Informatica®
10.2

# New Features Guide

Informatica New Features Guide
10.2
August 2017

© Copyright Informatica LLC 2009, 2019

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.

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

Portions of this software and/or documentation are subject to copyright held by third parties, including without limitation: Copyright DataDirect Technologies. All rights reserved. Copyright © Sun Microsystems. All rights reserved. Copyright © RSA Security Inc. All Rights Reserved. Copyright © Ordinal Technology Corp. All rights reserved. Copyright © Aandacht c.v. All rights reserved. Copyright Genivia, Inc. All rights reserved. Copyright Isomorphic Software. All rights reserved. Copyright © Meta Integration Technology, Inc. All rights reserved. Copyright © Intalio. All rights reserved. Copyright © Oracle. All rights reserved. Copyright © Adobe Systems Incorporated. All rights reserved. Copyright © DataArt, Inc. All rights reserved. Copyright © ComponentSource. All rights reserved. Copyright © Microsoft Corporation. All rights reserved. Copyright © Rogue Wave Software, Inc. All rights reserved. Copyright © Teradata Corporation. All rights reserved. Copyright © Yahoo! Inc. All rights reserved. Copyright © Glyph & Cog, LLC. All rights reserved. Copyright © Thinkmap, Inc. All rights reserved. Copyright © Clearpace Software Limited. All rights reserved. Copyright © Information Builders, Inc. All rights reserved. Copyright © OSS Nokalva, Inc. All rights reserved. Copyright Edifecs, Inc. All rights reserved. Copyright Cleo Communications, Inc. All rights reserved. Copyright © International Organization for Standardization 1986. All rights reserved. Copyright © ej-technologies GmbH. All rights reserved. Copyright © Jaspersoft Corporation. All rights reserved. Copyright © International Business Machines Corporation. All rights reserved. Copyright © yWorks GmbH. All rights reserved. Copyright © Lucent Technologies. All rights reserved. Copyright © University of Toronto. All rights reserved. Copyright © Daniel Veillard. All rights reserved. Copyright © Unicode, Inc. Copyright IBM Corp. All rights reserved. Copyright © MicroQuill Software Publishing, Inc. All rights reserved. Copyright © PassMark Software Pty Ltd. All rights reserved. Copyright © LogiXML, Inc. All rights reserved. Copyright © 2003-2010 Lorenzi Davide, All rights reserved. Copyright © Red Hat, Inc. All rights reserved. Copyright © The Board of Trustees of the Leland Stanford Junior University. All rights reserved. Copyright © EMC Corporation. All rights reserved. Copyright © Flexera Software. All rights reserved. Copyright © Jinfonet Software. All rights reserved. Copyright © Apple Inc. All rights reserved. Copyright © Telerik Inc. All rights reserved. Copyright © BEA Systems. All rights reserved. Copyright © PDFlib GmbH. All rights reserved. Copyright © Orientation in Objects GmbH. All rights reserved. Copyright © Tanuki Software, Ltd. All rights reserved. Copyright © Ricebridge. All rights reserved. Copyright © Sencha, Inc. All rights reserved. Copyright © Scalable Systems, Inc. All rights reserved. Copyright © jQWidgets. All rights reserved. Copyright © Tableau Software, Inc. All rights reserved. Copyright© MaxMind, Inc. All Rights Reserved. Copyright © TMate Software s.r.o. All rights reserved. Copyright © MapR Technologies Inc. All rights reserved. Copyright © Amazon Corporate LLC. All rights reserved. Copyright © Highsoft. All rights reserved. Copyright © Python Software Foundation. All rights reserved. Copyright © BeOpen.com. All rights reserved. Copyright © CNRI. All rights reserved.

This product includes software developed by the Apache Software Foundation (http://www.apache.org/), and/or other software which is licensed under various versions of the Apache License (the "License"). You may obtain a copy of these Licenses at http://www.apache.org/licenses/. Unless required by applicable law or agreed to in writing, software distributed under these Licenses is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the Licenses for the specific language governing permissions and limitations under the Licenses.

This product includes software which was developed by Mozilla (http://www.mozilla.org/), software copyright The JBoss Group, LLC, all rights reserved; software copyright © 1999-2006 by Bruno Lowagie and Paulo Soares and other software which is licensed under various versions of the GNU Lesser General Public License Agreement, which may be found at http:// www.gnu.org/licenses/lgpl.html. The materials are provided free of charge by Informatica, "as-is", without warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability and fitness for a particular purpose.

The product includes ACE(TM) and TAO(TM) software copyrighted by Douglas C. Schmidt and his research group at Washington University, University of California, Irvine, and Vanderbilt University, Copyright (©) 1993-2006, all rights reserved.

This product includes software developed by the OpenSSL Project for use in the OpenSSL Toolkit (copyright The OpenSSL Project. All Rights Reserved) and redistribution of this software is subject to terms available at http://www.openssl.org and http://www.openssl.org/source/license.html.

This product includes Curl software which is Copyright 1996-2013, Daniel Stenberg, <daniel@haxx.se>. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at http://curl.haxx.se/docs/copyright.html. Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

The product includes software copyright 2001-2005 (©) MetaStuff, Ltd. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at http://www.dom4j.org/ license.html.

The product includes software copyright © 2004-2007, The Dojo Foundation. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at http://dojotoolkit.org/license.

This product includes ICU software which is copyright International Business Machines Corporation and others. All rights reserved. Permissions and limitations regarding this software are subject to terms available at http://source.icu-project.org/repos/icu/icu/trunk/license.html.

This product includes software copyright © 1996-2006 Per Bothner. All rights reserved. Your right to use such materials is set forth in the license which may be found at http:// www.gnu.org/software/ kawa/Software-License.html.

This product includes OSSP UUID software which is Copyright © 2002 Ralf S. Engelschall, Copyright © 2002 The OSSP Project Copyright © 2002 Cable & Wireless Deutschland. Permissions and limitations regarding this software are subject to terms available at http://www.opensource.org/licenses/mit-license.php.

This product includes software developed by Boost (http://www.boost.org/) or under the Boost software license. Permissions and limitations regarding this software are subject to terms available at http:/ /www.boost.org/LICENSE_1_0.txt.

This product includes software copyright © 1997-2007 University of Cambridge. Permissions and limitations regarding this software are subject to terms available at http:// www.pcre.org/license.txt.

This product includes software copyright © 2007 The Eclipse Foundation. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at http:// www.eclipse.org/org/documents/epl-v10.php and at http://www.eclipse.org/org/documents/edl-v10.php.

This product includes software licensed under the terms at http://www.tcl.tk/software/tcltk/license.html, http://www.bosrup.com/web/overlib/?License, http://www.stlport.org/doc/ license.html, http://asm.ow2.org/license.html, http://www.cryptix.org/LICENSE.TXT, http://hsqldb.org/web/hsqlLicense.html, http://httpunit.sourceforge.net/doc/ license.html, http://jung.sourceforge.net/license.txt , http://www.gzip.org/zlib/zlib_license.html, http://www.openldap.org/software/release/license.html, http://www.libssh2.org, http://slf4j.org/license.html, http://www.sente.ch/software/OpenSourceLicense.html, http://fusesource.com/downloads/license-agreements/fuse-message-broker-v-5-3- license-agreement; http://antlr.org/license.html; http://aopalliance.sourceforge.net/; http://www.bouncycastle.org/licence.html; http://www.jgraph.com/jgraphdownload.html; http://www.jcraft.com/jsch/LICENSE.txt; http://jotm.objectweb.org/bsd_license.html; . http://www.w3.org/Consortium/Legal/2002/copyright-software-20021231; http://www.slf4j.org/license.html; http://nanoxml.sourceforge.net/orig/copyright.html; http://www.json.org/license.html; http://forge.ow2.org/projects/javaservice/, http://www.postgresql.org/about/licence.html, http://www.sqlite.org/copyright.html, http://www.tcl.tk/software/tcltk/license.html, http://www.jaxen.org/faq.html, http://www.jdom.org/docs/faq.html, http://www.slf4j.org/license.html; http://www.iodbc.org/dataspace/iodbc/wiki/iODBC/License; http://www.keplerproject.org/md5/license.html; http://www.toedter.com/en/jcalendar/license.html; http://www.edankert.com/bounce/index.html; http://www.net-snmp.org/about/license.html; http://www.openmdx.org/#FAQ; http://www.php.net/license/3_01.txt; http://srp.stanford.edu/license.txt;

http://www.schneier.com/blowfish.html; http://www.jmock.org/license.html; http://xsom.java.net; http://benalman.com/about/license/; https://github.com/CreateJS/EaselJS/blob/master/src/easeljs/display/Bitmap.js; http://www.h2database.com/html/license.html#summary; http://jsoncpp.sourceforge.net/LICENSE; http://jdbc.postgresql.org/license.html; http://protobuf.googlecode.com/svn/trunk/src/google/protobuf/descriptor.proto; https://github.com/rantav/hector/blob/master/LICENSE; http://web.mit.edu/Kerberos/krb5-current/doc/mitK5license.html; http://jibx.sourceforge.net/jibx-license.html; https://github.com/lyokato/libgeohash/blob/master/LICENSE; https://github.com/hjiang/jsonxx/blob/master/LICENSE; https://code.google.com/p/lz4/; https://github.com/jedisct1/libsodium/blob/master/LICENSE; http://one-jar.sourceforge.net/index.php?page=documents&file=license; https://github.com/EsotericSoftware/kryo/blob/master/license.txt; http://www.scala-lang.org/license.html; https://github.com/tinkerpop/blueprints/blob/master/LICENSE.txt; http://gee.cs.oswego.edu/dl/classes/EDU/oswego/cs/dl/util/concurrent/intro.html; https://aws.amazon.com/asl/; https://github.com/twbs/bootstrap/blob/master/LICENSE; https://sourceforge.net/p/xmlunit/code/HEAD/tree/trunk/LICENSE.txt; https://github.com/documentcloud/underscore-contrib/blob/master/LICENSE, and https://github.com/apache/hbase/blob/master/LICENSE.txt.

This product includes software licensed under the Academic Free License (http://www.opensource.org/licenses/afl-3.0.php), the Common Development and Distribution License (http://www.opensource.org/licenses/cddl1.php) the Common Public License (http://www.opensource.org/licenses/cpl1.0.php), the Sun Binary Code License Agreement Supplemental License Terms, the BSD License (http:// www.opensource.org/licenses/bsd-license.php), the new BSD License (http://opensource.org/licenses/BSD-3-Clause), the MIT License (http://www.opensource.org/licenses/mit-license.php), the Artistic License (http://www.opensource.org/licenses/artistic-license-1.0) and the Initial Developer's Public License Version 1.0 (http://www.firebirdsql.org/en/initial-developer-s-public-license-version-1-0/).

This product includes software copyright © 2003-2006 Joe Walnes, 2006-2007 XStream Committers. All rights reserved. Permissions and limitations regarding this software are subject to terms available at http://xstream.codehaus.org/license.html. This product includes software developed by the Indiana University Extreme! Lab. For further information please visit http://www.extreme.indiana.edu/.

This product includes software Copyright (c) 2013 Frank Balluffi and Markus Moeller. All rights reserved. Permissions and limitations regarding this software are subject to terms of the MIT license.

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

DISCLAIMER: Informatica LLC provides this documentation "as is" without warranty of any kind, either express or implied, including, but not limited to, the implied warranties of noninfringement, merchantability, or use for a particular purpose. Informatica LLC does not warrant that this software or documentation is error free. The information provided in this software or documentation may include technical inaccuracies or typographical errors. The information in this software and documentation is subject to change at any time without notice.

NOTICES

This Informatica product (the "Software") includes certain drivers (the "DataDirect Drivers") from DataDirect Technologies, an operating company of Progress Software Corporation ("DataDirect") which are subject to the following terms and conditions:

1. THE DATADIRECT DRIVERS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT.
2. IN NO EVENT WILL DATADIRECT OR ITS THIRD PARTY SUPPLIERS BE LIABLE TO THE END-USER CUSTOMER FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, CONSEQUENTIAL OR OTHER DAMAGES ARISING OUT OF THE USE OF THE ODBC DRIVERS, WHETHER OR NOT INFORMED OF THE POSSIBILITIES OF DAMAGES IN ADVANCE. THESE LIMITATIONS APPLY TO ALL CAUSES OF ACTION, INCLUDING, WITHOUT LIMITATION, BREACH OF CONTRACT, BREACH OF WARRANTY, NEGLIGENCE, STRICT LIABILITY, MISREPRESENTATION AND OTHER TORTS.

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: 2019-04-09

# Table of Contents

# Preface

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

# Informatica Resources

## Informatica Network

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

As a member, you can:

- Access all of your Informatica resources in one place.
- Search the Knowledge Base for product resources, including documentation, FAQs, and best practices.
- View product availability information.
- Review your support cases.
- Find your local Informatica User Group Network and collaborate with your peers.

## Informatica Knowledge Base

Use the Informatica Knowledge Base to search Informatica Network for product resources such as documentation, how-to articles, best practices, and PAMs.

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

## Informatica Documentation

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

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

## Informatica Product Availability Matrixes

Product Availability Matrixes (PAMs) indicate the versions of operating systems, databases, and other types of data sources and targets that a product release supports. If you are an Informatica Network member, you can access PAMs at https://network.informatica.com/community/informatica-network/product-availability-matrices.

## Informatica Velocity

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

If you are an Informatica Network member, you can access Informatica Velocity resources at http://velocity.informatica.com.

If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at ips@informatica.com.

## Informatica Marketplace

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

## Informatica Global Customer Support

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

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

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

CHAPTER 1

# New Features (10.2)

This chapter includes the following topics:



## Application Services

This section describes new application service features in 10.2.

### Model Repository Service

This section describes new Model Repository Service features in 10.2.

### Import Objects from Previous Versions

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

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

# Big Data

This section describes new big data features in 10.2.

## Big Data Management Installation

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

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

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

## Cluster Configuration

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

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

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

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

## Processing Hierarchical Data

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

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

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

- Convert data from one complex file format to another.

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

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

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

## Stateful Computing on the Spark Engine

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

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

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

## Data Integration Service Queuing

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

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

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

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

## Blaze Job Monitor

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

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

## Data Integration Service Properties for Hadoop Integration

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

The following table describes the new properties:

| Property | Description |
|---|---|
| Hadoop Staging Directory | The HDFS directory where the Data Integration Services pushes Informatica Hadoop binaries and stores temporary files during processing. Default is /tmp. |
| Hadoop Staging User | Required if the Data Integration Service user is empty. The HDFS user that performs operations on the Hadoop staging directory. The user needs write permissions on Hadoop staging directory. Default is the Data Integration Service user. |
| Custom Hadoop OS Path | The local path to the Informatica Hadoop binaries compatible with the Hadoop operating system. Required when the Hadoop cluster and the Data Integration Service are on different supported operating systems.<br>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.<br>The Data Integration Service can synchronize the following operating systems:<br>- SUSE 11 and Redhat 6.5<br>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.<br>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:<br>-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.<br>Added, changed, and removed Hadoop connection options. See infacmd isp createConnection. |
| getDomainSamlConfig | Renamed from getSamlConfig.<br>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.<br>The user activity data includes the following properties for each login attempt from an Informatica client:<br>- Application name<br>- Application version<br>- Host name or IP address of the application host<br>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.<br>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.<br>The -lu option is now obsolete. |
| SwitchToGatewayNode | The following options are added for configuring SAML authentication:<br>- asca. The alias name specified when importing the identity provider assertion signing certificate into the truststore file used for SAML authentication.<br>- saml. Enabled or disabled SAML authentication in the Informatica domain.<br>- std. The directory containing the custom truststore file required to use SAML authentication on gateway nodes within the domain.<br>- 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.<br>Imported into the cluster configuration as the property `yarn.resourcemanager.address`. |
| defaultFSURI | The URI to access the default Hadoop Distributed File System.<br>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.<br>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.<br>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:<br>- asca. The alias name specified when importing the identity provider assertion signing certificate into the truststore file used for SAML authentication.<br>- 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.<br>- std. The directory containing the custom truststore file required to use SAML authentication on gateway nodes within the domain.<br>- stp. The custom truststore password used for SAML authentication. |
| DefineGatewayNode | The following options are added for configuring SAML authentication:<br>- asca. The alias name specified when importing the identity provider assertion signing certificate into the truststore file used for SAML authentication.<br>- saml. Enables or disables SAML authentication in the Informatica domain.<br>- std. The directory containing the custom truststore file required to use SAML authentication on gateway nodes within the domain.<br>- stp. The custom truststore password used for SAML authentication. |

| Command | Description |
|---|---|
| UpdateDomainSamlConfig | Renamed from UpdateSamlConfig.<br>The following option is added for configuring SAML authentication:<br>- 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.<br>- asca. The alias name specified when importing the identity provider assertion signing certificate into the truststore file used for SAML authentication.<br>- saml. Enables or disables SAML authentication in the Informatica domain.<br>- std. The directory containing the custom truststore file required to use SAML authentication on gateway nodes within the domain.<br>- 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:<br>-w. Enables you to use a parameter in the password option. |
| ListObjectDependencies | Contains the following updated option:<br>-o. The object type list includes query and deploymentgroup. |
| UpdateConnection | Contains the following updated options:<br>-w. Enables you to use a parameter in the password option.<br>-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*.