Informatica adapters in version 10.4.0.

PowerExchange for Amazon Redshift

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

- You do not need to add the `GetBucketPolicy` permission in the Amazon S3 bucket policy to connect to Amazon Redshift. Previously, you had to add the `GetBucketPolicy` permission in the Amazon S3 bucket policy to connect to Amazon Redshift. The existing Amazon S3 bucket policies with the `GetBucketPolicy` permission continue to work without any change.
- The following advanced properties for an Amazon Redshift data object read operation are changed:

Old Property	New Property
S3 Server Side Encryption S3 Client Side Encryption	Encryption Type You can select one of the following values: - None - SSE-S3 - SSE-KMS - CSE-SMK

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

PowerExchange for Amazon S3

Effective in version 10.4.0, you do not need to add the `GetBucketPolicy` permission in the Amazon S3 bucket policy to connect to Amazon S3.

Previously, you had to add the `GetBucketPolicy` permission in the Amazon S3 bucket policy to connect to Amazon S3.

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

PowerExchange for Microsoft Azure Blob Storage

Effective in version 10.4.0, when you run a passthrough mapping on the Databricks Spark engine, only the connected fields are written to the flat file target. Previously, the Data Integration Service wrote unconnected field names with empty values to the flat file target.

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

PowerExchange for Microsoft Azure Data Lake Storage Gen1

Effective in version 10.4.0, when you run a passthrough mapping on the Databricks Spark engine, only the connected fields are written to the flat file target. Previously, the Data Integration Service wrote unconnected field names with empty values to the flat file target.

For more information, see the *Informatica 10.4.0 PowerExchange for Microsoft Azure Data Lake Storage Gen1 User Guide*.

Security

This section describes changes to security features in version 10.4.0.

infacmd isp Commands

The following table describes changed infacmd isp commands:

Command	Description
addNameSpace	The required **-ln** option is added to the command. You use the option to specify the name of the LDAP configuration.
listLDAPConnectivity	The required **-ln** option is added to the command.
setLDAPConnectivity	The command is renamed addLDAPConnectivity. Update any scripts that use setLDAPConnectivity with the new command syntax. The required **-ln** option is added to the command.
updateNameSpace	The required **-ln** option is added to the command.

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

LDAP Directory Services

Effective in version 10.4.0, Informatica supports the following LDAP directory services:

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

You can also configure a domain to use an LDAP directory service that Informatica does not certify.

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

LDAP Configurations

Effective in version 10.4.0, you can configure an Informatica domain to enable users imported from one or more LDAP directory services to log in to Informatica nodes, services, and application clients.

Previously, you could configure an Informatica domain to import users from a single LDAP directory service.

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

SAML Authentication

Effective in version 10.4.0, Informatica supports the following identity providers:

- Microsoft Active Directory Federation Services (AD FS) 2.0
- Microsoft Active Directory Federation Services (AD FS) 4.0
- PingFederate 9.2.2

Previously, Informatica only supported the AD FS 2.0 identity provider.

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

Part VI: Versions 10.2.2 - 10.2.2 HotFix 1

This part contains the following chapters:

- Version 10.2.2 HotFix 1, 203
- Version 10.2.2 Service Pack 1, 211
- 10.2.2 Important Notices, 218
- 10.2.2 What's New, 225
- 10.2.2 What's Changed, 256

CHAPTER 24

Version 10.2.2 HotFix 1

This chapter includes the following topics:

- Important Notices, 203
- What's New, 204
- What's Changed, 208

Important Notices

Support Changes

This section describes the support changes in version 10.2.2 HotFix 1 Service Pack 1.

Technical Preview Support

Technical Preview Initiated

Effective in version 10.2.2 HotFix 1 Service Pack 1, Informatica includes the following functionality for technical preview:

Hive Warehouse Connector and Hive LLAP

For Big Data products, use the Hive Warehouse Connector and Hive LLAP with Azure HDInsight 4.x and Hortonworks HDP 3.x to enable Spark code to interact with Hive tables and allow ACID-enabled Hive tables on the Spark engine.

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

Technical Preview Lifted

Effective in version 10.2.2 HotFix 1 Service Pack 1, Hortonworks HDP 3.1 is lifted from technical preview.

The upgrade to HDP 3.1 might affect managed Hive tables. Before you upgrade, review HDP 3.1 upgrade information and Cloudera known limitations in the KB article
What Should Big Data Customers Know About Upgrading to Hortonworks HDP 3.1?. Contact Informatica Global Customer Services or Cloudera Professional Services to validate your upgrade plans to HDP 3.1.

What's New

Command Line Programs

This section describes new commands in version 10.2.2 HotFix 1.

infacmd ldm Commands

The following table describes new infacmd ldm command options:

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

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

Big Data Management

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

Azure Data Lake Storage Gen2

Effective in version 10.2.2 HotFix 1 Service Pack 1, mappings can read and write to an Azure Data Lake Storage Gen2 storage account associated with an HDInsight cluster.

For more information on how to integrate Big Data Management with a cluster that uses ADLS Gen2 storage, see the *Big Data Management Integration Guide*. For information on how to use mappings with ADLS Gen2 sources and targets, see the *Big Data Management User Guide*.

Enterprise Data Catalog

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

Azure Data Lake Storage Gen2

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

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

Case Insensitive Linking

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

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

Enterprise Data Catalog Tableau Extension

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

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

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

New Resources

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

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

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

Profiling Warehouse Scanner

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

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

REST APIs

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

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

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

Search Enhancements

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

Double Quotes Search

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

Search Operators

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

Search Rank

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

Related Search

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

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

Search Tabs

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

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

Supported Resource Types for Standalone Scanner Utility

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

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

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

Technical Preview

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

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

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

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

PowerExchange Adapters for Informatica

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

PowerExchange for Google Analytics

Effective in version 10.2.2 HotFix 1 Service Pack 1, you can run PowerExchange for Google Analytics mappings on the Spark engine on the following distributions:

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

PowerExchange for Google BigQuery

Effective in version 10.2.2 HotFix 1 Service Pack 1, you can run PowerExchange for Google BigQuery mappings on the Spark engine on the following distributions:

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

PowerExchange for Google Cloud Spanner

Effective in version 10.2.2 HotFix 1 Service Pack 1, you can run PowerExchange for Google Cloud Spanner mappings on the Spark engine on the following distributions:

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

PowerExchange for Google Cloud Storage

Effective in version 10.2.2 HotFix 1 Service Pack 1, you can run PowerExchange for Google Cloud Storage mappings on the Spark engine on the following distributions:

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

What's Changed

Data Transformation

This section describes changes to Data Transformation in version 10.2.2 HotFix 1.

Business Intelligence and Reporting Tools (BIRT)

Effective in version 10.2.2 HotFix 1, Business Intelligence and Reporting Tools (BIRT) is no longer packaged with Data Transformation.

You can download BIRT from the location mentioned in the following file:

```
<Data Transformation installation directory>/readme_Birt.txt
```

If you try to use BIRT from Data Transformation before you download it, you might get the following error:

```
The Birt Report Engine was not found. See download instructions at [DT-home]/readme_Birt.txt.
```

Previously, BIRT was packaged with Data Transformation.

Enterprise Data Preparation

This section describes changes to Enterprise Data Preparation in version 10.2.2 HotFix 1.

Data Asset Search

Effective in version 10.2.2 HotFix 1, the data asset search capabilities and user interface are updated to reflect changes made in Enterprise Data Catalog.

For more information, see the *Informatica 10.2.2 HotFix 1 Enterprise Data Preparation User Guide*.

Enterprise Data Catalog

This section describes changes to Enterprise Data Catalog in version 10.2.2 HotFix 1.

Business Term Propagation

Effective in version 10.2.2 HotFix 1, you can now configure the **Business Term Association** settings for a resource to automatically associate business titles to assets. Previously, you could automatically associate business titles to assets only by configuring the **Run Similarity Profile** option. You can also automatically associate business titles to tables and file assets. Previously, you could only automatically associate business titles to columns and field assets.

For more information about business term propagation, see the "Enterprise Data Catalog Concepts" chapter in the *Informatica 10.2.2 HotFix 1 Catalog Administrator Guide*.

Custom Resources

Effective in version 10.2.2 HotFix 1, the custom resources include the following enhancements:

View Detailed or Summary Lineage for ETL Transformations

You can configure custom ETL resources to view the detailed and summary lineage for ETL transformations between multiple data sources.

Specify a Remote File Path for the Metadata ZIP File

When you configure a custom resource, you can provide a remote path to the ZIP file that includes the custom metadata that you want to ingest into the catalog. You can use the remote file path to automate and schedule custom resource scans.

Configure the Custom Resource to Use a User-defined Script to Extract Metadata

If you have automated the process to prepare custom metadata and generate the ZIP file using a script or a sequence of commands, you can provide the details when you configure the resource. Enterprise Data Catalog runs the script before running the resource.

Configure Custom Icons for Custom Resource Data Assets

You can customize and configure icons for classes that you define in the custom model. The icons appear in Enterprise Data Catalog to represent the data assets from a custom resource.

Configure Relationships Views for Custom Resources

You can configure the **Relationships Views** page in Enterprise Data Catalog for custom resources. As part of the configuration, you can define a set of configurations for classes in the custom model. Based on the definitions, you can filter or group related objects for each class type and view the objects on the **Relationships Views** page.

For more information, see the *Informatica 10.2.2 HotFix 1 Catalog Administrator Guide*.

Documentation Changes

Effective in version 10.2.2 HotFix 1, all information related to creating and configuring resources are moved from the Catalog Administrator Guide to a new guide titled Informatica 10.2.2 HotFix 1 Enterprise Data Catalog Scanner Configuration Guide.

For more information, see the *Informatica 10.2.2 HotFix 1 Enterprise Data Catalog Scanner Configuration Guide*.

Microsoft SQL Server Resource

Effective in version 10.2.2 HotFix 1, the Microsoft SQL Server resource supports Windows-based authentication using the Enterprise Data Catalog agent.

For more information, see the *Informatica 10.2.2 HotFix 1 Scanner Configuration Guide*.

PowerExchange Adapters for Informatica

This section describes changes to Informatica adapters in version 10.2.2 HotFix 1 Service Pack 1.

PowerExchange for Amazon Redshift

Effective in version 10.2.2 HotFix 1 Service Pack 1, you do not need to add the `GetBucketPolicy` permission in the Amazon S3 bucket policy to connect to Amazon Redshift.

Previously, you had to add the `GetBucketPolicy` permission in the Amazon S3 bucket policy to connect to Amazon Redshift.

For more information, see the *Informatica PowerExchange for Amazon Redshift 10.2.2 HotFix 1 User Guide*.

CHAPTER 25

Version 10.2.2 Service Pack 1

This chapter includes the following topics:

- Important Notices, 211
- What's New, 212
- What's Changed, 215

Important Notices

Support Changes

This section describes the support changes in version 10.2.2 Service Pack 1.

Deferred Support

Effective in version 10.2.2 Service Pack 1, Informatica deferred support for the following functionality:

Streaming for the Data Masking Transformation

With the changes in streaming support, Data Masking transformation in streaming mappings is deferred.

Informatica intends to reinstate it in an upcoming release, but might choose not to in accordance with changing market or technical circumstances.

Technical Preview Support

Effective in version 10.2.2 Service Pack 1, Informatica includes the following functionality for technical preview:

SSL enabled Kafka connections for streaming

When you create a Kafka connection, you can use SSL connection properties to configure the Kafka broker.

Support for Hortonworks HDP 3.1

You can use Informatica big data products with Hortonworks HDP version 3.1.

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

to in accordance with changing market or technical circumstances. For more information, contact Informatica Global Customer Support.

Product and Service Name Changes

This section describes changes to product and service names in version 10.2.2 Service Pack 1.

The following product and service names are changed:

- The product Enterprise Data Lake is renamed to Enterprise Data Preparation.
- The Enterprise Data Lake Service is renamed to Enterprise Data Preparation Service.
- The Data Preparation Service is renamed to Interactive Data Preparation Service.

Release Tasks

This section describes release tasks in version 10.2.2 Service Pack 1. Release tasks are tasks that you must perform after you upgrade to version 10.2.2 Service Pack 1.

Sqoop Connectivity

Effective in version 10.2.2 Service Pack 1, the following release tasks apply to Sqoop:

- When you use Cloudera Connector Powered by Teradata Connector to run existing Sqoop mappings on the Spark or Blaze engine and on Cloudera CDH version 6.1.x, you must download the junit-4.11.jar and sqoop-connector-teradata-1.7c6.jar files.
  Before you run existing Sqoop mappings on Cloudera CDH version 6.1.x, perform the following tasks:
  1. Download and copy the junit-4.11.jar file from the following URL: http://central.maven.org/maven2/junit/junit/4.11/junit-4.11.jar
  2. On the node where the Data Integration Service runs, add the junit-4.11.jar file to the following directory: `<Informatica installation directory>\externaljdbcjars`
  3. Download and extract the Cloudera Connector Powered by Teradata package from the Cloudera web site and copy the following file: sqoop-connector-teradata-1.7c6.jar
  4. On the node where the Data Integration Service runs, add the sqoop-connector-teradata-1.7c6.jar file to the following directory: `<Informatica installation directory>\externaljdbcjars`
- To run Sqoop mappings on the Blaze or Spark engine and on Cloudera CDH, you no longer need to set the mapreduce.application.classpath entries in the mapred-site.xml file for MapReduce applications.
  If you use Cloudera CDH version 6.1.x to run existing Sqoop mappings, remove the mapreduce.application.classpath entries from the mapred-site.xml file.

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

What's New

Big Data Management

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

Sqoop

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

Connect to SSL-enabled Oracle and Microsoft SQL Server databases

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

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

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

Support for connection-param-file Sqoop argument

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

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

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

Use the following syntax:

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

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

Big Data Streaming

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

Amazon S3 Target

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

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

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

TIME_RANGE Function

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

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

Syntax

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

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

Writing Data to Multiple HDFS Files

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

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

Enterprise Data Catalog

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

Supported Resource Types for Standalone Scanner Utility

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

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

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

Enterprise Data Preparation

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

Revert All Inferred Data Types

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

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

Prepare Avro and Parquet Files

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

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

PowerExchange Adapters for Informatica

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

PowerExchange for Hive

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

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

What's Changed

Big Data Management

This section describes the changes to Big Data Management in version 10.2.2 Service Pack 1.

Data Integration Service Queues

Effective in version 10.2.2 Service Pack 1, the Data Integration Service uses a distributed queue to store job information by default.

You can enable local queuing only by using a custom property. If you require this functionality, contact Informatica Global Support.

Previously, the Data Integration Service used a local queue on each node by default, and used the distributed queue only for Spark jobs when big data recovery was enabled.

For more information, see the "Data Integration Service Processing" chapter in the *Informatica Big Data Management 10.2.2 Service Pack 1 Administrator Guide*.

Mass Ingestion

Effective in version 10.2.2 Service Pack 1, you can select the cluster default as the storage format for a mass ingestion specification that ingests data to a Hive target. If you select the cluster default, the specification uses the default storage format on the Hadoop cluster.

Previously, the specification used the default storage format on the cluster when you selected the text storage format. In 10.2.2 Service Pack 1, selecting the text storage format ingests data to a standard text file.

Big Data Streaming

This section describes the changes to Big Data Streaming in version 10.2.2 Service Pack 1.

Transformations

This section describes the changes to transformations in Big Data Streaming in version 10.2.2 Service Pack 1.

Rank Transformation

Effective in version 10.2.2 Service Pack 1, a streaming mapping must meet the following additional requirements if it contains a Rank transformation:

- A streaming mapping cannot contain a Rank transformation and a passive Lookup transformation that is configured with an inequality lookup condition in the same pipeline. Previously, you could use a Rank transformation and a passive Lookup transformation that is configured with an inequality lookup condition in the same pipeline.
- A Rank transformation in a streaming mapping cannot have a downstream Joiner transformation. Previously, you could use a Rank transformation anywhere before a Joiner transformation in a streaming mapping.
- A streaming mapping cannot contain more than one Rank transformation in the same pipeline. Previously, you could use multiple Rank transformations in a streaming mapping.
- A streaming mapping cannot contain an Aggregator transformation and a Rank transformation in the same pipeline. Previously, you could use an Aggregator transformation and a Rank transformation in the same pipeline.

Sorter Transformation

Effective in version 10.2.2 Service Pack 1, a streaming mapping must meet the following additional requirements if it contains a Sorter transformation:

- A streaming mapping runs in complete output mode if it contains a Sorter transformation. Previously, a streaming mapping used to run in append output mode if it contains a Sorter transformation.
- The Sorter transformation in a streaming mapping must have an upstream Aggregator transformation. Previously, you could use a Sorter transformation without an upstream Aggregator transformation.
- The Window transformation upstream from an Aggregator transformation will be ignored if the mapping contains a Sorter transformation. Previously, the Window transformation upstream from an Aggregator transformation was not ignored if the mapping contains a Sorter transformation.

Informatica Analyst

This section describes changes to the Analyst tool in version 10.2.2 Service Pack 1.

Default View

Effective in version 10.2.2 Service Pack 1, the default view for flat file and table objects is the Properties tab. When you create or open a flat file or table data object, the object opens in the Properties tab. Previously, the default view was the Data Viewer tab.

For more information, see the *Informatica 10.2.2 Service Pack 1 Analyst Tool Guide*.

PowerExchange Adapters for Informatica

This section describes changes to Informatica adapters in version 10.2.2 Service Pack 1.

PowerExchange for Amazon Redshift

Effective in version 10.2.2 Service Pack 1, PowerExchange for Amazon Redshift has the following changes:

- PowerExchange for Amazon Redshift supports the **Server Side Encryption With KMS** encryption type on the following distributions:

  - Amazon EMR version 5.20

  - Cloudera CDH version 5.16 and 6.1

  Previously, the Data Integration Service supported the **Server Side Encryption With KMS** encryption type on the following distributions:

  - Amazon EMR version 5.16

  - Cloudera CDH version 5.15

- You cannot use the following distributions to run Amazon Redshift mappings:

  - MapR version 5.2

  - IBM BigInsight

  Previously, you could use the MapR version 5.2 and IBM BigInsight distributions to run Amazon Redshift mappings.

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

PowerExchange for Amazon S3

Effective in version 10.2.2 Service Pack 1, PowerExchange for Amazon S3 has the following changes:

- PowerExchange for Amazon S3 supports the **Server Side Encryption With KMS** encryption type on the following distributions:

  - Amazon EMR version 5.20

  - Cloudera CDH version 5.16 and 6.1

  Previously, PowerExchange for Amazon S3 supported the **Server Side Encryption With KMS** encryption type on the following distributions:

  - Amazon EMR version 5.16

  - Cloudera CDH version 5.15.

- You cannot use the following distributions to run Amazon S3 mappings:

  - MapR version 5.2

  - IBM BigInsight

  Previously, you could use the MapR version 5.2 and IBM BigInsight distributions to run Amazon S3 mappings.

For more information, see the *Informatica PowerExchange for Amazon S3 10.2.2 Service Pack 1 User Guide*.

CHAPTER 26

10.2.2 Important Notices

This chapter includes the following topics:

- Installation and Configuration, 218
- Support Changes, 219
- Release Tasks, 221
- New Products, 224

Installation and Configuration

This section describes notices related to big data installation and configuration in version 10.2.2.

OpenJDK

Effective in version 10.2.2, Informatica installer packages OpenJDK (AzulJDK). The supported Java version is Azul OpenJDK 1.8.192.

You can use the OpenJDK to deploy Enterprise Data Catalog on an embedded cluster. To deploy Enterprise Data Catalog on an existing cluster, you must install JDK 1.8 on all the cluster nodes.

Informatica dropped support of the Data Integration Service property for the execution option, JDK Home Directory. Sqoop mappings on the Spark engine use the Java Development Kit (JDK) packed with the Informatica installer.

Previously, the installer used the Oracle Java packaged with the installer. You also had to install the JDK and then specify the JDK installation directory in the Data Integration Service machine to run Sqoop mappings, mass ingestion specifications that use a Sqoop connection on the Spark engine, or to process a Java transformation on the Spark engine.

Installer Code Signing

Effective in version 10.2.2, Informatica uses a certificate based digital signature to sign the Informatica software code. The code sign helps to validate and secure code from any form of tampering after the digital signature of the installer code.

Informatica packages the public key, signature, and hash of the file in the installer bundle. After Informatica signs the software bundle, you can contact Informatica Global Customer Support to access the public key.

For more information about the installer code signing process or on how a customer can verify that the signed code is authentic, see the *Informatica Big Data Suite 10.2.2 Installation and Configuration Guide*.

Resume the Installer

Effective in version 10.2.2, you can resume the installation process from the point of failure or exit. If a service fails or if the installation process fails during a service creation, you can resume the installation process with the server installer. You cannot resume the installer if you are running it to configure services after the services have been created. When you join the domain, you also cannot resume the installer.

For more information about resuming the installer, see the *Informatica Big Data Suite 10.2.2 Installation and Configuration Guide*.

Informatica Docker Utility

Effective in version 10.2.2, you can use the Informatica Docker utility to quickly install the Informatica domain.

When you run the Informatica Docker utility, you can build the Informatica docker image with the base operating system and Informatica binaries. You can run the existing docker image to configure the Informatica domain. When you run the Informatica docker image, you can create a domain or join a domain. You can create the Model Repository Service, Data Integration Service, and cluster configuration during the container creation.

Installer

This section describes the changes to the Informatica installer in version 10.2.2.

Configure Enterprise Data Catalog

If you install the Enterprise Data Catalog binaries on a node on which the Informatica services are installed, you can use the installer to configure the Enterprise Data Catalog services on the node.

For more information, see the *Informatica Big Data Suite 10.2.2 Installation and Configuration Guide*.

Configure Enterprise Data Lake

If you install the Enterprise Data Lake binaries on a node on which the Informatica services are installed, you can use the installer to configure the Enterprise Data Lake services on the node.

For more information, see the *Informatica Big Data Suite 10.2.2 Installation and Configuration Guide*.

Support Changes

This section describes the support changes in version 10.2.2.

Hive Engine

Effective in version 10.2.2, Informatica dropped support for the Hive mode of execution for jobs run in the Hadoop environment. You cannot configure or run jobs on the Hive engine.

Informatica continues to support the Blaze and Spark engines in the Hadoop environment, and it added support for the Databricks Spark engine in the Databricks environment.

You need to update all mappings and profiles configured to run on the Hive engine before you upgrade.

For more information, see the Informatica 10.2.2 upgrade guides.

Distribution Support

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

Big Data Management added support for the Databricks environment and supports the Databricks distribution version 5.1.

The following table lists the supported Hadoop distribution versions for Informatica 10.2.2 big data products:

Product	EMR	HDI	CDH	HDP	MapR
Big Data Management	5.16	3.6.x	5.15 5.16	2.6.x	6.0.x MEP 5.0
Enterprise Data Catalog	N/A	3.6.x **Note:** HDI is supported only for WASB non-Kerberos.	5.15 5.16 **Note:** You can use OpenJDK 1.8.0 only on Enterprise Data Catalog deployed on a CDH 5.16 Hadoop distribution.	2.6.x	N/A
Enterprise Data Preparation	5.16	3.6.x **Note:** HDI is supported only for ADLS and WASB non-Kerberos.	5.15 5.16	2.6.x	6.0.x MEP 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.

Support Changes for Big Data Management

This section describes the support changes to Big Data Management in version 10.2.2.

Python Transformation

Effective in version 10.2.2, support for binary ports in the Python transformation is deferred. Support will be reinstated in a future release.

Support Changes for Big Data Streaming

This section describes the changes to Big Data Streaming in version 10.2.2.

Azure Event Hubs

Effective in version 10.2.2, upgraded streaming mappings become invalid. You must re-create the physical data objects to run the mappings in Spark engine that uses Spark Structured Streaming. After

you re-create the physical data objects, the following properties are not available for Azure Event Hubs data objects:

- Consumer Properties
- Partition Count

For more information, see the *Informatica 10.2.2 Upgrade Guide*.

Data object types

Effective in version 10.2.2, support for some data object types is deferred. Support will be reinstated in a future release.

The following table describes the deferred support for data object types in version 10.2.2:

Object Type	Object
Source	JMS MapR Streams
Target	MapR Streams
Transformation	Data Masking Joiner (master outer, detail outer, and full outer join types) Rank Sorter

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

Universal Connectivity Framework in Enterprise Data Catalog

Effective in version 10.2.2, the Universal Connectivity Framework is deprecated in Enterprise Data Catalog.

For more information, see the *Statement of Support for the Usage of Universal Connectivity Framework (UCF) with Informatica Enterprise Data Catalog*.

Release Tasks

This section describes release tasks in version 10.2.2. Release tasks are tasks that you must perform after you upgrade to version 10.2.2.

Big Data Management

This section describes release tasks for Big Data Management in version 10.2.2.

Decimal Data Types

If you upgrade to version 10.2.2, mappings that are enabled for high-precision mode and run on the Spark engine must use a scale argument for the TO_DECIMAL and TO_DECIMAL38 functions. If the functions do not have a scale argument, the mappings will fail.

For example, if a pre-upgraded mapping uses high-precision mode and contains the expression `TO_DECIMAL(3)`, you must specify a scale argument before you can run the upgraded mapping on the Spark engine. When the expression has a scale argument, the expression might be `TO_DECIMAL(3,2)`.

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

Mass Ingestion

Effective in version 10.2.2, you can use the Mass Ingestion tool to ingest data using an incremental load.

If you upgrade to version 10.2.2, mass ingestion specifications are upgraded to have incremental load disabled. Before you can run incremental loads on existing specifications, complete the following tasks:

1. Edit the specification.
2. On the **Definition** page, select **Enable Incremental Load**.
3. On the **Source** and **Target** pages, configure the incremental load options.
4. Save the specification.
5. Redeploy the specification to the Data Integration Service.

Note: The redeployed mass ingestion specification runs on the Spark engine.

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

Python Transformation

If you upgrade to version 10.2.2, the Python transformation can process data more efficiently in Big Data Management.

To experience the improvements in performance, configure the following Spark advanced properties in the Hadoop connection:

infaspark.pythontx.exec

Required to run a Python transformation on the Spark engine for Data Engineering Integration. The location of the Python executable binary on the worker nodes in the Hadoop cluster.

For example, set to:

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

If you use the installation of Python on the Data Integration Service machine, set the value to the Python executable binary in the Informatica installation directory on the Data Integration Service machine.

For example, set to:

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

infaspark.pythontx.executorEnv.PYTHONHOME

Required to run a Python transformation on the Spark engine for Data Engineering Integration and Data Engineering Streaming. The location of the Python installation directory on the worker nodes in the Hadoop cluster.

For example, set to:

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

If you use the installation of Python on the Data Integration Service machine, use the location of the Python installation directory on the Data Integration Service machine.

For example, set to:

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

After you configure the advanced properties, the Spark engine does not use Jep to run Python code in the Python transformation.

For information about installing Python, see the *Informatica Big Data Management 10.2.2 Integration Guide*.

Big Data Streaming

This section describes release tasks for Big Data Streaming in version 10.2.2.

Kafka Target

Effective in version 10.2.2, the data type of the key header port in the Kafka target is binary. Previously, the data type of the key header port was string.

After you upgrade, to run an existing streaming mapping, you must re-create the data object, and update the streaming mapping with the newly created data object.

For more information about re-creating the data object, see the *Big Data Management 10.2.2 Integration Guide*.

Kafka Connection Properties

After you upgrade, for a Kafka connection, configure the Kafka messaging broker version to 0.10.1.x-2.0.0.

PowerExchange Adapters for Informatica

This section describes releases tasks for Informatica adapters in version 10.2.2.

PowerExchange for HBase

Effective in version 10.2.2, you must run a mapping on the Spark engine to look up data in an HBase resource.

If you previously configured a mapping to run in the native environment to look up data in an HBase resource, you must update the execution engine to Spark after you upgrade to version 10.2.2. Otherwise, the mapping fails.

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

PowerExchange for Microsoft Azure SQL Data Warehouse

After you upgrade from a previous release to version 10.2.2, the existing mappings that contain the following data types fail on the Spark engine at run time:

- Binary
- Varbinary
- Datetime2
- Datetimeoffset

To run the existing mappings successfully, you must map these data types to the string data type or re-import the object.

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

For more information, see the *Informatica PowerExchange for PowerExchange for Microsoft Azure SQL Data Warehouse 10.2.2 User Guide*.

New Products

PowerExchange Adapters for Informatica

This section describes new Informatica adapters in version 10.2.2.

PowerExchange for Cassandra JDBC

Effective in version 10.2.2, you can create a Cassandra connection to connect to a Cassandra keyspace and include Cassandra tables in a data object. You can use the Cassandra connection to read data from Cassandra sources and write data to Cassandra targets. You can validate and run mappings on the Spark engine in the Hadoop environment.

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

PowerExchange for Google Cloud Spanner

Effective in version 10.2.2, you can create a Google Cloud Spanner connection to connect to a Google Cloud Spanner instance and include Google Cloud Spanner tables in a data object. You can use the Google Cloud Spanner connection to read data from Google Cloud Spanner sources and write data to Google Cloud Spanner targets. You can validate and run Google Cloud Spanner mappings in the native environment or on the Spark engine in the Hadoop environment.

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

PowerExchange for Tableau V3

Effective in version 10.2.2, you can use the Tableau V3 connection to read data from multiple sources, generate a Tableau `.hyper` output file, and write the data to Tableau. You can validate and run mappings in the native environment.

For more information, see the *Informatica PowerExchange for Tableau V3 User Guide*.

CHAPTER 27

10.2.2 What's New

This chapter includes the following topics:

- Application Services, 225
- Big Data Management, 226
- Big Data Streaming , 230
- Command Line Programs, 232
- Enterprise Data Catalog, 236
- Enterprise Data Lake, 239
- Informatica Developer, 244
- Informatica Mappings, 244
- Informatica Transformations, 246
- PowerExchange Adapters for Informatica, 249

Application Services

This section describes new application service features in version 10.2.2.

Mass Ingestion Service

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

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

Metadata Access Service

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

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

REST Operations Hub Service

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

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

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

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

Big Data Management

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

Azure Databricks Integration

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

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

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

Sources and Targets

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

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

Transformations

You can add the following transformations to a Databricks mapping:

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

Router

Sorter

Union

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

Data Types

The following data types are supported:

Array

Bigint

Date/time

Decimal

Double

Integer

Map

Struct

Text

String

Mappings

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

Workflows

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

For more information, refer to the following guides:

Big Data Management 10.2.2 Integration Guide

Big Data Management 10.2.2 Administrator Guide

Big Data Management 10.2.2 User Guide

Data Preview on the Spark Engine

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

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

- Array
- Struct
- Map

Data preview is available for technical preview. Technical preview functionality is supported for evaluation purposes but is unwarranted and is not supported in production environments or any environment that you plan to push to production. Informatica intends to include the preview functionality in an upcoming release

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

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

Hierarchical Data

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

Dynamic Complex Ports

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

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

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

High Availability

This section describes new high availability features in version 10.2.2.

Big Data Job Recovery

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

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

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

Distributed Data Integration Service Queues

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

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

Intelligent Structure Model

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

Aliases in XML Files

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

Data Types

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

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

Field Names

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

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

Processing Large XML Files

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

Data Drift

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

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

Mass Ingestion

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

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

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

Monitoring

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

Spark Monitoring

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

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

Security

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

Enterprise Security Package

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

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

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

Targets

This section describes new features for targets in version 10.2.2.

HDFS Flat File Targets

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

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

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

Big Data Streaming

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

Azure Event Hubs Data Objects

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

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

Cross-account IAM Role in Amazon Kinesis Connection

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

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

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

Intelligent Structure Model

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

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

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

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

Header Ports for Big Data Streaming Data Objects

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

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

AWS Credential Profile in Amazon Kinesis Connection

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

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

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

Spark Structured Streaming

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

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

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

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

Window Transformation

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

Watermark Delay

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

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

Window Port

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

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

Command Line Programs

This section describes new commands in version 10.2.2.

infacmd dis Commands

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

Update Service Options

The Update Service Options command updates Data Integration Service properties.

The following table describes new infacmd dis updateServiceOptions command options:

New Option	Description
-RecycleMode -rm	Optional. Recycle mode restarts the service and applies the latest service and service process properties. Select Abort or Complete. - Complete. Stops all applications and cancels all jobs within each application. Waits for all jobs to cancel before disabling the service. - Abort. Stops all applications and tries to cancel all jobs before aborting them and disabling the service. Default: Complete

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

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

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

infacmd ihs Commands

The following table describes new infacmd ihs command options:

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

The following table describes new infacmd ihs commands:

Command	Description
cleanCluster	Cleans the Informatica Cluster Service.

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

infacmd ipc Commands

The following table describes obsolete infacmd ipc commands:

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

infacmd ldm Commands

The following table describes new infacmd ldm command options:

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

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

The following table describes new infacmd ldm commands:

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

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

infacmd mi Commands

The following table describe changes to infacmd mi commands:

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

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

infacmd ms Commands

The following table describes new infacmd ms commands:

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

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

infacmd oie Commands

Effective in version 10.2.2, the oie plugin is deprecated and support for the plugin will be removed in a future release. The infacmd oie commands have migrated to the tools plugin. For details, see "infacmd tools Commands" on page 235.

infacmd tools Commands

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

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

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

The following table describes new infacmd tools commands:

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

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

infasetup Commands

The following table describes changed infasetup commands:

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

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

Enterprise Data Catalog

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

Automatically Assign Business Title to a Column

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

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

User Collaboration on Assets

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

You can collaborate on assets in the following ways:

Follow assets

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

Rate and review asset

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

Asset queries

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

Certify asset

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

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

Create Enterprise Data Catalog Application Services Using the Installer

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

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

Custom Metadata Validation Utility

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

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

Change Notifications

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

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

Business Glossary Assignment Report

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

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

Operating System Profiles

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

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

REST APIs

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

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

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

Source Metadata and Data Profile Filter

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

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

Scanner Utility

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

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

Resource Types

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

Google BigQuery

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

- Project
- Dataset

- Table
- View

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

Workday

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

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

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

Enterprise Data Lake

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

Apply Active Rules

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

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

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

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

Delete Duplicate Rows

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

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

Cluster and Categorize Column Data

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

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

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

CLAIRE-based Recommendations

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

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

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

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

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

Conditional Aggregation

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

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

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

Data Masking

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

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

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

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

Localization

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

Partitioned Sources and Targets

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

Add Comments to Recipe Steps

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

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

Save a Recipe as a Mapping

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

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

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

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

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

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

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

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

Statistical Functions

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

- AVG
- AVGIF
- COUNT
- COUNTIF
- COUNTDISTINCT

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

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

Date and Time Functions

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

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

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

Math Functions

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

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

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

Text Functions

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

- ENDSWITH
- ENDSWITH_IGNORE_CASE
- FIND_IGNORE_CASE
- FIND_REGEX
- FIRST_CHARACTER_TO_NUMBER
- NUMBER_TO_CHARACTER
- PROPER_CASE
- REMOVE_NON_ALPHANUMERIC_CHARACTERS
- STARTSWITH
- STARTSWITH_IGNORE_CASE
- SUBSTITUTE_REGEX
- TRIM_ALL
- Convert Date to Text
- Convert Number to Text
- Convert Text to Date
- Convert Text to Number

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

Window Functions

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

You can use window functions to perform the following tasks:

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

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

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

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

Purge Audit Events

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

Spark Execution Engine

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

Informatica Developer

This section describes new Developer tool features in version 10.2.2.

Applications

Effective in version 10.2.2, you can create incremental applications. An incremental application is an application that you can update by deploying an application patch to update a subset of application objects. The Data Integration Service updates the objects in the patch while other application objects continue running.

If you upgrade to version 10.2.2, existing applications are labeled "full applications." You can continue to create full applications in version 10.2.2, but you cannot convert a full application to an incremental application.

For more information, see the "Application Deployment" and the "Application Patch Deployment" chapters in the *Informatica 10.2.2 Developer Tool Guide*.

Informatica Mappings

This section describes new Informatica mapping features in version 10.2.2.

Data Types

Effective in version 10.2.2, you can enable high-precision mode in batch mappings that run on the Spark engine. The Spark engine can process decimal values with up to 38 digits of precision.

For more information, see the *Informatica Big Data Management 10.2.2 User Guide*.

Mapping Outputs

Effective in version 10.2.2, you can use mapping outputs in batch mappings that run as Mapping tasks in workflows on the Spark engine. You can persist the mapping outputs in the Model repository or bind the mapping outputs to workflow variables.

For more information, see the "Mapping Outputs" chapter in the *Informatica 10.2.2 Developer Mapping Guide* and the "Mapping Task" chapter in the *Informatica 10.2.2 Developer Workflow Guide*.

Mapping Parameters

Effective in version 10.2.2, you can assign expression parameters to port expressions in Aggregator, Expression, and Rank transformations that run in the native and non-native environments.

For more information, see the "Where to Assign Parameters" and "Dynamic Mappings" chapters in the *Informatica 10.2.2 Developer Mapping Guide*.

Optimizer Levels

Effective in version 10.2.2, you can configure the Auto optimizer level for mappings and mapping tasks. With the Auto optimization level, the Data Integration Service applies optimizations based on the execution mode and mapping contents.

The optimizer level default for new mappings is Auto.

When you upgrade to version 10.2.2, optimizer levels configured in mappings remain the same. To use the Auto optimizer level with upgraded mappings, you must manually change the optimizer level.

For more information, see the "Optimizer Levels" chapter in the *Informatica 10.2.2 Developer Mapping Guide*.

Sqoop

Effective in version 10.2.2, you can use the following new Sqoop features:

Incremental data extraction support

You can configure a Sqoop mapping to perform incremental data extraction based on an ID or timestamp. With incremental data extraction, Sqoop extracts only the data that changed since the last data extraction. Incremental data extraction increases the mapping performance.

Vertica connectivity support

You can configure Sqoop to read data from a Vertica source or write data to a Vertica target.

Spark engine optimization for Sqoop pass-through mappings

When you run a pass-through mapping with a Sqoop source on the Spark engine, the Data Integration Service optimizes mapping performance in the following scenarios:

- You write data to a Hive target that was created with a custom DDL query.
- You write data to an existing Hive target that is either partitioned with a custom DDL query or partitioned and bucketed with a custom DDL query.
- You write data to an existing Hive target that is both partitioned and bucketed.

--infaownername argument support

You can configure the --infaownername argument to indicate whether Sqoop must honor the owner name for a data object.

For more information, see the *Informatica Big Data Management 10.2.2 User Guide*.

Informatica Transformations

This section describes new features in Informatica transformations in version 10.2.2.

Address Validator Transformation

This section describes new Address Validator transformation features.

The Address Validator transformation contains additional address functionality for the following countries:

All Countries

Effective in version 10.2.2, the Address Validator transformation supports single-line address verification in every country for which Informatica provides reference address data.

In earlier versions, the transformation supported single-line address verification for 26 countries.

To verify a single-line address, enter the address in the Complete Address port. If the address identifies a country for which the default preferred script is not a Latin or Western script, use the default Preferred Script property on the transformation with the address.

Australia

Effective in version 10.2.2, you can configure the Address Validator transformation to add address enrichments to Australia addresses. You can use the enrichments to discover the geographic sectors and regions to which the Australia Bureau of Statistics assigns the addresses. The sectors and regions include census collection districts, mesh blocks, and statistical areas.

The transformation uses the following ports to deliver the enrichments:

- Census Collection District Code 2006
- Geocoded National Address File Identifier
- Greater Capital City Statistical Area 5-Digit
- Greater Capital City Statistical Area Name
- Level One Statistical Area 11-Digit
- Level One Statistical Area 7-Digit
- Level Two Statistical Area 9-Digit
- Level Two Statistical Area 5-Digit
- Level Two Statistical Area Name
- Level Three Statistical Area 5-Digit
- Level Three Statistical Area Name
- Level Four Statistical Area 3-Digit
- Level Four Statistical Area Name
- Mesh Block 11-Digit 2011
- Mesh Block 11-Digit 2016
- State or Territory Code
- State or Territory Name
- Supplementary AU Status

Find the ports in the AU Supplementary port group.

Bolivia

Effective in version 10.2.2, the Address Validator transformation improves the parsing and validation of Bolivia addresses. Additionally, Informatica updates the reference data for Bolivia.

The transformation also includes the following improvements for Bolivia:

- Address validation to street level.
- Geocoordinates at street mid-point level for addresses in major cities.

Canada

Informatica introduces the following features and enhancements for Canada:

Support for the Global Preferred Descriptor property in Canada Addresses

Effective in version 10.2.2, you can configure the Address Validator transformation to return the short or long form of an element descriptor.

The transformation can return the short or long form of the following descriptors:

- Street descriptors
- Directional values
- Building descriptors
- Sub-building descriptors

To specify the output format for the descriptors, configure the Global Preferred Descriptor property on the transformation. The property applies to English-language and French-language descriptors. By default, the transformation returns the descriptor in the format that the reference data specifies. If you select the PRESERVE INPUT option on the property, the Preferred Language property takes precedence over the Global Preferred Descriptor property.

Support for CH and CHAMBER as Sub-Building Descriptors

Effective in version 10.2.2, the Address Validator transformation recognizes CH and CHAMBER as sub-building descriptors in Canada addresses.

Colombia

Effective in version 10.2.2, the Address Validator transformation improves the processing of street data in Colombia addresses. Additionally, Informatica updates the reference data for Colombia.

France

Effective in version 10.2.2, Informatica introduces the following improvements for France addresses:

- Informatica improves the supplementary reference data for France.
- The Address Validator transformation assigns addresses to IRIS units in France with greater accuracy. The transformation uses the house number in the address to verify the IRIS unit to which the address belongs. The use of house numbers can improve the assignment accuracy when the address lies close to the border between different units.

Israel

Effective in version 10.2.2, Informatica introduces the following features and enhancements for Israel:

Multilanguage Support for Israel Addresses

You can configure the Address Validator transformation to return an Israel address in the English language or the Hebrew language.

Use the Preferred Language property to select the preferred language for the addresses that the transformation returns.

The default language for Israel addresses is Hebrew. To return address information in Hebrew, set the Preferred Language property to DATABASE or ALTERNATIVE_1. To return the address information in English, set the property to ENGLISH or ALTERNATIVE_2.

Support for Multiple Character Sets for Israel Addresses

The Address Validator transformation can read and write Israel addresses in Hebrew and Latin character sets.

Use the Preferred Script property to select the preferred character set for the address data.

The default character set for Israel addresses is Hebrew. When you set the Preferred Script property to Latin or Latin-1, the transformation transliterates Hebrew address data into Latin characters.

Peru

Effective in version 10.2.2, the Address Validator transformation validates a Peru address to house number level. Additionally, Informatica updates the reference data for Peru.

Sweden

Effective in version 10.2.2, the Address Validator transformation improves the verification of street names in Sweden addresses.

The transformation improves the verification of street names in the following ways:

- The transformation can recognize a street name that ends in the character G as an alias of the same name with the final characters GATAN.
- The transformation can recognize a street name that ends in the character V as an alias of the same name with the final characters VÄGEN.
- The Address Validator transformation can recognize and correct a street name with an incorrect descriptor when either the long form or the short form of the descriptor is used.

  For example, The transformation can correct RUNIUSV or RUNIUSVÄGEN to RUNIUSGATAN in the following address:

  RUNIUSGATAN 7

  SE-112 55 STOCKHOLM

United States

Effective in version 10.2 HotFix 2, you can configure the Address Validator transformation to identify United States addresses that do not receive mail on one or more days of the week.

To identify the addresses, use the *Non-Delivery Days* port. The port contains a seven-digit string that represents the days of the week from Sunday through Saturday. Each position in the string represents a different day.

The Address Validator transformation returns the first letter of a weekday in the corresponding position on the port if the address does not receive mail on that day. The transformation returns a dash symbol in the corresponding position for other days of the week.

For example, a value of `S----FS` on the Non-Delivery Days port indicates that an address does not receive mail on Sunday, Friday, and Saturday.

Find the Non-Delivery Days port in the US Specific port group in the Basic model. To receive data on the Non-Delivery Days port, run the Address Validator transformation in certified mode. The transformation reads the port values from the `USA5C129.MD` and `USA5C130.MD` database files.

For comprehensive information about the features and operations of the address verification software engine in version 10.2.2, see the *Informatica Address Verification 5.14.0 Developer Guide*.

Update Strategy Transformation

Effective in version 10.2.2, you can use an Update Strategy transformation in a mapping that runs on the Spark engine to update relational targets.

Previously, you could use an Update Strategy transformation in a mapping that runs on the Spark engine only to update Hive targets.

For more information, see the Update Strategy transformation chapter in the *Developer Transformation Guide*.

PowerExchange Adapters for Informatica

This section describes new Informatica adapter features in version 10.2.2.

PowerExchange for Amazon Redshift

Effective in version 10.2.2, PowerExchange for Amazon Redshift includes the following features:

- You can read data from or write data to the following regions:
  - China(Ningxia)
  - EU(Paris)
- You can use Amazon Redshift objects as dynamic sources and target in a mapping.
- You can use octal values of printable and non-printable ASCII characters as a DELIMITER or QUOTE.
- You can enter pre-SQL and post-SQL commands to run queries for source and target objects in a mapping.
- You can define an SQL query for read data objects in a mapping to override the default query. You can enter an SQL statement supported by the Amazon Redshift database.
- You can specify the maximum size of an Amazon S3 object in bytes when you download large Amazon S3 objects in multiple parts.
- You can read unique values when you read data from an Amazon Redshift source.
- When you upload an object to Amazon S3, you can specify the minimum size of the object and the number of threads to upload the objects in parallel as a set of independent parts.
- You can choose to retain an existing target table, replace a target table at runtime, or create a new target table if the table does not exist in the target.
- You can configure the Update Strategy transformations for an Amazon Redshift target in the native environment.
- When you write data to Amazon Redshift, you can override the Amazon Redshift target table schema and the table name during run time.
- When the connection type is ODBC, the Data Integration Service can push transformation logic to Amazon Redshift sources and targets using source-side and full pushdown optimization.
- You can use Server-Side Encryption with AWS KMS (AWS Key Management Service) on Amazon EMR version 5.16 and Cloudera CDH version 5.15 and 5.16.
- PowerExchange for Amazon Redshift supports AWS SDK for Java version 1.11.354.

For more information, see the *Informatica PowerExchange for Amazon Redshift 10.2.2 User Guide*.

PowerExchange for Amazon S3

Effective in version 10.2.2, PowerExchange for Amazon S3 includes the following features:

- You can read data from or write data to the following regions:
  - China(Ningxia)
  - EU(Paris)
  - AWS GovCloud (US)
- You can use Amazon S3 objects as dynamic sources and target in a mapping.
- When you run a mapping in the native environment or on the Spark engine to read data from an Avro, flat, JSON, ORC, or Parquet file, you can use wildcard characters to specify the source directory name or the source file name.
- You can add a single or multiple tags to the objects stored on the Amazon S3 bucket to categorize the objects. Each tag contains a key value pair. You can either enter the key value pairs or specify the absolute file path that contains the key value pairs.
- You can specify the maximum threshold size to download an Amazon S3 object in multiple parts.
- When you upload an object to Amazon S3, you can specify the minimum size of the object and the number of threads to upload the objects in parallel as a set of independent parts.
- When you create a data object read or write operation, you can read data present in the FileName port that contains the endpoint name and source path of the file.
- You can add new columns or modify the columns in the Port tab directly when you create a data object read or write operation.
- You can copy the columns of the source transformations, target transformations, or any other transformations from the Port tab and paste the columns in the data object read or write operation directly when you create a mapping to read or write an Avro, JSON, ORC, or Parquet file.
- You can update the Amazon S3 file format without losing the column metadata in the Schema field of the column projection properties even after you configure the column projection properties for another Amazon S3 file format.
- You can use Server-Side Encryption with AWS KMS (AWS Key Management Service) on Amazon EMR version 5.16 and Cloudera CDH version 5.15 and 5.16.
- PowerExchange for Amazon S3 supports AWS SDK for Java version 1.11.354.

For more information, see the *Informatica PowerExchange for Amazon S3 10.2.2 User Guide*.

PowerExchange for Google BigQuery

Effective in version 10.2.2, you can create a Google BigQuery target using the right-click **Create Target** option.

For more information, see the *Informatica PowerExchange for Google BigQuery 10.2.2 User Guide*.

PowerExchange for HBase

Effective in version 10.2.2, PowerExchange for HBase includes the following new features:

- When you create an HBase data object, you can select an operating system profile to increase security and to isolate the design-time user environment when you import and preview metadata from a Hadoop cluster.

  Note: You can choose an operating system profile if the Metadata Access Service is configured to use operating system profiles. The Metadata Access Service imports the metadata with the default operating system profile assigned to the user. You can change the operating system profile from the list of available operating system profiles.
- You can use the HBase objects as dynamic sources and targets in a mapping.
- You can run a mapping on the Spark engine to look up data in an HBase resource.

For more information, see the *Informatica PowerExchange for HBase 10.2.2 User Guide*.

PowerExchange for HDFS

Effective in version 10.2.2, PowerExchange for HDFS includes the following new features:

- When you create a complex file data object, you can select an operating system profile to increase security and to isolate the design-time user environment when you import and preview metadata from a Hadoop cluster.

  Note: You can choose an operating system profile if the Metadata Access Service is configured to use operating system profiles. The Metadata Access Service imports the metadata with the default operating system profile assigned to the user. You can change the operating system profile from the list of available operating system profiles.
- When you run a mapping in the native environment or on the Spark engine to read data from a complex file data object, you can use wildcard characters to specify the source directory name or the source file name.
  You can use the following wildcard characters:

  ? (Question mark)

  The question mark character (?) allows one occurrence of any character.

  *** (Asterisk)**

  The asterisk mark character (*) allows zero or more than one occurrence of any character.
- You can use complex file objects as dynamic sources and targets in a mapping.
- You can use complex file objects to read data from and write data to a complex file system.
- When you run a mapping in the native environment or on the Spark engine to write data to a complex file data object, you can overwrite target data, the Data Integration Service deletes the target data before writing new data.
- When you create a data object read or write operation, you can read the data present in the FileName port that contains the endpoint name and source path of the file.
- You can now view the data object operations immediately after you create the data object read or write operation.
- You can add new columns or modify the columns, when you create a data object read or write operation.
- You can copy the columns of the source transformations, target transformations, or any other transformations and paste the columns in the data object read or write operation directly when you read or write to an Avro, JSON, ORC, or Parquet file.

For more information, see the *Informatica PowerExchange for HDFS 10.2.2 User Guide*.

PowerExchange for Hive

Effective in version 10.2.2, PowerExchange for Hive includes the following new features:

- You can configure the following target schema strategy options for a Hive target:
  - RETAIN - Retain existing target schema
  - CREATE - Create or replace table at run time
  - APPLYNEWCOLUMNS - Alter table and apply new columns only
  - APPLYNEWSCHEMA - Alter table and apply new schema
  - FAIL - fail mapping if target schema is different
  - Assign Parameter
- You can truncate an internal or external partitioned Hive target before loading data. This option is applicable when you run the mapping in the Hadoop environment.
- You can create a read or write transformation for Hive in native mode to read data from Hive source or write data to Hive target.
- When you write data to a Hive target, you can configure the following properties in a Hive connection:
  - Hive Staging Directory on HDFS. Represents the HDFS directory for Hive staging tables. This option is applicable and required when you write data to a Hive target in the native environment.
  - Hive Staging Database Name. Represents the namespace for Hive staging tables. This option is applicable when you run a mapping in the native environment to write data to a Hive target. If you run the mapping on the Blaze or Spark engine, you do not need to configure the Hive staging database name in the Hive connection. The Data Integration Service uses the value that you configure in the Hadoop connection.

For more information, see the *Informatica PowerExchange for Hive 10.2.2 User Guide*.

PowerExchange for MapR-DB

Effective in version 10.2.2, when you create an HBase data object for MapR-DB, you can select an operating system profile to increase security and to isolate the design-time user environment when you import and preview metadata from a Hadoop cluster.
Note: You can choose an operating system profile if the Metadata Access Service is configured to use operating system profiles. The Metadata Access Service imports the metadata with the default operating system profile assigned to the user. You can change the operating system profile from the list of available operating system profiles.

For more information, see the *Informatica PowerExchange for MapR-DB 10.2.2 User Guide*.

PowerExchange for Microsoft Azure Blob Storage

Effective in version 10.2.2, PowerExchange for Microsoft Azure Blob Storage includes the following functionality:

- You can run mappings in the Azure Databricks environment.
- You can configure the US government Microsoft Azure end-points.

- You can compress data in the following formats when you read data from or write data to Microsoft Azure Blob Storage:
  - None
  - Deflate
  - Gzip
  - Bzip2
  - Lzo
  - Snappy
- You can use Microsoft Azure Blob Storage objects as dynamic sources and targets in a mapping.
- You can read the name of the file from which the Data Integration Service reads the data at run-time in the native environment.
- You can configure the relative path in **Blob Container Override** in the advanced source and target properties.

For more information, see the *Informatica PowerExchange for Microsoft Azure Blob Storage 10.2.2 User Guide*.

PowerExchange for Microsoft Azure Cosmos DB SQL API

Effective in version 10.2.2, PowerExchange for Microsoft Azure Cosmos DB SQL API includes the following functionality:

- You can run mappings in the Azure Databricks environment. Databricks support for PowerExchange for Microsoft Azure Cosmos DB SQL API is available for technical preview. Technical preview functionality is supported but is unwarranted and is not production-ready. Informatica recommends that you use these features in non-production environments only.

For more information, see the *Informatica PowerExchange for Microsoft Azure Cosmos DB SQL API 10.2.2 User Guide*.

PowerExchange for Microsoft Azure Data Lake Store

Effective in version 10.2.2, PowerExchange for Microsoft Azure Data Lake Store includes the following functionality:

- You can run mappings in the Azure Databricks environment.
- You can use complex data types, such as array, struct, and map, in mappings that run in the Hadoop environment. With complex data types, the respective engine directly reads, processes, and writes hierarchical data in Avro, JSON, and Parquet complex files. For an intelligent structure source, you can configure only the read operation.
- You can create mappings to read and write Avro and Parquet files that contain only primitive data types in the native environment.
- You can select a directory as a source in a mapping to read multiple files from the directory.
- You can use Microsoft Azure Data Lake Store objects as dynamic sources and targets in a mapping.
- You can create a Microsoft Azure Data Lake Store target using the Create Target option.

For more information, see the *Informatica PowerExchange for Microsoft Azure Data Lake Store 10.2.2 User Guide*.

PowerExchange for Microsoft Azure SQL Data Warehouse

Effective in version 10.2.2, PowerExchange for Microsoft Azure SQL Data Warehouse includes the following functionality:

- You can run mappings in the Azure Databricks environment.
- You can configure the US government Microsoft Azure end-points in mappings that run in the native environment and on the Spark engine.
- You can generate error files in the Microsoft Azure Blob Storage container. The error files contain rejected rows and the cause for the rejected rows.
- You can define the batch size in advance target properties in the native environment.
- You can configure full pushdown optimization to push transformation logic to source databases and target databases. Use pushdown optimization to improve task performance by using the database resources.
- You can use Microsoft Azure SQL Data Warehouse objects as dynamic sources and targets in a mapping.

The full pushdown optimization and the dynamic mappings functionality for PowerExchange for Microsoft Azure SQL Data Warehouse is available for technical preview. Technical preview functionality is supported but is unwarranted and is not production-ready. Informatica recommends that you use these features in non-production environments only.

For more information, see the *Informatica PowerExchange for Microsoft Azure SQL Data Warehouse 10.2.2 User Guide*.

PowerExchange for Salesforce

Effective in version 10.2.2, PowerExchange for Salesforce includes the following new features:

- You can use version 43.0 and 44.0 of Salesforce API to create a Salesforce connection and access Salesforce objects.
- You can configure OAuth for Salesforce connections.
- You can configure the native expression filter for the source data object operation.
- You can parameterize the following read operation properties for a Salesforce data object:
  - SOQL Filter Condition
  - CDC Start Timestamp
  - CDC End Timestamp
  - PK Chunking Size
  - PK Chunking startRow ID

  You can parameterize the following write operation properties for a Salesforce data object:
  - Set prefix for BULK success and error files
  - SFDC Success File Directory
  - Set the location of the BULK error files

For more information, see the *Informatica PowerExchange for Salesforce10.2.2 User Guide*.

PowerExchange for Snowflake

Effective in version 10.2.2, PowerExchange for Snowflake includes the following new features:

- You can configure Okta SSO authentication by specifying the authentication details in the JDBC URL parameters of the Snowflake connection.
- You can configure an SQL override to override the default SQL query used to extract data from the Snowflake source. Specify the SQL override in the Snowflake data object read operation properties.
- You can choose to compress the files before writing to Snowflake tables and optimize the write performance. In the advanced properties. You can set the compression parameter to On or Off in the **Additional Write Runtime Parameters** field in the Snowflake data object write operation advanced properties.
- The Data Integration Service uses the Snowflake Spark Connector APIs to run Snowflake mappings on the Spark engine.
- You can read data from and write data to Snowflake that is enabled for staging data in Azure or Amazon.

For more information, see the *Informatica PowerExchange for Snowflake10.2.2 User Guide*.

PowerExchange for Teradata Parallel Transporter API

Effective in version 10.2.2, PowerExchange for Teradata Parallel Transporter API includes the following functions in the advanced target properties:

- You can specify a replacement character to use in place of an unsupported Teradata unicode character in the Teradata database while loading data to targets.
- If you specified a character used in place of an unsupported character while loading data to Teradata targets, you can specify version 8.x - 13.x or 14.x and later for the target Teradata database. Use this attribute in conjunction with the Replacement Character attribute. The Data Integration Service ignores this attribute if you did not specify a replacement character while loading data to Teradata targets.
- When you write data to Teradata, you can override the Teradata target table schema and the table name during run time.

For more information, see the *Informatica PowerExchange for Teradata Parallel Transporter API 10.2.2 User Guide*.

CHAPTER 28

10.2.2 What's Changed

This chapter includes the following topics:

- Application Services, 256
- Big Data Management, 256
- Big Data Streaming, 259
- Enterprise Data Catalog, 260
- Enterprise Data Lake, 260
- Informatica Developer, 261
- Informatica Transformations, 261
- PowerExchange Adapters for Informatica, 262

Application Services

This section describes the changes to the application service features in version 10.2.2.

Metadata Access Service

Effective in version 10.2.2, you can use the same Metadata Access Service to import and preview metadata from multiple Hadoop distributions. Previously, you had to create multiple instances of Metadata Access Service for multiple Hadoop distributions.

Big Data Management

This section describes the changes to Big Data Management in version 10.2.2.

Hive Connection

Effective in version 10.2.2, the following Hive connection properties are renamed:

- The Observe Fine Grained SQL Authorization property is renamed Fine Grained Authorization.
- The User Name property is renamed LDAP username.

The following table describes the properties:

Property	Description
Fine Grained Authorization	When you select the option to observe fine grained authorization in a Hive source, the mapping observes the following: - Row and column level restrictions. Applies to Hadoop clusters where Sentry or Ranger security modes are enabled. - Data masking rules. Applies to masking rules set on columns containing sensitive data by Dynamic Data Masking. If you do not select the option, the Blaze and Spark engines ignore the restrictions and masking rules, and results include restricted or sensitive data.
LDAP username	LDAP user name of the user that the Data Integration Service impersonates to run mappings on a Hadoop cluster. The user name depends on the JDBC connection string that you specify in the Metadata Connection String or Data Access Connection String for the native environment. If the Hadoop cluster uses Kerberos authentication, the principal name for the JDBC connection string and the user name must be the same. Otherwise, the user name depends on the behavior of the JDBC driver. With Hive JDBC driver, you can specify a user name in many ways and the user name can become a part of the JDBC URL. If the Hadoop cluster does not use Kerberos authentication, the user name depends on the behavior of the JDBC driver. If you do not specify a user name, the Hadoop cluster authenticates jobs based on the following criteria: - The Hadoop cluster does not use Kerberos authentication. It authenticates jobs based on the operating system profile user name of the machine that runs the Data Integration Service. - The Hadoop cluster uses Kerberos authentication. It authenticates jobs based on the SPN of the Data Integration Service. LDAP username will be ignored.

For more information, see the *Informatica Big Data Management 10.2.2 User Guide*.

Mass Ingestion

Effective in version 10.2.2, deployed mass ingestion specifications run on the Spark engine. Upgraded mass ingestion specifications that were deployed prior to version 10.2.2 will continue to run on the Blaze and Spark engines until they are redeployed.

For more information, see the *Informatica Big Data Management 10.2.2 Mass Ingestion Guide*.

Spark Monitoring

Effective in version 10.2.2, the Spark monitoring is enabled by default.

Previously, Spark monitoring was disabled by default.

For more information about Spark monitoring, see the *Informatica Big Data Management 10.2.2 User Guide*.

Sqoop

Effective in version 10.2.2, the following changes apply to Sqoop:

- You can specify a file path in the Spark staging directory of the Hadoop connection to store temporary files for Sqoop jobs. When the Spark engine runs Sqoop jobs, the Data Integration Service creates a Sqoop staging directory within the Spark staging directory to store temporary files: `<Spark staging directory>/sqoop_staging`
  Previously, the Sqoop staging directory was hard-coded and the Data Integration Service used the following staging directory: `/tmp/sqoop_staging`

  For more information, see the *Informatica Big Data Management 10.2.2 User Guide*.
- Sqoop mappings on the Spark engine use the OpenJDK (AzulJDK) packaged with the Informatica installer. You no longer need to specify the **JDK Home Directory** property for the Data Integration Service. Previously, to run Sqoop mappings on the Spark engine, you installed the Java Development Kit (JDK) on the machine that runs the Data Integration Service. You then specified the location of the JDK installation directory in the **JDK Home Directory** property under the Data Integration Service execution options in Informatica Administrator.

Transformations in the Hadoop Environment

This section describes changes to transformations in the Hadoop environment in version 10.2.2.

Python Transformation

Effective in version 10.2.2, the Python transformation can process data more efficiently on the Spark engine compared to the Python transformation in version 10.2.1. Additionally, the Python transformation does not require you to install Jep, and you can use any version of Python to run the transformation.

Previously, the Python transformation supported only specific versions of Python that were compatible with Jep.

Note: The improvements are available only for Big Data Management.

For information about installing Python, see the *Informatica Big Data Management 10.2.2 Integration Guide*.

For more information about the Python transformation, see the "Python Transformation" chapter in the *Informatica 10.2.2 Developer Transformation Guide*.

Write Transformation

Effective in version 10.2.2, the **Create or Replace Target Tables** advanced property in a Write transformation for relational, Netezza, and Teradata data objects is renamed to **Target Schema Strategy**.

When you configure a Write transformation, you can choose from the following target schema strategy options for the target data object:

- RETAIN - Retain existing target schema. The Data Integration Service retains the existing target schema.
- CREATE - Create or replace table at run time. The Data Integration Service drops the target table at run time and replaces it with a table based on a target data object that you identify.
- Assign Parameter. Specify the **Target Schema Strategy** options as a parameter value.

Previously, you selected the **Create or Replace Target Tables** advanced property so that the Data Integration Service drops the target table at run time and replaces it with a table based on a target table that you identify. When you do not select the **Create or Replace Target Tables** advanced property, the Data Integration Service retains the existing schema for the target table.

In existing mappings where the **Create or Replace Target Tables** property was enabled, after the upgrade to version 10.2.2, by default, the **Target Schema Strategy** property shows enabled for the **CREATE - Create or replace table at run time** option. In mappings where the **Create or Replace Target Tables** option was not selected, after the upgrade, the **Target Schema Strategy** property is enabled for the **RETAIN - Retain existing target schema** option. After the upgrade, if the correct target schema strategy option is not selected, you must manually select the required option from the **Target Schema Strategy** list, and then run the mapping.

For more information about configuring the target schema strategy, see the "Write Transformation" chapter in the *Informatica Transformation Guide*, or the "Dynamic Mappings" chapter in the *Informatica Developer Mapping Guide*.

Big Data Streaming

This section describes changes to Big Data Streaming in version 10.2.2.

Big Data Streaming and Big Data Management Integration

Effective in version 10.2.2, to manage the target files in the target file system at run time, a temporary directory is created in the specified target file directory.

The temporary directory separates the target files to which the data is currently written and the target files that are closed after the rollover limit is reached.

Previously, all the target files were stored in the target file directory.

For more information, see the *Informatica Big Data Streaming 10.2.2 User Guide*.

Kafka Connection

Effective in version 10.2.2, Kafka broker maintains the configuration information for the Kafka messaging broker. Previously, Apache ZooKeeper maintained the configuration information for the Kafka messaging broker.

For more information, see the *Big Data Streaming 10.2.2 User Guide*.

Transformations

This section describes the changes to transformations in Big Data Streaming in version 10.2.2.

Aggregator Transformation

Effective in version 10.2.2, a streaming mapping must meet the following additional requirements if it contains an Aggregator transformation:

- A streaming mapping must have the Window transformation directly upstream from an Aggregator transformation. Previously, you could use an Aggregator transformation anywhere in the pipeline after the Window transformation.
- A streaming mapping can have a single Aggregator transformation. Previously, you could use multiple Aggregator transformations in a streaming mapping.

Joiner Transformation

Effective in version 10.2.2, a streaming mapping must meet the following additional requirements if it contains a Joiner transformation:

- A streaming mapping must have the Window transformation directly upstream from a Joiner transformation. Previously, you could use a Joiner transformation anywhere in the pipeline after a Window transformation.
- A streaming mapping can have a single Joiner transformation. Previously, you could use multiple Joiner transformations in a streaming mapping.
- A streaming mapping cannot contain an Aggregator transformation anywhere before a Joiner transformation in a streaming mapping. Previously, you could use an Aggregator transformation anywhere before a Joiner transformation in a streaming mapping.

Enterprise Data Catalog

This section describes the changes to Enterprise Data Catalog in version 10.2.2.

Java Development Kit Change

Effective in version 10.2.2, the Informatica installer packages OpenJDK (Azul JDK). Previously, the installer packaged the Oracle JDK. You can use the OpenJDK to deploy Enterprise Data Catalog on an embedded cluster.

To deploy Enterprise Data Catalog on an existing cluster, you must install JDK 1.8 on all the cluster nodes.

Enterprise Data Lake

This section describes the changes to Enterprise Data Lake in version 10.2.2.

MAX and MIN Functions

Effective in version 10.2.2, the behavior of the MAX and MIN aggregation functions is changed. To maintain the behavior of the MAX and MIN functions provided in version 10.2.1 or earlier in a worksheet, rename the functions to MAXINLIST and MININLIST in the worksheet recipe.

The following table describes the functions provided in version 10.2.2:

Function	Description
MAX (value)	Returns the maximum value among all rows in the worksheet, based on the columns included in the specified expression.
MIN (value)	Returns the minimum value among all rows in the worksheet, based on the columns included in the specified expression.

Function	Description
MAXINLIST (value, [value],...)	Returns the largest number or latest date in the specified list of expressions.
MININLIST (value, [value],...)	Returns the smallest number or earliest date in the specified list of expressions.

Informatica Developer

This section describes changes to Informatica Developer in version 10.2.2.

Informatica Developer Name Change

In version 10.2.2, Informatica Developer is now called Big Data Developer.

For big data releases, the tool is renamed to Big Data Developer. A big data release includes products such as Big Data Management and Big Data Quality.

For traditional releases, the tool name remains Informatica Developer. A traditional release includes products such as PowerCenter and Data Quality.

Informatica Transformations

This section describes changes to Informatica transformations in version 10.2.2.

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 10.2.2, the Address Validator transformation incorporates features from version 5.14.0 of the Informatica Address Verification software engine.

Previously, the transformation used version 5.12.0 of the Informatica Address Verification software engine.

Japan

Effective in version 10.2.2, Informatica improves the parsing and validation of Japan addresses based on customer feedback.

For example, in version 10.2.2, Informatica rejects a Japan address when the postal code is absent from the address or the postal code and the locality information do not match.

Previously, Informatica tried to correct the address.

Spain

Effective in version 10.2.2, Informatica improves the parsing and validation of Spain addresses.

For example, in version 10.2.2, the Address Validator transformation rejects a Spain address when the street information needs multiple corrections to create a match with the reference data.

Previously, the transformation performed multiple corrections to the street data, which might lead to an optimistic assessment of the input address accuracy.

Similarly, in version 10.2.2, if an address matches multiple candidates in the reference data, the Address Validator transformation returns a I3 result for the address in batch mode.

Previously, the transformation might try to correct the input address.

Parcel Centroid and Rooftop Geocodes

Effective October 2018, Informatica no longer supports the reference data files that contain Parcel Centroid and Rooftop geocode data.

For comprehensive information about the updates to the Informatica Address Verification software engine, see the *Informatica Address Verification 5.14.0 Release Guide*.

Write Transformation

Effective in version 10.2.2, the **Create or Replace Target Tables** advanced property in a Write transformation for relational, Netezza, and Teradata data objects is renamed to **Target Schema Strategy**.

When you configure a Write transformation, you can choose from the following target schema strategy options for the target data object:

- RETAIN - Retain existing target schema. The Data Integration Service retains the existing target schema.
- CREATE - Create or replace table at run time. The Data Integration Service drops the target table at run time and replaces it with a table based on a target data object that you identify.
- Assign Parameter. Specify the **Target Schema Strategy** options as a parameter value.

Previously, you selected the **Create or Replace Target Tables** advanced property so that the Data Integration Service drops the target table at run time and replaces it with a table based on a target table that you identify. When you do not select the **Create or Replace Target Tables** advanced property, the Data Integration Service retains the existing schema for the target table.

In existing mappings where the **Create or Replace Target Tables** property was enabled, after the upgrade to version 10.2.2, by default, the **Target Schema Strategy** property shows enabled for the **CREATE - Create or replace table at run time** option. In mappings where the **Create or Replace Target Tables** option was not selected, after the upgrade, the **Target Schema Strategy** property is enabled for the **RETAIN - Retain existing target schema** option. After the upgrade, if the correct target schema strategy option is not selected, you must manually select the required option from the **Target Schema Strategy** list, and then run the mapping.

For more information about configuring the target schema strategy, see the "Write Transformation" chapter in the *Informatica Transformation Guide*, or the "Dynamic Mappings" chapter in the *Informatica Developer Mapping Guide*.

PowerExchange Adapters for Informatica

This section describes changes to Informatica adapters in version 10.2.2.

PowerExchange for Amazon Redshift

Effective in version 10.2.2, PowerExchange for Amazon Redshift has the following changes:

- The names of the following advanced properties for an Amazon Redshift data object write operations are changed:

Old Property Name	New Property Name
Null value for CHAR and VARCHAR data types	Require Null Value For Char and Varchar
Wait time in seconds for file consistency on S3	WaitTime In Seconds For S3 File Consistency
Turn on S3 Server Side Encryption	S3 Server Side Encryption
Turn on S3 Client Side Encryption	S3 Client Side Encryption
Prefix to retain staging files on S3	Prefix To Retain For Staging Files On S3

- The default value for the following Unload command is changed:

Unload Command	Old Default Value	New Default Value
DELIMITER	pipe (|)	\036

- The default values for the following Copy commands are changed:

Copy Command	Old Default Value	New Default Value
DELIMITER	pipe (|)	\036
QUOTE	double quote (")	\037

- When you import an Amazon Redshift table in the Developer Tool, you cannot add nullable columns in the table as primary keys.
  Previously, you could add nullable columns in the table as primary keys in the Developer Tool.

For more information, see the *Informatica PowerExchange for Amazon Redshift 10.2.2 User Guide*.

PowerExchange for Amazon S3

Effective in version 10.2.2, PowerExchange for Amazon S3 has the following changes:

- The name of the **Download S3 File in Multiple Parts** advanced source session property is changed to **Multiple Download Threshold**.
- You do not need to add the `GetBucketAcl` permission in the Amazon S3 bucket policy to connect to Amazon S3.
  Previously, you had to add the `GetBucketAcl` permission in the Amazon S3 bucket policy to connect to Amazon S3.

For more information, see the *Informatica PowerExchange for Amazon S3 10.2.2 User Guide*.

PowerExchange for Google Analytics

Effective in version 10.2.2, PowerExchange for Google Analytics installs with the Informatica Services.

Previously, PowerExchange for PowerExchange for Google Analytics had a separate installer.

For more information, see the *Informatica PowerExchange for Google Analytics 10.2.2 User Guide*.

PowerExchange for Google Cloud Storage

Effective in version 10.2.2, PowerExchange for Google Cloud Storage installs with the Informatica Services.

Previously, PowerExchange for PowerExchange for Google Cloud Storage had a separate installer.

For more information, see the *Informatica PowerExchange for Google Cloud Storage 10.2.2 User Guide*.

PowerExchange for HBase

Effective in version 10.2.2, you must run a mapping on the Spark engine to look up data in an HBase resource.

Previously, you could run the mapping in the native environment or on the Spark engine to look up data in an HBase resource.

For more information, see the *Informatica PowerExchange for HBase 10.2.2 User Guide*.

PowerExchange for HDFS

Effective in version 10.2.2, When you run a mapping on the Blaze engine to write data to an HDFS flat file target, the Data Integration Service creates the target files with the following naming convention:

```
<FileName>-P1, <FileName>-P2,....<FileName>-P100....<FileName>-PN
```

Previously, the target files were generated in the following format:

```
Target1.out, Target2.out....Target<PartitionNo>.out
```

For more information, see the *Informatica PowerExchange for HDFS 10.2.2 User Guide*.

PowerExchange for Hive

Effective in version 10.2.2, the **User Name** property in a Hive connection is renamed to **LDAP username**.

PowerExchange for Microsoft Azure Blob Storage

Effective in version 10.2.2, you can navigate through the container structure when you import an object from Microsoft Azure Blob Storage and select objects from sub-directories. Previously, you could select objects present only in the container.

For more information, see the *Informatica PowerExchange for Microsoft Azure Blob Storage 10.2.2 User Guide*.

Part VII: Version 10.2.1

This part contains the following chapters:

- 10.2.1 What's New, 266
- 10.2.1 What's Changed, 300
- 10.2.1 Release Tasks, 319

CHAPTER 29

10.2.1 What's New

This chapter includes the following topics:

- Application Services, 266
- Big Data Management, 268
- Big Data Streaming, 276
- Command Line Programs, 277
- Enterprise Data Catalog, 282
- Enterprise Data Lake, 285
- Informatica Developer, 287
- Informatica Mappings, 289
- Informatica Transformation Language, 292
- Informatica Transformations, 293
- Informatica Workflows, 295
- PowerExchange Adapters for Informatica, 296
- Security, 299

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

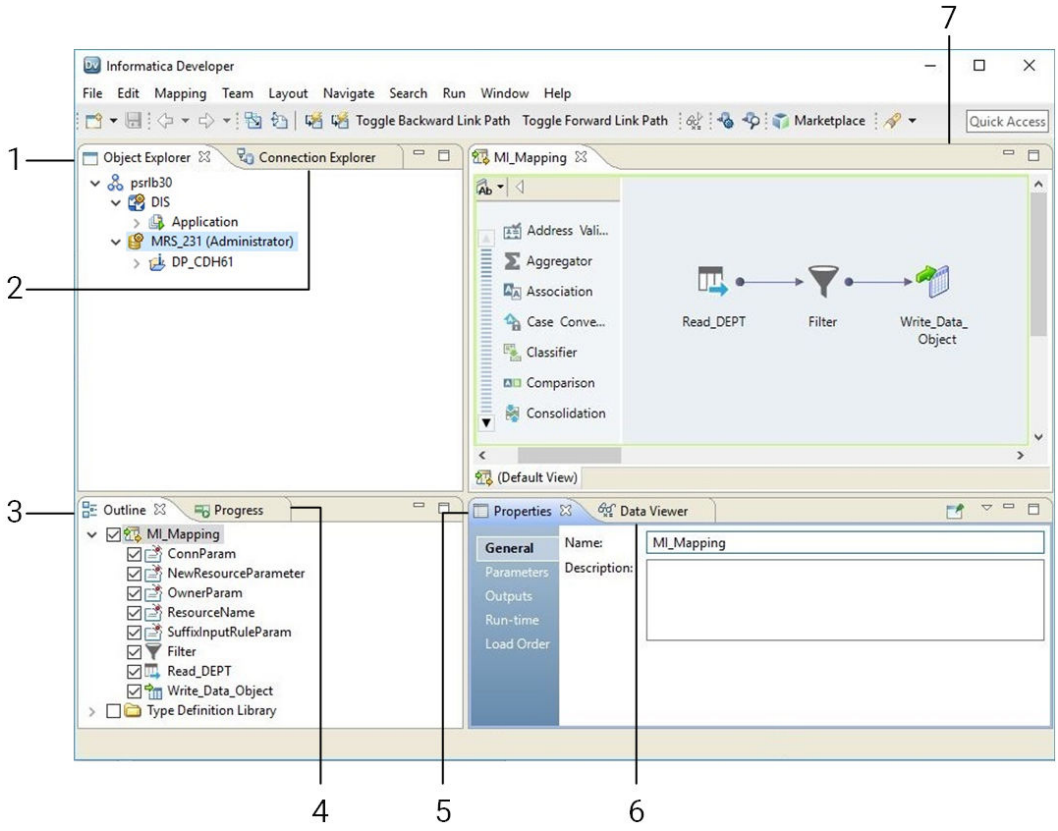
This section describes new Developer tool features in version 10.2.1.

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. Run the mapping in the native environment or on the Spark engine to look up data in an HBase resource. You can enable lookup caching and also parameterize the lookup condition.

The lookup feature for PowerExchange for HBase 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 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 30

10.2.1 What's Changed

This chapter includes the following topics:

- Support Changes, 300
- Installer Changes, 303
- Product Name Changes, 305
- Application Services, 305
- Big Data Management, 305
- Big Data Streaming, 310
- Command Line Programs , 311
- Content Installer, 312
- Enterprise Data Catalog , 312
- Informatica Analyst , 315
- Informatica Developer, 315
- Informatica Transformations, 316
- PowerExchange Adapters for Informatica, 317

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.

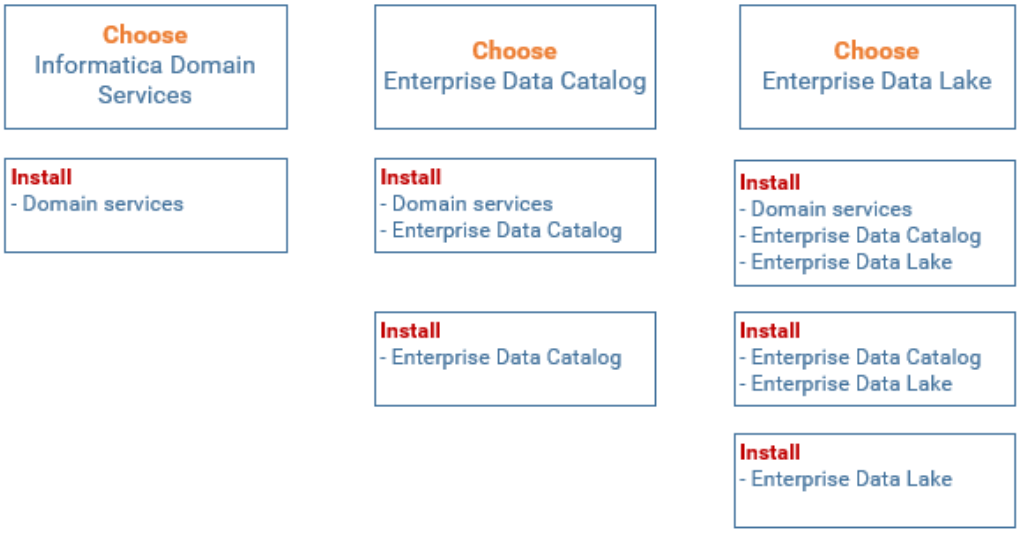
Installer Changes

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:



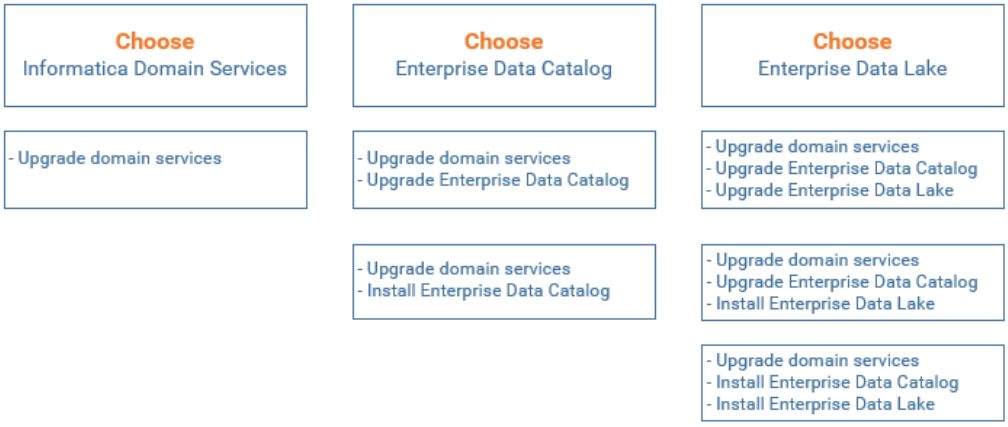
Note: When you install the domain services, the installer also installs application service binaries to support Big Data Management, Big Data Quality, and Big Data Streaming.

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:



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 31

10.2.1 Release Tasks

This chapter includes the following topic:

- PowerExchange Adapters for Amazon S3, 319

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 VIII: Version 10.2 - 10.2 HotFix 2

This part contains the following chapters:

- Version 10.2 HotFix 2, 321
- Version 10.2 HotFix 1, 333
- 10.2 Important Notices, 352
- 10.2 What's New, 355
- 10.2 What's Changed, 390

CHAPTER 32

Version 10.2 HotFix 2

This chapter includes the following topics:

- Important Notices, 321
- What's New, 323
- What's Changed, 329

Important Notices

This section describes important notices for version 10.2 HotFix 2.

Support Changes

This section describes the support changes in version 10.2 HotFix 2.

Verify the Hadoop Distribution Support

Verify the version of the Hadoop distribution in the Hadoop environment.

Big Data Management, Big Data Streaming, Big Data Quality, and PowerCenter supports the following Hadoop distributions:

- Amazon EMR
- Azure HDInsight
- Cloudera CDH
- Hortonworks HDP
- MapR

In each release, Informatica can add, defer, and drop support for the non-native distributions and 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 Customer Portal:

https://network.informatica.com/community/informatica-network/product-availability-matrices

OpenJDK

Effective in version 10.2 HotFix 2, Informatica installer packages OpenJDK (AzulJDK). The supported Java version is Azul OpenJDK 1.8.192.

Sqoop mappings on the Spark engine do not use the Azul OpenJDK packaged with the Informatica installer. Sqoop mappings on the Spark and Blaze engines continue to use the `infapdo.env.entry.hadoop_node_jdk_home` property in the `hadoopEnv.properties` file. The `HADOOP_NODE_JDK_HOME` represents the directory from which you run the cluster services and the JDK version that the cluster nodes use.

Previously, the installer used the Oracle Java packaged with the installer.

DataDirect SQL Server Legacy ODBC Driver

Effective in version 10.2 HotFix 2, Informatica dropped support for the DataDirect SQL Server Legacy ODBC Driver file `DWmsss27.x` because DataDirect no longer supports this driver.

When you use an ODBC connection to connect to Microsoft SQL Server, you can use the DataDirect 8.0 SQL Server Wire Protocol packaged with the Informatica installer, or any ODBC driver from a third-party vendor.

PowerExchange for SAP NetWeaver

Effective in version 10.2 HotFix 2, PowerExchange for SAP NetWeaver supports the SAP NetWeaver RFC SDK 7.50 libraries for PowerCenter.

The existing mappings that use the SAP NetWeaver RFC SDK 7.20 libraries will not fail. However, Informatica recommends that you download and install the SAP NetWeaver RFC SDK 7.50 libraries.

New Products

PowerExchange Adapters for PowerCenter

This section describes new PowerCenter adapters in version 10.2 HotFix 2.

PowerExchange for Tableau V3

Effective in version 10.2 HotFix 2, you can use the Tableau V3 connection to read data from multiple sources, generate a Tableau `.hyper` output file, and write the data to Tableau.

For more information, see the *Informatica PowerExchange for Tableau V3 10.2 HotFix 2 User Guide for PowerCenter*.

Release Tasks

PowerExchange Adapters for Informatica

This section describes release tasks for Informatica adapters in version 10.2 HotFix 2.

PowerExchange for Snowflake

Effective in version 10.2 HotFix 2, the Snowflake JDBC driver version 3.6.26 installs with the Informatica Services. Existing Snowflake mappings from 10.2 HotFix 1 release might fail after the upgrade because the installation location now contains the Snowflake JDBC driver version 3.6.4 from version 10.2 HotFix 1 and Snowflake JDBC driver version 3.6.26 from version 10.2 HotFix 2.

To run existing Snowflake mappings from the Developer Tool, you must delete the Snowflake JDBC driver version 3.6.26 and retain only version 3.6.4 in the following directory in the Data Integration Service machine: `<Informatica installation directory>\connectors\thirdparty`

PowerExchange Adapters for PowerCenter

This section describes release tasks for PowerCenter adapters in version 10.2 HotFix 2.

PowerExchange for Google BigQuery

Effective in version 10.2 HotFix 2, the PowerExchange for Google BigQuery plug-in has been updated with new elements for **Custom Query**, **SQL Override**, and **Quote Char** properties. Existing Google BigQuery mappings from the earlier release might fail after the upgrade because the existing plug-in does not contain the new elements.

To run existing Google BigQuery mappings from the PowerCenter client and enable the **Custom Query**, **SQL Override**, and **Quote Char** properties, you must re-register the `bigqueryPlugin.xml` plug-in with the PowerCenter repository.

What's New

Command Line Programs

This section describes new commands in version 10.2 HotFix 2.

pmrep Commands

You can use the pmrep createconnection and pmrep updateconnection command option `-S` with argument `odbc_subtype` to enable the ODBC subtype option when you create or update an ODBC connection.

The following table describes the new pmrep createconnection and pmrep updateconnection command option:

Option	Argument	Description
-S	odbc_subtype	Optional. Enables the ODBC subtype for an ODBC connection. An ODBC subtype can be one of the following types: - AWS Redshift - Azure DW - Greenplum - Google Big Query - PostgreSQL - SAP HANA - Snowflake - None Default is None.

For more information, see the "pmrep Command Reference" chapter in the *Informatica 10.2 HotFix 2 Command Reference*.

Informatica Analyst

This section describes new Analyst tool features in version 10.2 HotFix 2.

Scorecards

Effective in version 10.2 HotFix 2, you can enter the number of invalid rows that you want to export from a scorecard trend chart.

For more information, see the "Scorecards in Informatica Analyst" chapter in the *Informatica 10.2 HotFix 2 Data Discovery Guide*.

Informatica Transformations

This section describes new features in Informatica transformations in version 10.2 HotFix 2.

Address Validator Transformation

The Address Validator transformation contains additional address functionality for the following countries:

Australia

Effective in version 10.2 HotFix 2, you can configure the Address Validator transformation to add address enrichments to Australia addresses. You can use the enrichments to discover the geographic sectors and regions to which the Australia Bureau of Statistics assigns the addresses. The sectors and regions include census collection districts, mesh blocks, and statistical areas.

The transformation uses the following ports to deliver the enrichments:

- Census Collection District Code 2006
- Geocoded National Address File Identifier
- Greater Capital City Statistical Area 5-Digit
- Greater Capital City Statistical Area Name
- Level One Statistical Area 11-Digit
- Level One Statistical Area 7-Digit
- Level Two Statistical Area 9-Digit
- Level Two Statistical Area 5-Digit
- Level Two Statistical Area Name
- Level Three Statistical Area 5-Digit
- Level Three Statistical Area Name
- Level Four Statistical Area 3-Digit
- Level Four Statistical Area Name
- Mesh Block 11-Digit 2011
- Mesh Block 11-Digit 2016
- State or Territory Code
- State or Territory Name
- Supplementary AU Status

Find the ports in the AU Supplementary port group.

Israel

Effective in version 10.2 HotFix 2, Informatica introduces the following features and enhancements for Israel:

Multilanguage support for Israel addresses

You can configure the Address Validator transformation to return an Israel address in the English language or the Hebrew language.

Use the Preferred Language property to select the preferred language for the addresses that the transformation returns.

The default language for Israel addresses is Hebrew. To return address information in Hebrew, set the Preferred Language property to DATABASE or ALTERNATIVE_1. To return the address information in English, set the property to ENGLISH or ALTERNATIVE_2.

Support for multiple character sets in Israel addresses

The Address Validator transformation can read and write Israel addresses in Hebrew and Latin character sets.

Use the Preferred Script property to select the preferred character set for the address data.

The default character set for Israel addresses is Hebrew. When you set the Preferred Script property to Latin or Latin-1, the transformation transliterates Hebrew address data into Latin characters.

United States

Effective in version 10.2 HotFix 2, Informatica introduces the following features and enhancements for the United States:

Additional information on undeliverable mail

You can configure the Address Validator transformation to return information on why the United States Postal Service does not deliver mail to an address that appears to be valid.

The United States Postal Service might not deliver mail to an address for one of the following reasons:

- The address site is under construction.
- The customer receives mail as a part of mail forwarding service.
- The postal carrier returns the mail undelivered.

The United States Postal Service maintains a table of addresses that do not permit delivery. The table is called the No-Statistics table. To return a code that indicates why the United States Postal Service adds the address to the No-Statistics table, select the Delivery Sequence File Second Generation No Statistics Reason port. Find the port in the US Specific port group in the Basic model.

The transformation reads the No-Statistics table data from the `USA5C131.MD` database file.

Improved correction for addresses with trailing house number characters

You can configure the Address Validator transformation to identify a valid street address that includes a house number value with an unrecognized trailing alphabetic character. If the address can yield a valid delivery point without the trailing character, the transformation returns the value TA in the Delivery Point Validation Footnote ports.

Information on street addresses that forward mail to a Post Office Box

You can configure the Address Validator transformation to identify a street address for which the United States Postal Service redirects mail to a Post Office Box. To identify the address, select the Delivery Point Validation Throwback port. Find the port in the US Specific port group in the Basic model.

The transformation reads the throwback data from the`USA5C132.MD` database file.

Information on addresses that cannot accept mail on one or more days of the week

You can configure the Address Validator transformation to identify United States addresses that do not receive mail on one or more days of the week.

To identify the addresses, use the *Non-Delivery Days* port. The port contains a seven-digit string that represents the days of the week from Sunday through Saturday. Each position in the string represents a different day.

Find the Non-Delivery Days port in the US Specific port group in the Basic model. To receive data on the Non-Delivery Days port, run the Address Validator transformation in certified mode. The transformation reads the port values from the `USA5C129.MD` and `USA5C130.MD` database files.

For more information, see the *Informatica 10.2 HotFix 2 Developer Transformation Guide* and the *Informatica 10.2 HotFix 2 Address Validator Port Reference*.

For comprehensive information about the features and operations of the address verification software engine in version 10.2 HotFix 2, see the *Informatica Address Verification 5.14.0 Developer Guide*.

Metadata Manager

This section describes new Metadata Manager features in version 10.2 HotFix 2.

Cognos

Effective in version 10.2 HotFix 2, you can set up the following configuration properties for a Cognos resource:

- Cognos SDK directory. Enter the Cognos SDK directory location.
- Miscellaneous. Specify multiple miscellaneous options separated by commas.

For more information, see the "Business Intelligence Resources" chapter in the *Informatica 10.2 HotFix 2 Metadata Manager Administrator Guide*.

Microstrategy

Effective in version 10.2 HotFix 2, you can set up the **Miscellaneous** configuration property for a Microstrategy resource. Use the **Miscellaneous** configuration property to specify multiple miscellaneous options separated by commas.

For more information, see the "Business Intelligence Resources" chapter in the *Informatica 10.2 HotFix 2 Metadata Manager Administrator Guide*.

PowerCenter

This section describes new PowerCenter features in version 10.2 HotFix 2.

Pushdown Optimization for PostgreSQL

Effective in version 10.2 HotFix 2, when the connection type is ODBC, you can select the ODBC subtype as **PostgreSQL** to push transformation logic to PostgreSQL. You can configure source-side or full pushdown optimization to push the transformation logic to PostgreSQL.

For more information about the supported functions and transformations that you can push to the PostgreSQL database, see the *Informatica PowerCenter 10.2 HotFix 2 Advanced Workflow Guide*.

PowerCenter Binary Functions

Effective in version 10.2 HotFix 2, you can use the following functions for binary expressions:

- EBCDIC_ISO88591. Converts a binary value encoded in EBCDIC to a string value encoded in ISO-8859-1.
- BINARY_COMPARE. Compares two binary values and returns TRUE (1) if they are the same and FALSE (0) if they are different.
- BINARY_CONCAT. Concatenates two or more binary values together and returns the concatenated value.
- BINARY_LENGTH. Returns the length of a binary value.
- BINARY_SECTION. Returns a portion of a binary value.
- DEC_HEX. Decodes a hex encoded value and returns a binary value with the binary representation of the data.
- ENC_HEX. Encodes binary data to string data using hexadecimal encoding.
- SHA256. Calculates the SHA-256 digest of the input value.

For more information about the custom functions, see the *Informatica PowerCenter 10.2 HotFix 2 Transformation Language Reference*.

PowerExchange Adapters for PowerCenter

This section describes new PowerCenter adapter features in version 10.2 HotFix 2.

PowerExchange for Amazon Redshift

Effective in version 10.2 HotFix 2, you can retain the null values when you read data from Amazon Redshift.

For more information, see the *Informatica PowerExchange for Amazon Redshift 10.2 HotFix 2 User Guide for PowerCenter*.

PowerExchange for Google BigQuery

Effective in version 10.2 HotFix 2, PowerExchange for Google BigQuery includes the following new features:

- You can configure a custom SQL query to configure a Google BigQuery source.
- You can configure a SQL override to override the default SQL query used to extract data from the Google BigQuery source.
- You can specify a quote character that defines the boundaries of text strings in a .csv file. You can configure parameters such as single quote or double quote.
- You can configure full pushdown optimization to push the transformation logic to Google BigQuery when you use the ODBC connection type.
  For information about the operators and functions that the PowerCenter Integration Service can push to Google BigQuery, see the *Informatica PowerCenter 10.2 HotFix 2 Advanced Workflow Guide*.

For more information, see the *Informatica PowerExchange for Google BigQuery 10.2 HotFix 2 User Guide for PowerCenter*.

PowerExchange for Kafka

Effective in version 10.2 HotFix 2, PowerExchange for Kafka includes the following new features:

- You can import a Kafka source or target definition in Avro format.
- You can configure SSL configuration properties to securely connect to Kafka.
- You can preview data of a Kafka source or target definition.

For more information, see the *Informatica PowerExchange for Kafka 10.2 HotFix 2 User Guide for PowerCenter*.

PowerExchange for Microsoft Azure SQL Data Warehouse V3

Effective in version 10.2 HotFix 2, PowerExchange for Microsoft Azure SQL Data Warehouse V3 includes the following new features:

- You can use custom queries when you read a source object.
- You can override the source object and the source object schema in the source session properties. The source object and the source object schema defined in the source session properties take the precedence.
- You can override the target object and the target object schema in the target session properties. The target object and the target object schema defined in the target session properties take the precedence.
- You can create a mapping to read the real-time or changed data from a Change Data Capture (CDC) source and load data into Microsoft Azure SQL Data Warehouse. You must select **Data Driven** as the target operation to capture changed data. You can resume the extraction of changed data from the point of interruption when a mapping fails or is stopped before completing the session.
- You can utilize the following pushdown functions when you use an ODBC connection to connect to Microsoft Azure SQL Data Warehouse:
  - `Date_diff()`
  - `First()`
  - `Instr()`
  - `Last()`
  - `MD5()`
  - `ReplaceChr()`
  - `ReplaceStr()`
- To enhance the performance, you can compress staging files in the .gzip format when you write to a target object.

For more information, see the *Informatica PowerExchange for Microsoft Azure SQL Data Warehouse V310.2 HotFix 2 User Guide for PowerCenter*.

PowerExchange for Salesforce

Effective in version 10.2 HotFix 2, PowerExchange for Salesforce includes the following new features:

- You can use version 43.0, 44.0, and 45.0 of Salesforce API to create a Salesforce connection and access Salesforce objects.
- You can enable primary key chunking for queries on a shared object that represents a sharing entry on the parent object. Primary key chunking is supported for shared objects only if the parent object is supported. For example, if you want to query on CaseHistory, primary key chunking must be supported for the parent object Case.
- You can create assignment rules to reassign attributes in records when you insert, update, or upsert records for Lead and Case target objects using the standard API.
- You can parameterize the Salesforce service URL for standard and OAuth connection.

For more information, see the *Informatica PowerExchange for Salesforce 10.2 HotFix 2 User Guide for PowerCenter*.

PowerExchange for Snowflake

Effective in version 10.2 HotFix 2, PowerExchange for Snowflake includes the following new features:

- You can configure the PowerCenter Integration Service to write rejected records to a rejected file. If the file already exists, the PowerCenter Integration Service appends the rejected records to that file. If you do not specify the rejected file path, the PowerCenter Integration Service does not write the rejected records.
- You can use an Update Strategy transformation in a mapping to insert, update, delete, or reject data in Snowflake targets. When you configure an Update Strategy transformation, the **Treat Source Rows As** session property is marked **Data Driven** by default.
- You can use the SQL editor to create or edit the pre-SQL and post-SQL statements in the Snowflake source and target sessions.

For more information, see the *Informatica PowerExchange for Snowflake 10.2 HotFix 2 User Guide for PowerCenter*.

Security

This section describes new security features in 10.2 HotFix 2.

Kerberos Cross Realm Authentication

Effective in version 10.2 HotFix 2, you can configure an Informatica domain to use Kerberos cross realm authentication. Kerberos cross realm authentication enables Informatica clients that belong to one Kerberos realm to authenticate with nodes and application services that belong to another Kerberos realm.

For more information, see the *Informatica 10.2 HotFix 2 Security Guide*.

What's Changed

Analyst Tool

This section describes changes to the Analyst tool in version 10.2 HotFix 2.

Default View

Effective in version 10.2 Hotfix 2, the default view for flat file and table objects is the Properties tab. When you create or open a flat file or table data object, the object opens in the Properties tab. Previously, the default view was the Data Viewer tab.

For more information, see the *Informatica 10.2 Hotfix 2 Analyst Tool Guide*.

Scorecards

Effective in version 10.2 HotFix 2, you can export a maximum of 100,000 invalid rows. When you export more than 100 invalid rows for a metric, the Data Integration Service creates a folder for the scorecard, subfolder for the metric, and Microsoft Excel files to export the rest of the invalid rows.

For more information, see the "Scorecards" chapter in the *Informatica 10.2 HotFix 2 Data Discovery Guide*.

Command Line Programs

This section describes changes to commands in version 10.2 HotFix 2.

infasetup Commands

Effective in version 10.2 HotFix 2, the valid values for the -srn and -urn options used in infasetup commands are changed.

If you configure a domain to use Kerberos authentication, the values for the -srn and -urn options do not need to be identical. If you configure a domain to use Kerberos cross realm authentication, you can specify a string containing the name of each Kerberos realm that the domain uses to authenticate users, separated by a comma.

Previously, you could specify the name of a single Kerberos realm as the value for both the -srn and -urn options.

For more information, see the *Informatica 10.2 HotFix 2 Command Reference*.

Informatica Transformations

This section describes changes to Informatica transformations in version 10.2 HotFix 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 HotFix 2, the Address Validator transformation uses version 5.14.0 of the Informatica Address Verification software engine.

Previously, the transformation used version 5.13.0 of the Informatica Address Verification software engine.

Parcel Centroid and Rooftop Geocodes

Effective in version 10.2 HotFix 2, the Geocode Data Type property on the Address Validator transformation does not include options for Parcel Centroid or Rooftop geocode analysis.

For more information, see the *Informatica 10.2 HotFix 2 Developer Transformation Guide* and the *Informatica 10.2 HotFix 2 Address Validator Port Reference*.

For comprehensive information about the updates to the Informatica Address Verification software engine, see the *Informatica Address Verification 5.14.0 Release Guide*.

Metadata Manager

SAP Business Warehouse (Deprecated)

Effective in version 10.2 HotFix 2, Informatica deprecated the SAP Business Warehouse resource in Metadata Manager.

Cognos Version 8 or Earlier

Effective in version 10.2 HotFix 2, if the Cognos server version is later than version 8, remove the *axis* and *saaj* jar files from the `<MM_Agent_Home>\MetadataManagerAgent\java\CognosRepository\` location.

For more information, see the "Business Intelligence Resources" chapter in the *Informatica 10.2 HotFix 2 Metadata Manager Administrator Guide*.

PowerCenter

This section describes the PowerCenter changes in version 10.2 HotFix 2.

Pushdown Optimization for SAP HANA

Effective in version 10.2 HotFix 2, you must have a license to use the **SAP HANA** ODBC subtype in ODBC connections. If you do not have a license, a warning message appears. The sessions do not fail in version 10.2 HotFix 2. However, Informatica will start failing sessions in a future release. if you use the **SAP HANA** ODBC subtype connection without a license. Hence, Informatica recommends that you contact Informatica Global Customer Support to get a license.

PowerExchange Adapters for Informatica

This section describes changes to Informatica adapters in version 10.2 HotFix 2.

PowerExchange for MongoDB

Effective in version 10.2 HotFix 2, the Schema Editor is no longer available in the following directory:

```
Informatica installation directory>/clients/tools/mongodb/Tools
```

To use the Schema Editor on a Windows machine, apply the Informatica EBF-13871.

For more information, see the *Informatica PowerExchange for MongoDB 10.2 HotFix 2 User Guide*.

PowerExchange Adapters for PowerCenter

This section describes changes to PowerCenter adapters in version 10.2 HotFix 2.

PowerExchange for Google Analytics

Effective in version 10.2 HotFix 2, PowerExchange for Google Analytics installs with the Informatica services.

Previously, PowerExchange for PowerExchange for Google Analytics had a separate installer.

For more information, see the *Informatica PowerExchange for Google Analytics 10.2 HotFix 2 User Guide for PowerCenter*.

PowerExchange for Google Cloud Spanner

Effective in version 10.2 HotFix 2, PowerExchange for Google Cloud Spanner installs with the Informatica services.

Previously, PowerExchange for PowerExchange for Google Cloud Spanner had a separate installer.

For more information, see the *Informatica PowerExchange for Google Cloud Spanner 10.2 HotFix 2 User Guide for PowerCenter*.

PowerExchange for Google Cloud Storage

Effective in version 10.2 HotFix 2, PowerExchange for Google Cloud Spanner installs with the Informatica services.

Previously, PowerExchange for PowerExchange for Google Cloud Spanner had a separate installer.

For more information, see the *Informatica PowerExchange for Google Cloud Spanner 10.2 HotFix 2 User Guide for PowerCenter*.

PowerExchange for Kafka

Effective in version 10.2 HotFix 2, PowerExchange for Kafka installs with the Informatica services.

Previously, PowerExchange for PowerExchange for Kafka had a separate installer.

For more information, see the *Informatica PowerExchange for Kafka 10.2 HotFix 2 User Guide for PowerCenter*.

PowerExchange for MongoDB

Effective in version 10.2 HotFix 2, the Schema Editor is no longer available in the following directory:

```
Informatica installation directory>/clients/tools/mongodb/Tools
```

To use the Schema Editor on a Windows machine, apply the Informatica EBF-13871.

For more information, see the *Informatica PowerExchange for MongoDB 10.2 HotFix 2 User Guide for PowerCenter*.

PowerExchange for Salesforce

Effective in version 10.2 HotFix 2, the null values in the source are treated as null in the target.

Previously, the null values in the source were treated as false in the target.

For more information, see the *Informatica PowerExchange for Salesforce 10.2 HotFix 2 User Guide for PowerCenter*.

PowerExchange for Snowflake

Effective in version 10.2 HotFix 1, PowerExchange for Snowflake has the following changes:

- You can override the database and schema name used while importing the tables through the Snowflake connection. To override, enter the database and schema name in the **Additional JDBC URL Parameters** field in the Snowflake connection properties in the following format:
  `DB=<DB_name>&Schema=<schema_name>`
  Previously, you could specify an override for the database and schema name only in the session properties.
- You can read or write data that is case-sensitive or contains special characters. You cannot use the following special characters: @ ~ \
  Previously, you had to ensure that the source and target table names contained only uppercase letters.
- PowerExchange for Snowflake uses the Snowflake JDBC driver version 3.6.26.
  Previously, PowerExchange for Snowflake used the Snowflake JDBC driver version 3.6.4.
- When you run a custom query to read data from Snowflake, the Integration Service runs the query and the performance of importing the metadata is optimized.
  Previously, when you configured a custom query to read data from Snowflake, the Integration Service sent the query to the Snowflake endpoint and the number of records to import was limited to 10 records.

For more information, see the *Informatica PowerExchange for Snowflake 10.2 HotFix 2 User Guide for PowerCenter*.

PowerExchange for Teradata Parallel Transporter API

Effective in 10.2 HotFix 2, you can run a Teradata ODBC mapping on the AIX machine. You must install the Teradata ODBC driver version 16.20.00.50 or later when you use the Teradata client version 16.20.x.

Previously, you could not run a Teradata ODBC mapping on the AIX machine when you used the Teradata ODBC driver versions earlier than 16.20.00.50.

CHAPTER 33

Version 10.2 HotFix 1

This chapter includes the following topics:

- What's New, 333
- What's Changed, 344
- Release Tasks, 350

What's New

Application Services

This section describes new application service features in version 10.2 HotFix 1.

Model Repository Service

Git Version Control System

Effective in version 10.2 HotFix1, 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.

For more information, see the "Model Repository Service" chapter in the *Informatica 10.2 HotFix 1 Application Service Guide*.

Business Glossary

This section describes new Business Glossary features in version 10.2 HotFix 1.

Import Glossary Assets as Plain Text

Effective in 10.2 HotFix 1, you can import business glossary assets as plain text into the Analyst tool.

For more information about export and import of glossary assets, see the "Glossary Administration" chapter in the *Informatica 10.2 HotFix 1 Business Glossary Guide*.

Command Line Programs

This section describes new commands in version 10.2 HotFix 1.

infacmd isp Commands

The following table describes new infacmd isp commands:

Command	Description
PingDomain	Pings a domain, service, domain gateway host, or node.
ListPasswordRules	Lists the rules to set up complex passwords.
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 HotFix 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*.

Connectivity

This section describes new connectivity features in version 10.2 HotFix 1.

Connectivity to Cloud Databases

Effective in version 10.2 HotFix 1, you can configure relational connections in Informatica Developer and PowerCenter to connect to Cloud databases.

You can configure the following connections:

- Oracle connection to connect to Oracle Autonomous Data Warehouse Cloud version 18C
- Oracle connection to connect to Oracle Database Cloud Service version 12C
- Microsoft SQL Server connection to connect to Azure SQL Database
- IBM DB2 connection to connect to DashDB

For more information, see the *Informatica 10.2 HotFix 1 Installation and Configuration Guide*.

Data Types

This section describes new data type features in 10.2 HotFix 1.

Microsoft SQL Server Data Types

Effective in version 10.2 HotFix 1, you can read or write the date data type when you use the Microsoft SQL Server connection in an Informatica mapping.

For more information, see the Data Type Reference appendix in the *Informatica 10.2 HotFix 1 Developer Tool Guide*.

Installer

This section describes new Installer feature in version 10.2 HotFix 1.

Docker Utility

Effective in version 10.2 HotFix 1, you can use the Informatica PowerCenter Docker utility to create the Informatica domain services. You can build the Informatica docker image with base operating system and Informatica binaries and run the existing docker image to create the Informatica domain within a container.

Informatica Transformations

This section describes new features in Informatica transformations in version 10.2 HotFix 1.

Address Validator Transformation

This section describes new Address Validator transformation features.

The Address Validator transformation contains additional address functionality for the following countries:

All Countries

Effective in version 10.2 HotFix 1, the Address Validator transformation supports single-line address verification in every country for which Informatica provides reference address data.

In earlier versions, the transformation supported single-line address verification for 26 countries.

To verify a single-line address, enter the address in the Complete Address port. If the address identifies a country for which the default preferred script is not a Latin or Western script, use the default Preferred Script property on the transformation with the address.

Austria, Germany, and Switzerland

Effective in version 10.2 HotFix 1, the Address Validator transformation supports the uppercase character ß in Austria, Germany, and Switzerland addresses.

The transformation supports the character ß in the following ways:

- If you set the Casing property to UPPER, the transformation returns the German character ß as ß. If you set the Casing property to LOWER, the transformation returns the German character ß as ß.
- The transformation treats ß and ß as equally valid characters in an address. In reference data matches, the transformation can identify a perfect match when the same values contain either ß or ß.
- The transformation treats ß and ss as equally valid characters in an address. In reference data matches, the transformation can identify a standardized match when the same values contain either ß or ss.
- If you set the Preferred Script property to ASCII_SIMPLIFIED, the transformation returns the character ß as S.
- If you set the Preferred Script property to ASCII_EXTENDED, the transformation returns the character ß as SS.

Bolivia

Effective in version 10.2 HotFix 1, the Address Validator transformation improves the parsing and validation of Bolivia addresses. Additionally, Informatica updates the reference data for Bolivia.

The transformation also includes the following improvements for Bolivia:

- Address validation to street level.
- Geocoordinates at street mid-point level for addresses in major cities.

Canada

Informatica introduces the following features and enhancements for Canada:

Support for the Global Preferred Descriptor property in Canada Addresses

Effective in version 10.2 HotFix 1, you can configure the transformation to return the short or long form of an element descriptor.

Address Verification can return the short or long form of the following descriptors:

- Street descriptors
- Directional values
- Building descriptors
- Sub-building descriptors

To specify the output format for the descriptors, configure the Global Preferred Descriptor property on the transformation. The property applies to English-language and French-language descriptors. By default, the transformation returns the descriptor in the format that the reference data specifies. If you select the PRESERVE INPUT option on the property, the Preferred Language property takes precedence over the Global Preferred Descriptor property.

Support for CH and CHAMBER as Sub-Building Descriptors

Effective in version 10.2 HotFix 1, Address Validator transformation recognizes CH and CHAMBER as sub-building descriptors in Canada addresses.

Colombia

Effective in version 10.2 HotFix1, the Address Validator transformation improves the processing of street data in Colombia addresses. Additionally, Informatica updates the reference data for Colombia.

The Address Validator transformation validates an address in Colombia to house number level. The transformation can verify a Colombia address that includes information for the street on which the house is located and also for the nearest cross street to the house.

For example, the transformation can verify the following addresses:

AVENIDA 31 65 29 APTO 1626

AVENIDA 31 DIAGONAL 65 29 APTO 1626

Consider the following guidelines for Colombia addresses:

- The Address Validator transformation can verify the address with and without the cross street descriptor DIAGONAL.
- The transformation can verify the address with and without a dash symbol between the cross street number and the house number. The transformation does not include a dash symbol in this position in the output address.

- The transformation can recognize the # symbol before the cross street number in an input address. The transformation recognizes the address as a reference data match and not a correction.

  For example, address validation recognizes the following input address as a reference data match and omits the # symbol from the verified output address:

  AVENIDA 31 #65 29 APTO 1626

France

Effective in version 10.2 HotFix 1, Informatica introduces the following improvements for France addresses:

- Informatica improves the supplementary reference data for France.
- The Address Validator transformation assigns addresses to IRIS units in France with greater accuracy. The transformation uses the house number in the address to verify the IRIS unit to which the address belongs. The use of house numbers can improve the assignment accuracy when the address lies close to the border between different units.

India

Effective in version 10.2 HotFix 1, the Address Validator transformation validates an address in India to house number level.

Peru

Effective in version 10.2 HotFix 1, the Address Validator transformation validates a Peru address to house number level. Additionally, Informatica updates the reference data for Peru.

South Africa

Effective in version 10.2 HotFix 1, the Address Validator transformation improves the parsing and verification of delivery service descriptors in South Africa addresses.

The transformation 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 HotFix 1, the Address Validator transformation introduces the following features and enhancements for South Korea:

- The South Korea address reference data includes building information. The transformation can read, verify, and correct building information in a South Korea address.
- The transformation 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, the transformation 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.

Sweden

Effective in version 10.2 HotFix 1, the Address Validator transformation improves the verification of street names in Sweden addresses.

The transformation improves the verification of street names in the following ways:

- The transformation can recognize a street name that ends in the character G as an alias of the same name with the final characters GATAN.
- The transformation can recognize a street name that ends in the character V as an alias of the same name with the final characters VÄGEN.
- The Address Validator transformation can recognize and correct a street name with an incorrect descriptor when either the long form or the short form of the descriptor is used.

  For example, The transformation can correct RUNIUSV or RUNIUSVÄGEN to RUNIUSGATAN in the following address:

  RUNIUSGATAN 7

  SE-112 55 STOCKHOLM

Thailand

Effective in version 10.2 HotFix 1, the Address Validator transformation introduces the following features and enhancements for Thailand:

Improvements to Thailand Addresses

The transformation improves the parsing and validation of Thailand addresses in a Latin script.

Additionally, the transformation validates an address to house number level.

Native Support for Thailand Addresses

The Address Validator transformation 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 HotFix 1, the Address Validator transformation 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 HotFix 1, the Address Validator transformation can return a United Kingdom territory name.

The transformation returns the territory name in the Country_2 element and 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 HotFix 1, the Address Validator transformation can recognize up to three sub-building levels in a United States address.

In compliance with the United States Postal Service requirements, the transformation matches the information in a single sub-building element with the reference data. If the Sub-building_1 information does not match, the transformation compares the Sub-building_2 information. If the Sub-building_2 information does not match, the transformation compares the Sub-building_3 information. The transformation copies the unmatched sub-building information from the input address to the output address.

For comprehensive information about the features and operations of the address verification software engine version that Informatica embeds in version 10.2 HotFix 1 see the *Informatica Address Verification 5.13.0 Developer Guide*.

Metadata Manager

This section describes new Metadata Manager features in version 10.2 HotFix 1.

SAML Authentication for Metadata Manager

Effective in version 10.2 HotFix 1, Metadata Manager supports Security Assertion Markup Language (SAML)-based single sign-on. SAML-based single sign-on authenticates users against account credentials stored in Microsoft Active Directory. Accounts are imported from Active Directory into a security domain within the Informatica domain.

For information, see the "SAML Authentication for Informatica Web Applications" chapter in the *Informatica 10.2 HotFix 1 Security Guide*.

Skip Lineage During Metadata Manager Repository Backup or Restore Operations

Effective in version 10.2 HotFix 1, use the `[<-sl|--skipLineage> skipLineage]` option in the `backupRepository` and `restoreRepository` command to skip lineage during Metadata Manager repository back up and restore operations.

For information, see the *Informatica 10.2 HotFix 1 Metadata Manager Command Reference Guide*.

PowerCenter

This section describes new PowerCenter features in version 10.2 HotFix 1.

Pushdown Optimization for SAP HANA

Effective in version 10.2 HotFix 1, when the connection type is ODBC, you can select the ODBC provider sub-type as **SAP HANA** to push transformation logic to SAP HANA. You can configure source-side, target-side, or full pushdown optimization to push the transformation logic to SAP HANA.

For more information, see the *Informatica PowerCenter 10.2 HotFix 1 Advanced Workflow Guide*.

Pushdown Optimization for Snowflake

Effective in version 10.2 HotFix 1, when the connection type is ODBC, you can configure source-side or full pushdown optimization to push the transformation logic to Snowflake.

For information, see the *Informatica PowerCenter 10.2 HotFix 1 Advanced Workflow Guide*.

PowerExchange Adapters for Informatica

This section describes new Informatica adapter features in version 10.2 HotFix 1.

PowerExchange for Microsoft Azure Blob Storage

Effective in version 10.2 HotFix 1, PowerExchange for Microsoft Azure Blob Storage includes the following functionality:

- You can run mappings on the Spark engine.
- You can use the file port to store the file name from which the Data Integration Service reads the data at run-time.
- You can read and write .csv, Avro, and Parquet files when you run a mapping in the native environment.
- You can read a directory when you run a mapping in the native environment.
- 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 .csv files 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 in the native environment.

For more information, see the *Informatica PowerExchange for Microsoft Azure Blob Storage 10.2 HotFix 1 User Guide*.

PowerExchange for Microsoft Azure SQL Data Warehouse

Effective in version 10.2 HotFix 1, PowerExchange for Microsoft Azure SQL Data Warehouse includes the following features:

- 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 an uncached lookup operation in the native environment.

For more information, see the *Informatica PowerExchange for Microsoft Azure SQL Data Warehouse 10.2 HotFix 1 User Guide*.

PowerExchange for Netezza

Effective in version 10.2 HotFix 1, you can configure dynamic mappings to change Netezza sources and targets at run time based on parameters and rules that you define.

When you configure dynamic mappings, you can also create or replace the target at run time. You can select the **Create or replace table at run time** option in the advanced properties of the Netezza data object write operation.

PowerExchange for Teradata Parallel Transporter API

Effective in version 10.2 HotFix 1, you can configure dynamic mappings to change Teradata sources and targets at run time based on parameters and rules that you define.

When you configure dynamic mappings, you can also create or replace the Teradata target at run time. You can select the **Create or replace table at run time** option in the advanced properties of the Teradata data object write operation.

PowerExchange Adapters for PowerCenter

This section describes new PowerCenter adapter features in version 10.2 HotFix 1.

PowerExchange for Amazon Redshift

Effective in version 10.2 HotFix 1, PowerExchange for Amazon Redshift includes the following new features:

- In addition to the existing regions, you can also read data from or write data to the `AWS GovCloud` region.
- You can specify the part size of an object to download the object from Amazon S3 in multiple parts.
- You can encrypt data while fetching the file from Amazon Redshift using the AWS-managed encryption keys or AWS KMS-managed customer master key for server-side encryption.
- You can provide the number of files to calculate the number of the staging files for each batch. If you do not provide the number of files, PowerExchange for Amazon Redshift calculates the number of the staging files.
- You can use the TRUNCATECOLUMNS option in the copy command to truncate the data of the VARCHAR and CHAR data types column before writing data to the target.
- PowerExchange for Amazon Redshift supports version 11 and 12 of SuSe Linux Enterprise Server operating system.

For more information, see the *Informatica PowerExchange for Amazon Redshift 10.2 HotFix 1 User Guide for PowerCenter*.

PowerExchange for Amazon S3

Effective in version 10.2 HotFix 1, PowerExchange for Amazon S3 includes the following new features:

- In addition to the existing regions, you can also read data from or write data to the `AWS GovCloud` region.
- You can specify the line that you want to use as the header when you read data from Amazon S3. You can specify the line number in the **Header Line Number** property under the source session properties.
- You can specify the line number from where you want the PowerCenter Integration Service to read data. You can configure the **Read Data From Line** property under the source session properties.
- You can specify an asterisk (*) wildcard in the file name to fetch files from the Amazon S3 bucket. You can specify the asterisk (*) wildcard to fetch all the files or only the files that match the name pattern.
- You can add a single or multiple tags to the objects stored on the Amazon S3 bucket to categorize the objects. Each tag contains a key value pair. You can either enter the key value pairs or specify the absolute file path that contains the key value pairs.
- You can specify the part size of an object to download the object from Amazon S3 in multiple parts.

- You can configure partitioning for Amazon S3 sources. Partitioning optimizes the mapping performance at run time when you read data from Amazon S3 sources.
- PowerExchange for Amazon S3 supports version 11 and 12 of SuSe Linux Enterprise Server operating system.

For more information, see the *Informatica PowerExchange for Amazon S3 10.2 HotFix 1 User Guide for PowerCenter*.

PowerExchange for Cassandra

Effective in version 10.2 HotFix 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 HotFix 1 User Guide for PowerCenter*.

PowerExchange for Microsoft Dynamics CRM

Effective in version 10.2 HotFix 1, PowerExchange for Microsoft Dynamics CRM includes the following new features:

- You can select either **Discovery Service** or **Organization Service** as a service type for passport authentication in the Microsoft Dynamics CRM run-time connection.
- You can configure an alternate key in update, upsert, and delete operations.
- You can specify an alternate key as a reference for Lookup, Customer, Owner, and PartyList data types.

For more information, see the *Informatica PowerExchange for Microsoft Dynamics CRM 10.2 HotFix 1 User Guide for PowerCenter*.

PowerExchange for Salesforce

Effective in version 10.2 HotFix 1, PowerExchange for Salesforce includes the following new features:

- You can use version 42.0 of Salesforce API to create a Salesforce connection and access Salesforce objects.
- You can configure OAuth for Salesforce connections.

For more information, see the*Informatica PowerExchange for Salesforce 10.2 HotFix 1 User Guide for PowerCenter*.

PowerExchange for SAP NetWeaver

Effective in version 10.2 HotFix 1, PowerExchange for SAP NetWeaver includes the following new features:

Connection resiliency support for business content integration mappings

You can configure the following connection resiliency parameters in the listener session of business content integration mappings:

- **Number of Retries for Connection Resiliency**. Defines the number of connection retries the PowerCenter Integration Service must attempt in the event of an unsuccessful connection with SAP.

- **Delay between Retries for Connection Resiliency**. Defines the time interval in seconds between the connection retries.

New SAP data types support

You can use the following new SAP data types based on the integration method you use:

Data Type	Data Integration using the ABAP Program (Table Reader and Ttable Writer)	Data Integration using BAPI/RFC Functions	IDoc Integration using ALE
INT8	Supported	Not supported	Not supported
DF16_DEC	Supported	Supported	Not supported
DF34_DEC	Supported	Supported	Not supported
DF16_DEC	Supported	Supported	Not supported
DF34_DEC	Supported	Supported	Not supported
DF16_RAW	Supported	Supported	Not supported
DF34_RAW	Supported	Supported	Not supported
RAWSTRING	Supported	Not supported	Supported
STRING	Supported	Supported	Supported
SSTRING	Supported	Supported	Not supported

For more information, see the *Informatica PowerExchange for SAP NetWeaver 10.2 HotFix 1 User Guide for PowerCenter*.

PowerExchange for Snowflake

Effective in version 10.2 HotFix 1, PowerExchange for Snowflake includes the following new features:

- You can configure a custom SQL query to configure a Snowflake source.
- You can override the database and schema name when you create temporary tables in the Snowflake staging database. You must add the database and schema name that you want to override in the JDBC URL parameters of the Snowflake connection.
- You can override the imported Snowflake source table name by specifying the table name in the Snowflake session properties. You can also configure a SQL override to override the default SQL query used to extract data from the Snowflake source.
- You can override the Snowflake target table name by specifying the table name in the Snowflake target session properties.
- You can configure source-side or full pushdown optimization to push the transformation logic to Snowflake when you use the ODBC connection type. For information about the operators and functions that the PowerCenter Integration Service can push to Snowflake, see the *Informatica PowerCenter 10.2 HotFix 1 Advanced Workflow Guide*.
- You can join multiple Snowflake source tables by specifying a join condition.
- You can configure an unconnected lookup transformation for the source qualifier in a mapping.

- You can configure pass-through partitioning for a Snowflake session. After you add the number of partitions, you can specify an SQL override or a filter override condition for each of the partitions.
- You can configure the HTTP proxy server authentication settings at design time or at run time to read data from or write data to Snowflake.
- You can configure Okta SSO authentication by specifying the authentication details in the JDBC URL parameters of the Snowflake connection.
- You can read data from and write data to Snowflake that is enabled for staging data in Azure or Amazon.

For more information, see the *Informatica PowerExchange for Snowflake 10.2 HotFix 1 User Guide for PowerCenter*.

PowerExchange for Teradata Parallel Transporter API

Effective in version 10.2 HotFix 1, when you use the Load operator, you can set the maximum buffer size in kilobytes that the Teradata PT API uses for writing data. You can set the maximum buffer size value in the Teradata target session properties.

Security

This section describes new security features in version 10.2 HotFix 1.

Security - Password Complexity

Effective in version 10.2 HotFix 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 HotFix 1 Security Guide*.

What's Changed

Support Changes

This section describes the support changes in 10.2 HotFix 1.

Big Data Management Hadoop Distributions

The following table lists the supported Hadoop distribution versions and changes in Big Data Management 10.2 HotFix 1:

Hadoop Distribution	Supported Distribution Versions	10.2 HotFix 1 Changes
Amazon EMR	5.8	Dropped support for version 5.4.
Azure HDInsight	3.5.x 3.6.x	No change.

Hadoop Distribution	Supported Distribution Versions	10.2 HotFix 1 Changes
Cloudera CDH	5.10.x 5.11.x 5.12.x 5.13.x	Dropped support for version 5.9.
Hortonworks HDP	2.5x 2.6x	Dropped support for version 2.4.
IBM BigInsights	Not supported.	Dropped support for IBM BigInsights.
MapR	5.2 MEP 3.0.x	Dropped support for version 5.2 MEP 2.0.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

Intelligent Streaming Hadoop Distributions

The following table lists the supported Hadoop distribution versions and changes in Intelligent Streaming 10.2 HotFix 1:

Distribution	Supported Versions	Changes Since 10.2
Amazon EMR	5.8	Dropped support for version 5.4.
Cloudera CDH	5.10 5.11 5.12 5.13	No change.
Hortonworks HDP	2.5.x 2.6.x	No change.
MapR	5.2 MEP 3.0	Dropped 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`.

Application Services

This section describes changes to application services in version 10.2 HotFix 1.

Model Repository Service

Monitoring Model Repository Service

Effective in version 10.2 HotFix 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 HotFix 1 Application Service Guide*.

Big Data Management

This section describes the changes to Big Data Management in version 10.2 HotFix 1.

Precision and Scale on the Hive Engine

Effective in version 10.2 HotFix 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 HotFix 1 User Guide*.

Business Glossary

This section describes changes to Business Glossary in version 10.2 HotFix 1.

Alerts and Notifications

Effective in 10.2 HotFix 1, the Analyst tool displays glossary name in addition to the names of the sender and asset submitted for review. Previously, the Analyst tool displayed only the name of the sender and name of the asset that was submitted for review.

For more information, see the "Finding Glossary Content" chapter in the *Informatica 10.2 HotFix 1 Business Glossary Guide*.

Documentation

This section describes changes to guides in Informatica documentation in version 10.2 HotFix 1.

Informatica PowerCenter 10.2 HotFix 1 Repository Guide

Effective in version 10.2 HotFix 1, the *Informatica PowerCenter 10.2 HotFix 1 Repository Guide* includes appendices MX Views Reference and PowerCenter Reports Reference. The *Informatica PowerCenter 10.2 HotFix 1 PowerCenter Reports Guide* and the *Informatica PowerCenter 10.2 HotFix 1 Using PowerCenter Reports Guide* are moved into the *Informatica PowerCenter 10.2 HotFix 1 Repository Guide* as one appendix.

For more information, see the *Informatica PowerCenter 10.2 HotFix 1 Repository Guide*.

Informatica Development Platform

This section describes changes to Informatica Development Platform in version 10.2 HotFix 1.

Informatica Connector Toolkit

Effective in version 10.2 HotFix 1, you can use the Informatica Connector Toolkit to build connector for Informatica Intelligent Cloud Services by default. If you want to publish a connector developed using Informatica Connector Toolkit versions earlier than 10.2 HotFix 1, set `YES` as the value of the `CTK_ADAPTER` system environment variable. For example, `CTK_ADAPTER =YES`.

Previously, you could use the Informatica Connector Toolkit to build connector for Informatica Cloud Services.

For more information, see the *Informatica Development Platform 10.2 HotFix 1 User Guide*.

Informatica Transformations

This section describes the changes to the Informatica transformations in version 10.2 HotFix 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:

All Countries

Effective in version 10.2 HotFix 1, the Address Validator transformation uses version 5.13.0 of the Informatica Address Verification software engine. The engine enables the features that Informatica adds to the Address Validator transformation in version 10.2 HotFix 1.

Previously, the transformation used version 5.11.0 of the Informatica Address Verification software engine.

Parcel Centroid and Rooftop Geocodes

Effective June 2018, Informatica ceases to update the reference data files that contain Parcel Centroid and Rooftop geocode data, and Informatica ceases to provide the files to first-time users. Current reference data files for Parcel Centroid and Rooftop geocodes remain operational.

For more information, see the *Informatica 10.2 HotFix 1 Developer Transformation Guide* and the *Informatica 10.2 HotFix 1 Address Validator Port Reference*.

For comprehensive information about the updates to the Informatica Address Verification software engine from version 5.11.0 through version 5.13.0, see the *Informatica Address Verification 5.13.0 Release Guide*.

PowerCenter

This section describes changes to PowerCenter in version 10.2 HotFix 1.

Microsoft Analyst for Excel

Effective in version 10.2 HotFix 1, Informatica supports Mapping Analyst for Excel with Microsoft Excel 2016. Mapping Analyst for Excel includes an Excel add-in that you can use to configure mapping specifications in Microsoft Excel 2016.

Previously, Informatica supported Mapping Analyst for Excel with Microsoft Excel 2007 and Microsoft Excel 2010.

For more information about installing the add-in for Microsoft Excel 2016, see the *Informatica PowerCenter 10.2 HotFix 1 Mapping Analyst for Excel Guide*.

PowerExchange Adapters for PowerCenter

This section describes changes to PowerCenter adapters in version 10.2 HotFix 1.

PowerExchange for Amazon Redshift

Effective in version 10.2 HotFix 1, PowerExchange for Amazon Redshift has the following changes:

- You can provide the number of files in the **Number of Files per Batch** field under the target session properties to calculate the number of the staging files for each batch.
  Previously, the number of the staging files for each batch was calculated based on the values you provided in the **Cluster Node Type** and **Number of Nodes in the Cluster** fields under the connection properties.
- The session log contains information about the individual time taken to upload data to the local staging area, upload data to Amazon S3 from the local staging area, and then upload data to an Amazon Redshift target by issuing copy command.
  Previously, the session log contained only the total time taken to write data from source to target.

For more information, see the *Informatica PowerExchange for Amazon Redshift 10.2 HotFix 1 User Guide for PowerCenter*.

PowerExchange for Cassandra

Effective in version 10.2 HotFix 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 HotFix 1 User Guide*.

PowerExchange for Google BigQuery

Effective in version 10.2 HotFix 1, PowerExchange for Google BigQuery installs with the Informatica services.

Previously, PowerExchange for PowerExchange for Google BigQuery had a separate installer.

For more information, see the *Informatica PowerExchange for Google BigQuery 10.2 HotFix 1 User Guide for PowerCenter*.

PowerExchange for Salesforce

Effective in version 10.2 HotFix 1, the error messages that you encounter during a Salesforce session might differ because of a change in the Salesforce API.

For example, when you reconnect to Salesforce the following error message appears:

```
[ERROR] Reattempt the Salesforce request [getBatchInfo] due to the error [Server error returned in unknown format].
```

Previously, for the same scenario the following error message was displayed:

```
[ERROR] Reattempt the Salesforce request [getBatchInfo] due to the error [input stream can not be null].
```

PowerExchange for Salesforce Analytics

Effective in version 10.2 HotFix 1, PowerExchange for Salesforce Analytics installs with Informatica services.

Previously, PowerExchange for Salesforce Analytics had a separate installer.

For more information, see the *Informatica PowerExchange for Salesforce Analytics 10.2 HotFix 1 User Guide for PowerCenter*

PowerExchange for Snowflake

Effective in version 10.2 HotFix 1, PowerExchange for Snowflake installs with the Informatica services.

Previously, PowerExchange for PowerExchange for Snowflake had a separate installer.

For more information, see the *Informatica PowerExchange for Snowflake 10.2 HotFix 1 User Guide for PowerCenter*.

Reference Data

This section describes changes in reference data operations in version 10.2 HotFix 1.

Content Installer

Effective Spring 2018, 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 might use the Content Installer to extract and copy the files to the Informatica directories.

For more information, see the *Informatica 10.2 HotFix 1 Content Guide*.

Properties Files in PowerCenter

Effective in version 10.2 HotFix 1, the upgrade process and the HotFix installation process maintain the contents of the reference data properties files from an earlier PowerCenter version. You do not need to edit any reference data properties file after you install the HotFix or upgrade to version 10.2 HotFix 1.

PowerCenter reads configuration information for reference data from the following properties files:

- AD50.cfg. Contains properties for address reference data.
- CLASSIFIER.properties. Contains properties for classifier models.
- IDQTx.cfg. Contains properties for identity populations.
- NER.properties. Contains properties for probabilistic models.

The HotFix installation or upgrade process writes backup versions of the properties files to the Informatica 10.2 HotFix 1 directories. The backup files are the default versions of the files and do not contain any value that you set in the earlier installation. Each backup file name ends with the extension .bak.

Previously, the upgrade operation renamed any reference data properties file that it found with the extension .bak. The upgrade operation also created default versions of any properties file that it renamed.

Note: Previously, if you installed a HotFix for an installation, so that the Informatica directory structure did not change, the installation process preserved the AD50.cfg file. The HotFix installation otherwise added the extension .bak to each reference data properties file that it found and created a default version of each file.

For more information, see the *Informatica 10.2 HotFix 1 Content Guide*.

Release Tasks

PowerExchange Adapters for Informatica

This section describes release tasks for Informatica adapters in version 10.2 HotFix 1.

PowerExchange for Netezza

Effective in version 10.2 HotFix 1, if you want to run dynamic mappings for Netezza objects, you must add the Netezza JDBC jar in the following location: `<Informatica installation directory>/externaljdbcjars`.

For more information, see the *Informatica 10.2 HotFix 1 PowerExchange for Netezza User Guide*.

PowerExchange for Teradata Parallel Transporter API

Effective in version 10.2 HotFix 1, if you want to run dynamic mappings for Teradata objects, you must add the Teradata JDBC jar in the following location: `<Informatica installation directory>/externaljdbcjars`.

For more information, see the *Informatica 10.2 HotFix 1 PowerExchange for Teradata Parallel Transporter API User Guide*.

PowerExchange Adapters for PowerCenter

This section describes release tasks for PowerCenter adapters in version 10.2 HotFix 1.

PowerExchange for Amazon Redshift

Effective in version 10.2 HotFix 1, PowerExchange for Amazon Redshift has the following release tasks:

- The **Cluster Node Type** and **Number of Nodes in the Cluster** fields are not available in the connection properties. After upgrade, PowerExchange for Amazon Redshift calculates the number of the staging files and ignores the value that you specified in the previous version for the existing mapping.
  You can specify the number of files in the **Number of Files per Batch** field under the target session properties to calculate the number of the staging files for each batch.
- The AWS SDK for Java is updated to the version 1.11.250.
- The following third party jars are updated to the latest version 2.9.5:
  - jackson-annotations
  - jackson-databind
  - jackson-core

For more information, see the *Informatica 10.2 HotFix 1 PowerExchange for Amazon Redshift User Guide for PowerCenter*.

PowerExchange for Amazon S3

Effective in version 10.2 HotFix 1, the AWS SDK for Java is updated to the version 1.11.250.

For more information, see the *Informatica 10.2 HotFix 1 PowerExchange for Amazon S3 User Guide for PowerCenter*.

PowerExchange for Teradata Parallel Transporter API

Effective in version 10.2 HotFix 1, you can define a maximum buffer size in kilobytes that the Teradata PT API uses for writing data.

When you upgrade from an earlier version, you must re-register the `TeradataPT.xml` plug-with the PowerCenter Repository Service to enable the maximum buffer size property. After you register, you can define the maximum buffer size in the Teradata target session properties.

For more information about configuring the maximum buffer size, see the *Informatica 10.2 HotFix 1 PowerExchange for Teradata Parallel Transporter API User Guide for PowerCenter*.

CHAPTER 34

10.2 Important Notices

This chapter includes the following topics:

- New Products, 352
- Release Tasks, 352

New Products

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*.

Release Tasks

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.

CHAPTER 35

10.2 What's New

This chapter includes the following topics:

- Application Services, 355
- Big Data , 356
- Command Line Programs, 359
- Data Types, 367
- Documentation, 368
- Enterprise Information Catalog, 369
- Informatica Analyst, 372
- Intelligent Data Lake, 373
- Informatica Developer, 375
- Informatica Installation, 375
- Intelligent Streaming, 375
- Metadata Manager, 377
- PowerCenter, 377
- PowerExchange Adapters, 378
- Rule Specifications, 382
- Security, 382
- Transformation Language, 383
- Transformations, 384
- Workflows, 388

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
UpdateDomainSamlConfig	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.
UpdateGatewayNode	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 36

10.2 What's Changed

This chapter includes the following topics:

- Support Changes, 390
- Application Services, 393
- Big Data, 394
- Command Line Programs, 400
- Enterprise Information Catalog, 401
- Informatica Analyst, 401
- Intelligent Streaming, 401
- PowerExchange Adapters, 402
- Security, 404
- Transformations, 404
- Workflows, 405

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	5.8	Added support for version 5.8. Dropped support for versions 5.0 and 5.4. **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	5.10.x 5.11.x 5.12.x 5.13.x	Added support for versions 5.12, 5.13 Dropped support for version 5.8 and 5.9.

Hadoop Distribution	Supported Distribution Versions	10.2 Changes
Hortonworks HDP	2.5x 2.6x	Dropped support for versions 2.3 and 2.4. **Note:** To use Hortonworks 2.5 with Big Data Management 10.2, you must apply a Emergency Bug Fix patch. See the following Knowledge Base article: - Hortonworks 2.5 support: KB 521847.
MapR	5.2 MEP 3.0.x	Added support for version 5.2 MEP 3.0. Dropped support for version 5.2 MEP 1.x and 5.2 MEP 2.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`.

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.
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.

Hadoop Distribution	Supported Distribution Versions	Changes since 10.1.1 HotFix1
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.
Blaze User Name	Blaze Service User Name CadiUserName	The owner of the Blaze service and Blaze service logs. Appears in Blaze Configuration.

Current Name	Previous Name	Description
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.

Property	Description
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*.