



Informatica® Data Quality Integration for  
PowerCenter

10.1

# User Guide

© Copyright Informatica LLC 2009, 2018

This software and documentation contain proprietary information of Informatica LLC and are provided under a license agreement containing restrictions on use and disclosure and are also protected by copyright law. Reverse engineering of the software is prohibited. 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. This Software may be protected by U.S. and/or international Patents and other Patents Pending.

Use, duplication, or disclosure of the Software by the U.S. Government is subject to the restrictions set forth in the applicable software license agreement and as provided in DFARS 227.7202-1(a) and 227.7702-3(a) (1995), DFARS 252.227-7013(1)(ii) (OCT 1988), FAR 12.212(a) (1995), FAR 52.227-19, or FAR 52.227-14 (ALT III), as applicable.

The information in this product or documentation is subject to change without notice. If you find any problems in this product or documentation, please report them to us in writing.

Informatica, Informatica Platform, Informatica Data Services, PowerCenter, PowerCenterRT, PowerCenter Connect, PowerCenter Data Analyzer, PowerExchange, PowerMart, Metadata Manager, Informatica Data Quality, Informatica Data Explorer, Informatica B2B Data Transformation, Informatica B2B Data Exchange Informatica On Demand, Informatica Identity Resolution, Informatica Application Information Lifecycle Management, Informatica Complex Event Processing, Ultra Messaging, Informatica Master Data Management, and Live Data Map are trademarks or registered trademarks of Informatica LLC in the United States and in jurisdictions throughout the world. All 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 (c) 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 © TMat 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](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/hsqLicense.html>, <http://httpunit.sourceforge.net/doc/license.html>, <http://jung.sourceforge.net/license.txt>, [http://www.gzip.org/zlib/zlib\\_license.html](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://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://www.php.net/license/3_01.txt), <http://srp.stanford.edu/license.txt>, <http://www.schneider.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/ssl/>, <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.

Publication Date: 2018-07-03

# Table of Contents

<b>Preface .....</b>	<b>5</b>
Informatica Resources. ....	5
Informatica Network. ....	5
Informatica Knowledge Base. ....	5
Informatica Documentation. ....	5
Informatica Product Availability Matrixes. ....	6
Informatica Velocity. ....	6
Informatica Marketplace. ....	6
Informatica Global Customer Support. ....	6
 <b>Chapter 1: Informatica Data Quality Integration.....</b>	 <b>7</b>
Data Quality Integration Overview. ....	7
Data Quality Features and Functionality. ....	8
Integration Installers. ....	8
Object Export to PowerCenter. ....	9
Reference Data Installation. ....	9
Integration Matrix. ....	9
Rules and Guidelines for Product Integration. ....	10
Data Quality Transformation Behavior in PowerCenter. ....	10
 <b>Chapter 2: Installation.....</b>	 <b>14</b>
Installation Overview. ....	14
System Requirements. ....	14
Before You Install. ....	15
Installing the Server Plug-in. ....	15
Installing on Windows. ....	15
Installing on UNIX. ....	16
Installing the Client Plug-in. ....	16
Registering the Plug-In with PowerCenter. ....	17
 <b>Chapter 3: Reference Data Requirements.....</b>	 <b>18</b>
Reference Data Overview. ....	18
Reference Data File Requirements. ....	19
Address Reference Data Requirements. ....	20
Identity Population Data Requirements. ....	21
Probabilistic Model and Classifier Model Requirements. ....	22
Probabilistic Model and Classifier Model File Locations. ....	23
PowerCenter Integration Service Memory Allocation. ....	23

# Preface

PowerCenter users can work with mappings and reference data objects that they import from an Informatica Data Quality environment. The *Informatica Data Quality Integration for PowerCenter User Guide* is written for the PowerCenter users who run data quality mappings and manage the associated reference data objects.

## 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](mailto: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](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](mailto: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](mailto: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

## Informatica Data Quality Integration

This chapter includes the following topics:

- [Data Quality Integration Overview, 7](#)
- [Data Quality Features and Functionality, 8](#)
- [Integration Installers, 8](#)
- [Object Export to PowerCenter, 9](#)
- [Reference Data Installation, 9](#)
- [Integration Matrix, 9](#)
- [Rules and Guidelines for Product Integration, 10](#)
- [Data Quality Transformation Behavior in PowerCenter, 10](#)

### Data Quality Integration Overview

Informatica users can create and run mappings in Informatica Data Quality. The users save the mappings and other objects to the Model repository. The users can export the objects for use in PowerCenter. As a PowerCenter user, you can import the objects to the PowerCenter repository and you can add the objects to sessions.

Use the objects to analyze and enhance the content and structure of your data.

You can use the objects that the users created in Data Quality 10.1 with the following versions of PowerCenter:

- PowerCenter 10.1.
- PowerCenter 9.6.1 with the latest hotfix installed.
- PowerCenter 9.5.1 with the latest hotfix installed.

If you use the current versions of Data Quality and PowerCenter, you can import objects that originate in the Model repository without any additional installation. If you use an earlier version of PowerCenter, install and register the current version of the Informatica Data Quality Integration plug-in.

# Data Quality Features and Functionality

Informatica Data Quality uses Informatica applications to create objects, save objects to the Model repository, and run mappings.

You may need to install the Informatica Data Quality Plug-in for PowerCenter if you do not use the current product versions.

Use Informatica Data Quality to design and run mappings that achieve the following objectives:

- Profile data. Profiling reveals the content and structure of your data. Profiling is a key step in any data project as it can identify strengths and weaknesses in your data and help you define your project plan.
- Create scorecards to review data quality. A scorecard is a graphical representation of the quality measurements in a profile.
- Standardize data values. Standardize data to remove errors and inconsistencies that you find when you run a profile. You can standardize variations in punctuation, formatting, and spelling. For example, you can ensure that the city, state, and ZIP code values are consistent.
- Parse records. Parse data records to improve record structure and derive additional information from your data. You can split a single field of freeform data into fields that contain different information types. You can also add information to your records. For example, you can flag customer records as personal or business customers.
- Validate postal addresses. Address validation evaluates and enhances the accuracy and deliverability of your postal address data. Address validation corrects errors in addresses and completes partial addresses by comparing address records against reference data from national postal carriers. Address validation can also add postal information that speeds mail delivery and reduces mail costs.
- Find duplicate records. Duplicate record analysis compares a set of records against each other to find similar or matching values in selected data columns. You set the level of similarity that indicates a good match between field values. You can also set the relative weight fixed to each column in match calculations. For example, you can prioritize surname information over forename information.
- Create and run data quality rules. Informatica provides pre-built rules that you can run or edit to suit your project objectives. You can create rules in the Developer tool.
- Collaborate with Informatica users. The rules and reference data tables you add to the Model repository are available to users in the Developer tool and the Analyst tool. Users can collaborate on projects, and different users can take ownership of objects at different stages of a project.
- Export mappings to PowerCenter. You can export mappings to PowerCenter to reuse the metadata for physical data integration or to create web services.

## Integration Installers

If you use different versions of Data Quality and PowerCenter to create and store objects, install the Informatica Data Quality Plug-in for PowerCenter. Download the plug-in installers from Informatica.

Install the plug-in version for the version of PowerCenter that you want to use. The plug-in has Client and Server installers.

Run the Client installer on the PowerCenter client machine to view mapplets and mappings in the repository and add them to sessions. Run the Server installer on the PowerCenter Integration Service machine to run the sessions in workflows.



After you install the plug-in on the Integration Service machine, register the plug-in with the PowerCenter repository.

## Object Export to PowerCenter

Use the Developer tool to export objects such as mapplets and mappings from the Informatica Model repository. You can export objects from the Model repository to the PowerCenter repository, and you can export objects from the Model repository to the file system. Use the PowerCenter Repository Manager to import XML objects to the PowerCenter repository.

- For information about exporting objects from the Model repository, see the *Informatica Developer Mapping Guide* and the *Informatica Developer Tool Guide*.
- For information about importing objects to the PowerCenter repository, see the *PowerCenter Repository Guide*.

## Reference Data Installation

If a maplet or mapping reads reference data files, install the reference data files on the PowerCenter Integration Service machine.

You can use the Developer tool to export reference data from the Model repository when you export a maplet or mapping. The Developer exports the reference data as a compressed file. You extract the file and copy the reference data to the Integration Service machine.

You can also use the Data Quality Content Installer to add reference data to the Integration Service machine. Run the Content Installer to install address reference data files and identity population data files.

## Integration Matrix

The level of integration between Data Quality and PowerCenter depends on the respective product versions. You can export objects and reference data from multiple versions of Data Quality to multiple versions of PowerCenter. You cannot export objects or reference data from PowerCenter to Data Quality.

The following table shows the product versions that you can integrate:

Product Name	PowerCenter 10.1	PowerCenter 10.0	PowerCenter 9.6.1	PowerCenter 9.5.1
Data Quality 10.1	Yes	No	Install plug-ins	Install plug-ins
Data Quality 10.0	No	Yes	Install plug-ins	Install plug-ins
Data Quality 9.6.1	No	No	Yes	Install plug-ins
Data Quality 9.5.1	No	No	No	Yes

The table values represent the following types of integration:

### **Yes**

You can export objects and files from a Data Quality environment to a PowerCenter environment. You do not need to install or register a plug-in. You can run data quality objects in a PowerCenter repository that you upgrade to the specified version.

### **Install plug-ins**

You can export objects and files from a Data Quality environment to a PowerCenter environment. Install and register the client and server plug-ins for the version of PowerCenter that you use. You can run data quality objects in a PowerCenter repository that you upgrade to the specified version.

### **No**

You cannot use the objects and files that you define in a Data Quality environment to a PowerCenter environment.

## Rules and Guidelines for Product Integration

You can export data quality objects from different versions of Informatica Data Quality to different versions of PowerCenter. Install the Integration plug-in for the version of Informatica Data Quality from which you export the objects.

Consider the following rules and guidelines when you integrate Informatica Data Quality and PowerCenter:

- Do not install the Integration plug-in if the PowerCenter repository product version matches the Model repository product version.
- You cannot run two Data Quality Integration plug-ins concurrently on a single PowerCenter machine. PowerCenter uses the plug-in that you installed most recently. For example, if you install the Data Quality 9.6.1 Integration plug-in on a PowerCenter 9.5.1 machine, you can no longer import and run data quality objects from a version 9.5.1 Model repository. You can continue to run any data quality object that you imported to the PowerCenter repository before you installed the 9.6.1 plug-in.
- You cannot edit data quality mapplets or mappings in a PowerCenter repository.

## Data Quality Transformation Behavior in PowerCenter

When you import a data quality mapplet or mapping from the Model repository to the PowerCenter repository, the data quality transformation behavior can change.

### **Consolidation Transformation Considerations**

Informatica Data Quality updated the default strategy on the Consolidation transformation in version 9.5.1. In version 9.5.1 and later versions, the Consolidation transformation selects the record with the highest row ID as the survivor record. In version 9.5 and earlier versions, the Consolidation transformation uses the most frequent nonblank value in a column that you specify to identify the survivor record.

If you upgrade a PowerCenter repository that contains a Consolidation transformation to version 9.5.1 or later, the upgrade maintains the most frequent nonblank value as the default setting. If you import a mapping that contains a Consolidation transformation with default settings from a 9.5.1 Model repository or a later repository, the transformation uses the highest row ID to identify the survivor record.

## Conversion to Mapplets

The following transformations can convert to mapplets on import to the PowerCenter repository:

- Consolidation
- Exception
- Key Generator
- Match

When you import a mapping that contains one or more of the transformations, PowerCenter converts each transformation to a mapplet that contains the expanded configuration.

When you import a mapplet that contains one or more of the transformations, PowerCenter expands the configuration of each transformation into multiple transformations.

PowerCenter might add transformations to the data quality transformation or mapplet expansion to complete the transformation logic. For example, PowerCenter expands the Match transformation with additional transformations.

## Decision Transformation Considerations

You cannot export a mapping or a mapplet that contains a Decision transformation to PowerCenter if the transformation script includes a system parameter. The export operation cannot convert the system parameter to a value that PowerCenter can use. Before you export a mapping or a mapplet with a Decision transformation that uses a system parameter, replace the parameter with an appropriate value.

## Exception Transformation Considerations

PowerCenter cannot use an Exception transformation that originates in a PowerCenter repository that is older than version 9.5. If you upgrade a PowerCenter repository to version 9.5 or later and the repository contains a mapping with an Exception transformation, export the mapping again. Create the mapping again in the Developer tool if required.

When you run a mapping that includes an Exception transformation, you can update the bad record or duplicate record output tables in the Analyst tool. If you plan to update a bad record table in the Analyst tool, you must enable high precision on the session that runs the mapping.

## Match Transformation Considerations

You cannot run an identity match mapping in PowerCenter if the mapping reads or writes identity index data to database tables. To perform identity match analysis in PowerCenter, configure the Match transformation in the Developer tool to write identity index data to temporary files.

## Multi-Strategy Transformations

You can define multiple data transformation strategies within each of the following transformations in the Developer tool:

- Case
- Classifier
- Decision
- Key Generator
- Labeler
- Match
- Merge
- Parser

- Standardizer

When you import a mapplet or mapping that contains a multi-strategy transformation, each strategy converts to a separate transformation.

## Partition Settings

Before you run a session task that contains a data quality mapping on a grid, identify the PowerCenter version in which you created the grid nodes. If you created the nodes in PowerCenter 9.5.1 HotFix 2 or in an earlier version of PowerCenter, the mapping might fail.

To verify that data quality mappings can run on all nodes on the grid, perform one of the following tasks:

- Review the advanced properties on any PowerCenter Integration Service that distributes tasks to nodes on the grid. Configure each PowerCenter Integration Service to ignore resource requirements when it distributes tasks. By default, the advanced properties specify that the PowerCenter Integration Service ignores resource requirements.
- Re-create any node that you created in PowerCenter 9.5.1 HotFix 2 or in an earlier PowerCenter version. When you re-create the node, PowerCenter applies the current node configuration files to the node.

When you run a session task on a grid, the Data Transformation Manager process (DTM) distributes the session threads across the nodes of the grid. Some transformations are not partitionable across a grid. When a transformation is not partitionable across a grid, the DTM creates a single partition group for the transformation threads and runs the threads on a single node.

The following table describes the partition settings for data quality transformations:

Transformation	Partition Setting
Address Validation	Across grid
Association	Not partitionable
AV Report	Locally partitionable
Case Converter	Across grid
Classifier	Locally partitionable
Cluster	Not partitionable
Comparison	Across grid
Consolidation	Not partitionable
Decision	Across grid
Exception	Not partitionable
Execution Point	Across grid
Key Generator	Not partitionable
Key Store	Not partitionable
Labeler	Across grid
Merge	Across grid

Transformation	Partition Setting
Pair Generator	Not partitionable
Parser (probabilistic analysis)	Locally partitionable
Parser	Across grid
Token Standardizer	Across grid
Weight Based Analyzer	Across grid

## CHAPTER 2

# Installation

This chapter includes the following topics:

- [Installation Overview, 14](#)
- [System Requirements, 14](#)
- [Before You Install, 15](#)
- [Installing the Server Plug-in, 15](#)
- [Installing the Client Plug-in, 16](#)
- [Registering the Plug-In with PowerCenter, 17](#)

## Installation Overview

The Data Quality Integration plug-in has a server component and a client component.

Complete the following steps to install and register the plug-in:

1. Run the Server installer on the PowerCenter Integration Service machine.
2. Run the Client installer on the PowerCenter client machine.
3. Register the `IDQZIntegration.xml` file with the PowerCenter repository.

## System Requirements

The Informatica Data Quality Integration plug-in has the same system requirements as the PowerCenter applications and services to which it installs.

For information about PowerCenter prerequisites and system requirements, consult the PowerCenter installation documentation.

# Before You Install

Consider the following installer behavior if you use reference data in PowerCenter.

The Integration plug-in installer writes property files to the PowerCenter Integration Service machine. The property files specify the locations of the reference data files that data quality mappings can read.

The installer writes the following files:

- AD50.cfg. Stores configuration properties for address reference data.
- CLASSIFIER.properties. Stores configuration properties for classifier model data.
- IDQTx.cfg. Stores configuration properties for identity population data.
- NER.properties. Stores configuration properties for probabilistic model data.

The installer adds the files to the following directory:

```
[PowerCenter_installation]/server/bin
```

If the installer finds a property file in the `bin` directory, it creates a backup copy of the file and installs a new file. To keep the current reference data configuration after you install, merge the contents of the old and new files.

If you install PowerCenter 9.5.1 or a later version, the server installation process adds the property files to the `bin` directory by default.

## Address Reference Data Prerequisite

Before you install address reference data, stop the PowerCenter Integration Service. Restart the service after you install the data.

# Installing the Server Plug-in

The Informatica Data Quality Integration Server plug-in installer adds data quality transformation application files to the PowerCenter Integration Service directory structure.

## Installing on Windows

Run the Windows Server installer to add the plug-in to a PowerCenter Integration Service machine on a Windows platform.

Close all applications before installing.

1. Log in to the machine with a system user account.
2. Extract the Integration Server file for Windows.
3. Browse to the directory where you extracted the files and run `install.exe`.
4. Select a language and click **OK**.
5. On the **Welcome** screen, click **Next**.
6. Verify that PowerCenter services are not running and click **OK**.
7. Specify the path to the PowerCenter installation directory. The installer provides a default path to this directory. To select another path, click **Choose**.
8. Click **Next**.

9. Review the pre-installation summary. Click **Previous** to fix any error.
10. Click **Install**.
11. Review the post-installation summary.  
For more information about installation tasks and to view configuration properties for the installed components, view the installation log files.
12. Click **Done**.

## Installing on UNIX

Run the UNIX Server installer to add the plug-in to a PowerCenter Integration Service machine on a UNIX platform.

1. Log in to the machine with a system user account.
2. Close all other applications.
3. Extract the Integration Server file for UNIX.
4. Open a UNIX shell and navigate to the directory where you extracted the installation files.
5. Enter `sh install.bin`.
6. To accept English as the installation language, press **Enter**.
7. Read the Welcome text and press **Enter**.
8. Type an absolute path to the PowerCenter installation directory, or press **Enter** to accept the default path.
9. Review the pre-installation summary. To fix any error, type `back` and press **Enter**.
10. Press **Enter** to start the installation.
11. Review the post-installation summary.  
For more information about installation tasks and to view configuration properties for the installed components, view the installation log files.
12. Press **Enter** to complete the installation process.

## Installing the Client Plug-in

Run the Data Quality Integration Client installer on PowerCenter Client machines.

1. Close all other applications.
2. Extract the Integration Client file.
3. Open the directory where you extracted the files.
4. Run `install.exe`.
5. On the **Welcome** screen, click **Next**.  
The installer displays a prompt to close PowerCenter client applications before continuing the installation.
6. Verify that no PowerCenter client applications are running and click **OK**.
7. Specify the path to the PowerCenter Client directory. The installer provides a default path to this directory. To select another path, click **Choose**.



8. Click **Next**.
9. Review the pre-installation summary. Click **Previous** to fix any error.
10. Click **Install**.
11. Review the post-installation summary.  
For more information about installation tasks and to view configuration properties for the installed components, view the installation log files.
12. Click **Done**.

## Registering the Plug-In with PowerCenter

Register the plug-in with any PowerCenter repository that you want to contain data quality objects.

The Integration installer writes `IDQZIntegration.xml` to the `$INFA_HOME\server\bin\native` directory on the PowerCenter Integration Service machine. Verify that you can access this file. If required, copy or FTP this file to the PowerCenter Integration Service machine.

If you do not have the correct privileges to register the plug-in, contact the PowerCenter Repository Service administrator.

1. Open the PowerCenter Administration Console.
2. In the Navigator, select the repository you want to use for data quality objects.
3. Run the Repository Service in exclusive mode.
4. Click the **Plug-ins** tab.
5. Click the **Register Repository Service plug-in** link.
6. Click **Browse** to find the `IDQZIntegration.xml` file.
7. To overwrite the registration of an older Integration plug-in, select the option to update the plug-in registration.
8. Enter your repository user name and password.
9. Click **OK**.

The Repository Service registers the plug-in with the repository. The results of the registration operation appear in the activity log.

10. Run the Repository Service in normal mode.

## CHAPTER 3

# Reference Data Requirements

This chapter includes the following topics:

- [Reference Data Overview, 18](#)
- [Reference Data File Requirements, 19](#)
- [Address Reference Data Requirements, 20](#)
- [Identity Population Data Requirements, 21](#)
- [Probabilistic Model and Classifier Model Requirements, 22](#)

## Reference Data Overview

Data quality transformations can read reference data. If any transformation in the PowerCenter repository reads reference data, verify that the reference data files are present in the PowerCenter environment.

Data quality transformations in PowerCenter can use the following types of reference data:

### **Reference data files**

Text files that contain the standard versions of business terms and alternative versions of the terms. You can edit reference data files.

### **Address reference data files**

Data files that contain complete data for all postal addresses in a country. You install address reference data files on the PowerCenter Integration Service machine. You cannot edit address reference data files.

### **Identity population files**

Data files that define different types of identity and contain algorithms used by the Match transformation. You install population files on the PowerCenter services machine. You cannot edit population files.

### **Content sets**

Data files that contain reference data specified by one or more transformations. When you export a mapping that reads content sets, the export process exports the content set data with the transformation metadata. In the case of classifier and probabilistic model files, the process exports the file data from the Data Quality directory structure.

# Reference Data File Requirements

A data quality mapping can read reference data objects that are defined in the Model repository.

When an Informatica Data Quality user exports a mapping that reads a reference data object, the export process writes the reference data to the file system as one or more reference data files. When you use the mapping in PowerCenter, you must verify that the files are installed at a location that the Integration Service can read.

You specify the file location for the reference data files when you export the mapping. Export the files into the PowerCenter directory structure, or export the files to the file system and copy the files to the PowerCenter Integration Service machine.

The reference data file locations in the PowerCenter directory structure must correspond to the reference table locations in the Model repository structure. The directory structure must follow this format:

```
<PowerCenter_Root_Directory>\services\<Model_Repository_Project_Name>\<Model_Repository_Project_Folder_Name>
```

**Note:** PowerCenter reads reference data files in UTF-8 format. The Developer tool exports reference data to PowerCenter in UTF-8 format.

## Installing Reference Data from the Data Quality Content Installer

The Data Quality Content Installer file set includes reference data files that it to the PowerCenter directory structure.

You identify the PowerCenter installation directory when you run the Content Installer. By default, the Content Installer appends the following path to the root:

```
\services\DQContent\INFA_Content\
```

Edit the path to match the Model repository structure, or copy the reference data files to the required path after you run the Content Installer.

## Using an Environment Variable to Define the reference data File Location

You can define an *INFA\_CONTENT* environment variable on the PowerCenter Integration Service machine to set the path that the Integration Service uses to read reference data files.

Use this environment variable if you cannot install the reference data files in the required location on the PowerCenter Integration Service machine. Install the files in a location that the PowerCenter Integration Service can read, and create the *INFA\_CONTENT* environment variable with the installed path.

# Address Reference Data Requirements

A mapping that performs address validation reads address reference data. You purchase and download address reference data from Informatica. Copy the data files to the PowerCenter Integration Service machine.

The following table describes the types of address reference data you can install:

Type of Reference Data	When To Use
Batch	Use for general address validation operations.
CAMEO	Use to add consumer demographic data to the address records.
Certified	<p>Use to verify that address records meet the certification standards that a mail carrier defines.</p> <p>The following countries define certification standards:</p> <ul style="list-style-type: none"><li>- Australia. Certifies mail according to the Address Matching Approval System (AMAS) standard.</li><li>- Canada. Certifies mail according to the Software Evaluation And Recognition Program (SERP) standard.</li><li>- France. Certifies mail according to the National Address Management Service (SNA) standard.</li><li>- New Zealand. Certifies mail according to the SendRight standard.</li><li>- United States. Certifies mail according to the Coding Accuracy Support System (CASS) standard.</li></ul>
Suggestion list	Use to find alternative valid versions of an incomplete address record. Use suggestion list data when you configure an address validation mapping to process single address records in real time.
Geocode	Use to add latitude coordinates and longitude coordinates to the address records.
Supplementary	<p>Use to identify the geographical area or population area that contains an address.</p> <p>The Address Validator transformation can add supplementary data to address records from the following countries:</p> <ul style="list-style-type: none"><li>- Japan. Identifies the urban district that contains the address.</li><li>- Serbia. Identifies the street-level postal code for the address.</li><li>- United Kingdom. Adds a delivery point identifier to the address.</li><li>- United States. Identifies the population center that contains the address.</li></ul>

You can also use the Data Quality Content Installer to install the address reference data. Run the Server Content Installer on the PowerCenter Integration Service machine to install the files.

The Data Quality Content Installer creates an `/av/` directory in the PowerCenter directory structure and writes address reference data to that directory. The PowerCenter Integration Service reads the path to the address reference data from the `AD50.cfg` configuration file.

The PowerCenter installer writes the `AD50.cfg` file to the following path:

```
[PowerCenter_Installation]/server/bin
```

When you install address reference data, verify the following properties are set in `AD50.cfg`:

## Reference Data Location

Enter the path to the address reference data in the `ReferenceDataLocation` property.

### License Key Data

Enter license data in the `LicenseKey` property. If you have more than one license, enter each license in a comma-separated string.

### Preload Values

Enter at least one country abbreviation as a preload value for each type of reference data that a mapping reads. Enter ALL to apply a preload setting for all countries.

The Integration Service can use a different method to load data for each country. For example, you can specify full preload for United States batch/interactive data and partial preload for United Kingdom batch/interactive data. The Integration Service can also use a different preload method for each type of data. For example, you can specify full preload for United States batch/interactive data and partial preload for United States geocoding data.

Full preload settings supersede partial preload settings, and partial preload settings supersede settings that indicate no data preload. For example, if you enter ALL for no data preload and enter USA for full preload, the Integration Service loads all United States data into memory and does not load data for any other country. If you do not have a preload requirement, enter ALL for no data preload for any type of reference data that you plan to use.

You do not specify a preload value for Supplementary data.

You can optionally edit other properties in `AD50.cfg`.

**Note:** Informatica Data Quality 9.x does not read the `AD50.cfg` file. Informatica Data Quality 9.x users enter the reference data location, license key information, and preload values through the Administrator tool.

## Identity Population Data Requirements

A mapping that performs identity match analysis reads reference data files called identity populations. You purchase and download the identity population files from Informatica. Copy the population files to the PowerCenter Integration Service machine or to a shared directory on a machine that the PowerCenter Integration Service can access.

You can also use the Data Quality Content Installer to install the population files. Run the Server Content Installer on the PowerCenter Integration Service machine to install the files.

You must copy or install the population files to a directory with the name `default`.

The PowerCenter Integration Service can read the location of the population file directory from the `IDQTx.cfg` configuration file or from the SSAPR environment variable. The PowerCenter Integration Service looks for the location in the `IDQTx.cfg` file. If the `IDQTx.cfg` file is not present or does not specify a location, the PowerCenter Integration Service looks for the location in the SSAPR environment variable.

**Note:** The `IDQTx.cfg` file and the SSAPR environment variable specify the path to the parent directory of the `default` directory. The path does not include the directory name.

The PowerCenter installer writes the `IDQTx.cfg` file to the following path:

```
[PowerCenter_Installation]/server/bin
```

The `IDQTx.cfg` file uses the `IdentityReferenceDataLocation` property to specify the location of the population file directory. Update the property with the current location of the directory.

For example, you might set the following path on the `IdentityReferenceDataLocation` property:

```
[PowerCenter_Installation]/services/DQContent/INFA_Content/identity/
```

You can set an absolute path or a relative path to the default directory on the *IdentityReferenceDataLocation* property. Set an absolute path to a directory on the PowerCenter Integration Service machine or to a shared directory on a machine that the service can access. Set a relative path to a directory on the PowerCenter Integration Service machine.

The PowerCenter Integration Service reads a relative path from the following directory:

```
[PowerCenter_Installation]/server/bin
```

For example, to specify a relative path to a population file directory in the `bin` directory, set the following relative path:

```
./
```

## IDQTx.cfg Properties

When you set the *IdentityReferenceDataLocation* property in the IDQTx.cfg file, you can also set the following properties:

### IdentityCacheDir

Path to the directory that stores temporary data files created when the mapping runs. The path identifies a parent directory. The PowerCenter Integration Service writes the temporary files to directories below the location that you specify.

The property has the following default value:

```
./identityCache
```

### IdentityIndexDir

Path to the directory that contains the temporary index files created when the mapping runs. Identity match analysis uses the index to sort records into groups before match analysis. The path identifies a parent directory. The PowerCenter Integration Service writes the index files to directories below the location that you specify.

The property has the following default value:

```
./identityIndex
```

**Note:** A Developer tool user can set the cache directory path and the index directory path in the Match transformation. The PowerCenter Integration Service uses the path in the IDQTx.cfg file if the Match transformation does not specify the path.

# Probabilistic Model and Classifier Model Requirements

Probabilistic models and classifier models are reference data files that identify the types of information in a data string. Before you run a mapping that reads a probabilistic model or classifier model file, you configure PowerCenter to read the files.

Perform the following configuration tasks:

- Verify the model file locations in the PowerCenter directory structure.
- Update the Java memory allocation on the PowerCenter Integration Service.

## Probabilistic Model and Classifier Model File Locations

When you export a mapping that reads probabilistic model or classifier model data, you specify a directory location for the exported model data file. Export the file to the PowerCenter directory structure, or export the file to the file system and copy it to the PowerCenter services machine.

By default, the PowerCenter Integration Service reads the model files from the following directory:

```
[PowerCenter_Installation]/server/bin
```

If the files are not in the default location, the PowerCenter Integration Service reads the file locations from property files. You update the property files with the paths to the model files.

The following table describes the property files:

File Name	Description
CLASSIFIER.properties	Contains the <i>ClassifierFileLocation</i> property. Update the property to identify the path to the classifier model files.
NER.properties	Contains the <i>NERFileLocation</i> property. Update the property to identify the path to the probabilistic model files.

The PowerCenter Integration Service reads the property files from the following directory:

```
[PowerCenter_Installation]/server/bin
```

You can set an absolute path or a relative path to identify the model file locations in each property file. If you set a relative path, the PowerCenter Integration Service adds the path to the following directory:

```
[PowerCenter_Installation]/server/bin
```

**Note:** Use forward slashes to define the directory paths to the model files on all platforms.

## PowerCenter Integration Service Memory Allocation

A PowerCenter Integration Service uses large amounts of memory to read a classifier model file or a probabilistic model file. If the PowerCenter Integration Service cannot access the required memory, it fails to read the model file and it cannot process the data.

Review the memory allocation on the PowerCenter Integration Service. If you run sessions that reference a classifier model file or a probabilistic model file, set the memory allocation to at least 512 MB.

You can use the *Java SDK Maximum Memory* process property to allocate the memory. To review or update the property, log in to the Administrator tool and select the PowerCenter Integration Service that runs the session tasks. You find the *Java SDK Maximum Memory* property on the **Processes** view.