



Informatica® PowerExchange for Greenplum  
10.1.1

# User Guide for PowerCenter

© Copyright Informatica LLC 2012, 2018

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, 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](mailto: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/license.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/license.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.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, please report them to us in writing at Informatica LLC 2100 Seaport Blvd. Redwood City, CA 94063.

INFORMATICA LLC 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: 2018-09-27

# 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: Introduction to PowerExchange for Greenplum.....</b>	<b>7</b>
PowerExchange for Greenplum Overview. ....	7
 <b>Chapter 2: PowerExchange for Greenplum Configuration .....</b>	<b>8</b>
PowerExchange for Greenplum Configuration Overview. ....	8
Prerequisites. ....	8
Registering the Plug-in. ....	9
PowerExchange for Greenplum Upgrade. ....	9
Upgrading PowerExchange for Greenplum from Versions Earlier than 9.5.1. ....	9
Upgrading PowerExchange for Greenplum from Version 9.5.1 or Later. ....	9
 <b>Chapter 3: Greenplum Targets.....</b>	<b>11</b>
Greenplum Targets Overview. ....	11
Importing Greenplum Target Definitions. ....	11
 <b>Chapter 4: Greenplum Sessions and Workflows.....</b>	<b>12</b>
Greenplum Sessions and Workflows. ....	12
SSL Authentication for Greenplum Targets. ....	12
Configuring a PowerExchange for Greenplum Connection. ....	13
Session Configuration with a Greenplum Target. ....	14
Guidelines for Configuring Greenplum Session Properties. ....	17
Match and Update Columns. ....	17
Partitioning Greenplum Targets. ....	18
Parameterization. ....	18
Error Handling for Greenplum Targets. ....	19
 <b>Appendix A: Greenplum Datatype Reference.....</b>	<b>20</b>
Greenplum and ODBC Data Types. ....	20
 <b>Index.....</b>	<b>22</b>

# Preface

The *Informatica PowerExchange® for Greenplum User Guide for PowerCenter®* provides information about loading data into a Greenplum target. It is written for database administrators and developers who are responsible for loading data into Greenplum. This book assumes you have knowledge of relational database concepts and database engines, Greenplum, and PowerCenter.

## 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

# Introduction to PowerExchange for Greenplum

This chapter includes the following topic:

- [PowerExchange for Greenplum Overview, 7](#)

## PowerExchange for Greenplum Overview

PowerExchange for Greenplum provides connectivity between PowerCenter and the Greenplum Database to load data.

When you run a Greenplum session, the PowerCenter Integration Service creates a control file to provide load specifications to the Greenplum gpload bulk loading utility, invokes the Greenplum gpload bulk loading utility, and writes data to the named pipe. The Greenplum gpload bulk loading utility launches gpfdist, which is Greenplum's file distribution program, that reads data from the named pipe and loads data into the Greenplum target.

You can use the Greenplum Database to store and analyze terabyte to petabytes of data on large clusters of powerful and inexpensive servers, storage, and ethernet switches. You can read data from any source system and use PowerExchange for Greenplum to load data into the Greenplum distributed storage system.

## CHAPTER 2

# PowerExchange for Greenplum Configuration

This chapter includes the following topics:

- [PowerExchange for Greenplum Configuration Overview, 8](#)
- [Registering the Plug-in, 9](#)
- [PowerExchange for Greenplum Upgrade, 9](#)

## PowerExchange for Greenplum Configuration Overview

PowerExchange for Greenplum installs with PowerCenter. After you install or upgrade Informatica Services, you must register the PowerExchange for Greenplum plug-in with the PowerCenter repository.

### Prerequisites

Before you can use PowerExchange for Greenplum, perform the following tasks:

1. Verify that the DataDirect Greenplum Wire Protocol exists in the ODBC data source administrator on the Windows machine that hosts the PowerCenter Client. Use the DataDirect Greenplum driver to import metadata from Greenplum.
2. Install the Greenplum loaders package on the machine where the PowerCenter Integration Service runs. The loaders package contains the gpload utility.

For more information about setting up the gpload utility, see the Greenplum documentation.

3. Verify that you can connect to Greenplum Database with the gpload utility.
4. Verify that the PATH, GPHOME\_LOADERS, PYTHONPATH, and LIBRARY\_PATH environment variables are set after you install the loaders package.
5. Restart both the PowerCenter Repository Service and the PowerCenter Integration Service after you update the environment variables.
6. Verify that you have read, write, and execute permissions on the following directories:

`<PowerCenter Installation Directory>/server/bin`

`<PowerCenter Installation Directory>/server/bin/Plugin`



# Registering the Plug-in

After you install Informatica Services, you must register the plug-in with the PowerCenter repository.

A plug-in is an XML file that defines the functionality of PowerExchange for Greenplum. To register the plug-in, the repository must be running in exclusive mode. Use the Administrator tool or the pmrep RegisterPlugin command to register the plug-in.

The plug-in file for PowerExchange for Greenplum is `GreenplumConnector.xml` and is available at the following directory after you install Informatica Services:

```
<PowerCenter Installation Directory>\server\bin\Plugin
```

**Note:** If you do not have the Manage Services privilege to register the plug-in, contact the user who manages the PowerCenter Repository Service.

## PowerExchange for Greenplum Upgrade

The steps to upgrade PowerExchange for Greenplum depend on the version that you upgrade from.

### Upgrading PowerExchange for Greenplum from Versions Earlier than 9.5.1

When you upgrade PowerExchange for Greenplum from versions earlier than 9.5.1, you must run the pmrep upgradeGreenplumToPartitionable command.

1. Upgrade PowerCenter and the PowerCenter Repository Service.
2. Verify that the plug-in file `GreenplumConnector.xml` is located in the following location: `<PowerCenter Installation Directory>/server/bin/Plugin`
3. Start the PowerCenter Repository Service in exclusive mode.
4. Run the pmrep connect command to connect to the PowerCenter Repository Service.
5. Run the pmrep upgradeGreenplumToPartitionable command to upgrade PowerExchange for Greenplum.

Enter the following command:

```
pmrep upgradeGreenplumToPartitionable
```

6. Restart the PowerCenter Repository Service to run in normal mode.

You can view the log file generated by the pmrep utility to get more information about the total number of rows that were updated. The log file is copied to the following directory:

```
<PowerCenter Installation Directory>/server/infa_shared/Log
```

### Upgrading PowerExchange for Greenplum from Version 9.5.1 or Later

When you upgrade PowerExchange for Greenplum from version 9.5.1 or later, you must update the existing plug-in file registration.

1. Upgrade PowerCenter and the PowerCenter Repository Service.
2. Start the PowerCenter Repository Service in exclusive mode.
3. In Informatica Administrator, click the **Plug-Ins** tab.

4. Click the **Register Plug-in** icon.
5. Browse and select the Greenplum plug-in file and select the option to update the existing plug-in file registration.

The Greenplum plug-in file `GreenplumConnector.xml` is located in the following location:

`<PowerCenter Installation Directory>/server/bin/Plugin`

6. Enter your repository user name and password.
7. Click **OK**.

The PowerCenter Repository Service updates the existing Greenplum plug-in file registration.
8. Restart the PowerCenter Repository Service to run in normal mode.

## CHAPTER 3

# Greenplum Targets

This chapter includes the following topics:

- [Greenplum Targets Overview, 11](#)
- [Importing Greenplum Target Definitions, 11](#)

## Greenplum Targets Overview

Greenplum target definitions represent metadata for Greenplum tables.

You can import Greenplum tables from the Greenplum database into the Designer before loading the table. You cannot use the Designer to create new tables in the Greenplum database. When you import Greenplum tables as target definitions, the ODBC datatypes corresponding to the Greenplum datatypes appear in the Designer.

## Importing Greenplum Target Definitions

You must import the Greenplum target definition into the Target Designer. Use an ODBC data source to import a Greenplum target definition.

Use the Greenplum drivers to create a DSN connection to the Greenplum database.

1. In the Target Designer, click **Targets > Import from Database**.
2. Select the Greenplum data source used to connect to the target database.  
If you need to create a Greenplum data source, click the **Browse** button to open the **ODBC Administrator**, create the Greenplum data source, and select the Greenplum data source.
3. Enter the user name and password to connect to the database, and click **Connect**.  
If you are not the owner of the table that you want to use as a target, specify the owner name.
4. Drill down through the list of database objects to view the available tables as targets.
5. Select the Greenplum table or tables to import into the repository.  
You can press and hold the Shift key to select a block of tables, or press and hold the Ctrl key to make non-contiguous selections. You can also use the **Select All** and **Select None** buttons to select or clear all available targets.
6. Click **OK**.  
The selected target definitions appear in the Navigator under the Targets node.

## CHAPTER 4

# Greenplum Sessions and Workflows

This chapter includes the following topics:

- [Greenplum Sessions and Workflows, 12](#)
- [SSL Authentication for Greenplum Targets, 12](#)
- [Configuring a PowerExchange for Greenplum Connection, 13](#)
- [Session Configuration with a Greenplum Target, 14](#)
- [Partitioning Greenplum Targets, 18](#)
- [Parameterization, 18](#)
- [Error Handling for Greenplum Targets, 19](#)

## Greenplum Sessions and Workflows

In the Workflow Manager, create a session and associate it with a Greenplum mapping. Configure a Greenplum relational connection before you run the workflow. You can also define properties in a session to determine how the gpload utility writes data to a Greenplum target. When you run a session, the PowerCenter Integration Service invokes the gpload utility to load data into the Greenplum database.

## SSL Authentication for Greenplum Targets

You can configure secure communication between the gpload utility and the Greenplum server by using the Secure Sockets Layer (SSL) protocol. SSL is a protocol that ensures secure data transfer between a client and a server.

To enable PowerExchange for Greenplum to secure communication between the gpload utility and the Greenplum server, select the **Enable SSL** option in the Greenplum connection. In the Greenplum connection, you must also define the path where the SSL certificates for the Greenplum server are stored.

For information about configuring SSL for the gpload utility, see the gpload documentation.

# Configuring a PowerExchange for Greenplum Connection

Before you run a Greenplum workflow, configure a Greenplum relational connection in the Workflow Manager. When you configure a Greenplum relational connection, you define the connection attributes that the gpload utility uses to connect to the Greenplum database.

1. In the Workflow Manager, connect to a repository.
2. Click **Connections > Relational**.  
The **Relational Connection Browser** dialog box appears.
3. Click **New**.  
The **Select Subtype** dialog box appears.
4. Select **Greenplum Connection** from the **Select Subtype** list.
5. Click **OK**.  
The **Connection Object Definition** dialog box appears.
6. Enter the following connection information:

Connection Attribute	Description
Name	Name for the Greenplum relational connection.
User Name	User name with permissions to access the Greenplum database. You can connect to a database that runs on a network that uses Kerberos authentication. To configure Kerberos authentication for the database connection, set the user name to the reserved word <i>PmKerberosUser</i> . If you use Kerberos authentication, the connection uses the credentials of the user account that runs the session to connect to the database. The user account must have a user principal on the Kerberos network where the database runs.
Password	Password to connect to the Greenplum database. If you set the user name to <i>PmKerberosUser</i> to use Kerberos authentication for the database connection, set the password to the reserved word <i>PmKerberosPassword</i> . The connection uses the credentials of the user account that runs the session to connect to the database.
Host Name	Host name or IP address of the Greenplum server.
Port	Greenplum server port number. If you enter 0, the gpload utility reads from the environment variable \$PGPORT. Default is 5432.
Database	Name of the database.
Enable SSL	Select this option to establish secure communication between the gpload utility and the Greenplum server over SSL.
Certificate Path	Path where the SSL certificates for the Greenplum server are stored. For information about the files that need to be present in the certificates path, see the gpload documentation.
Schema	Name of the schema that contains the metadata for Greenplum targets. Default is public.

7. Click **OK**.

The database connection appears in the **Connection Browser** list.

## Session Configuration with a Greenplum Target

You can configure the session properties for Greenplum targets in the **Transformations** view on the **Mapping** tab. Define the properties for each target instance in the session.

You can configure the following session properties for Greenplum targets:

### Method

Determines how the gpload utility processes the data from the named pipe:

- **Insert.** Inserts rows into the target.
- **Update.** Updates rows in the target.
- **Merge.** If the rows exist in the target, updates the existing rows. If the rows do not exist in the target, inserts the rows into the target.

### Match Columns

Matches rows based on the comma-separated list of column names. Enclose the column names in double quotation marks and ensure that there are no leading or trailing spaces between the column names.

### Update Columns

Updates the columns specified in the comma-separated list of column names. Enclose the column names in double quotes and ensure that there are no leading and trailing spaces between the column names.

### Update Condition

Updates a row based on the condition that you specify. The gpload utility performs an update or merge operation based on the update condition.

### Format

The PowerCenter Integration Service writes data in a format that is compatible with the gpload utility. Select one of the following values:

- **Text.** In the text format, the PowerCenter Integration Service separates data using the delimiter character that you specify in the session properties. If the data contains the delimiter or escape characters specified in the session properties, you can choose to ignore the escape character or specify delimiter and escape character values that are not a part of the data.
- **CSV.** In the CSV format, the PowerCenter Integration Service encloses the data with the quote character that you specify in the session properties. The PowerCenter Integration Service also separates the data using the delimiter character specified in the session properties. If the data contains the quote or escape characters specified in the session properties, you can choose to ignore the escape character or specify quote and escape character values that are not a part of the data.

Default is Text.

**Note:** If the data contains newline characters, you must use the CSV format. If you use the text format and the data contains newline characters, the data after the newline character is treated as a new record. In such situations, the gpload utility might reject or insert incorrect data into the tables.

**Delimiter**

Delimiter separates successive input fields. For data in the text format, use any 7-bit ASCII value except a-z, A-Z, and 0-9. For data in the CSV format, use any 7-bit ASCII value except \n, \r, and \\. Default is pipe (|).

You can also specify a non-printable ASCII character through an escape sequence by using the decimal representation of the ASCII character. For example, \014 represents the shift out character.

**Escape**

Character that treats special characters in the data as regular characters. In the text format, special characters comprise delimiter and escape characters. In the CSV format, special characters comprise quotes and escape characters. Use any 7-bit ASCII value as an escape character. Default is backslash (\).

**Note:** You can improve the session performance if the data does not contain escape characters.

**Skip Escaping**

Skips escaping special characters in the data. Clear this option to treat special characters in the data as regular characters.

**Null As**

String that represents a null value. In the source data, any data item that matches the string is treated as a null value. Default is backslash N (\N).

**Quote**

Character that encloses the data in the CSV format. The PowerCenter Integration Service encloses data by the specified character and passes the data to the gpload utility. The quote character is ignored for data in the text format. Use any 7-bit ASCII value that is not equal to the delimiter or null value. Default is double quotes (").

**Error Table**

Name of the error table where the gpload utility logs rejected rows when reading data that is processed by the PowerCenter Integration Service. The naming convention for the table name is <schema name>.<table name>, where schema name is the name of the schema that contains the table.

**Error Limit**

For each Greenplum segment, across all partitions if pass-through partitioning is used, the number of rows that the gpload utility discards or logs in the error table because of format errors. The gpload utility fails the session if the error limit is reached for any Greenplum segment. Default is zero. The maximum error limit is 2,147,483,647.

**Greenplum Pre SQL**

The SQL command to run before loading data to the target.

**Greenplum Post SQL**

The SQL command to run after loading data to the target.

**Truncate Target Option**

Truncates the Greenplum target table before loading data to the target.

**Reuse Table**

Determines if the gpload utility drops the external table objects and staging table objects it creates. The gpload utility reuses the objects for future load operations that use the same load specifications.

Default is unchecked.

**Greenplum Target Table**

Overrides the default target table name.

**Greenplum Loader Logging**

Sets the logging level for the gpload utility. You can select one of the following values:

- None
- Verbose
- Very Verbose

Default is None.

**Greenplum gpfdist Timeout**

The number of seconds that elapse before the gpfdist process times out when attempting to connect to the target. The default value is 30 seconds.

**Window Pipe Buffer Size**

The size (in kilobytes) that the PowerCenter Integration Service allocates to buffer data before writing to the Greenplum bulk loader. Enter a value between 1 and 2048. The default value is 2048 KB. You might need to test different settings for optimal performance. This attribute is applicable for Informatica servers that run on Windows.

**Delete Control File**

Determines if the PowerCenter Integration Service must delete the gpload control file after the session is complete.

Default is selected.

**Gpload Log File Location**

The file system location where the gpload utility generates the gpload log file.

Default is `$PMRootDir/Temp`.

**Gpload Control File Location**

The file system location where the PowerCenter Integration Service generates the gpload control file.

Default is `$PMRootDir/Temp`.

**Encoding**

Character set encoding of the source data.

PowerExchange for Greenplum supports only the UTF-8 character set encoding.

**Pipe Location**

The file system location where the pipes that are used for data transfer are created. This attribute is not applicable to Informatica servers that run on Windows.

Default is `$PMRootDir/Temp`.

**Schema Override**

Overrides the schema that is specified in the Greenplum connection object.

If you do not configure this property, the Data Integration Service uses the schema value specified in the Greenplum connection object.



### Max\_Line\_Length

The Max\_Line\_Length integer specifies the maximum length of a line in the XML transformation data that is passed to gpload.

The session that you configure for Greenplum targets depends on the third party implementation.

Use the Greenplum target session properties instead of the ODBC object properties, such as Table Name Prefix, Pre SQL, Post SQL, and Target Table Name. The PowerCenter Integration Service will not process the ODBC object properties.

The following table compares the Greenplum target session properties that you should use instead of the corresponding ODBC object properties that appear in the **Transformations** view on the **Mapping** tab:

Greenplum Property	ODBC Object Property
Schema Override	Table Name Prefix
Greenplum Pre SQL	Pre SQL
Greenplum Post SQL	Post SQL
Greenplum Target Table	Target Table Name

## Guidelines for Configuring Greenplum Session Properties

Use the following guidelines when you configure session properties for loading data into a Greenplum target:

- Use the recommended values for the delimiter, escape, and quote characters. If you specify values that are not valid, the gpload utility fails the session.
- Use the default values for the delimiter, escape, and quote characters to improve the session performance. When you use the default values, the PowerCenter Integration Service validates the character and unchar datatypes for special characters. The PowerCenter Integration Service does not validate the rest of the datatypes.
- Verify that the column names specified in the Match Columns and Update Columns session properties do not have leading or trailing spaces. When you import target definitions, the Designer strips leading and trailing spaces from the column names. Therefore, if the Match Columns and Update Columns session properties have leading or trailing spaces, the gpload utility logs an error and the session fails. The gpload utility cannot load data into the Greenplum target because the column names in the target definition do not match the column names specified in the session properties.

## Match and Update Columns

Before you run a session that loads data to a Greenplum target, you can configure the match and update columns.

### Match Columns

You can specify the columns to use as the join condition for the update. The attribute value in the specified target columns must be equal to that of the corresponding source data columns in order for the row to be updated in the target table. You must specify the match columns if the method to process data from the named pipe is update or merge.

## Update Columns

You can specify the columns to update for the rows that meet the criteria for match columns and the update column property. Update columns cannot be columns that are used for the Greenplum distribution key for the table. You must specify the match columns if the method to process data from the named pipe is update or merge.

The following table describes the session properties you can configure for match and update columns:

Target Property	Description
Match Column	Matches rows based on the comma-separated list of column names. Enclose the column names in double quotes and ensure that there are no leading and trailing spaces between the column names.
Update Column	Updates the columns specified in the comma-separated list of column names. Enclose the column names in double quotes and ensure that there are no leading and trailing spaces between the column names.

## Partitioning Greenplum Targets

If you need to load a large amount of data to a Greenplum target, you can configure partitioning for the session to improve session performance.

You can configure pass-through partitioning for Greenplum sessions. The PowerCenter Integration Service processes partitions concurrently.

When you run a Greenplum session, the PowerCenter Integration Service creates a control file to provide load specifications to the gpload utility, invokes the gpload utility, and writes data to the named pipe. Each partition creates a pipe. The gpload utility launches gpfdist, which is the file distribution program of Greenplum, that reads data from the named pipe and loads data into the Greenplum target.

The session properties that you define for a Greenplum target apply across all partitions. The error limit you specify is applicable across all partitions and not individually on each partition. The gpload utility generates a single gpload log file that contains the aggregated results of all the partition loads.

## Parameterization

You can parameterize Greenplum session properties to override the properties during run time.

You can parameterize the following session properties:

- Greenplum Target Table
- Match Columns
- Update Columns
- Update Condition
- Delimiter
- Escape Character

- Null As
- Quote
- Error Table
- Greenplum Pre SQL
- Greenplum Post SQL

## Error Handling for Greenplum Targets

You can set the error limit for a Greenplum segment to specify the number of rows that the gpload utility can discard before it fails a session. If you specify an error table, the gpload utility logs the discarded rows in the error table.

The error limit includes rows with format errors. The default value is 0. By default, the gpload utility stops a session when it encounters a row with format errors.

Use the following naming convention for the error table name: `<schema name>.<table name>`

If you do not specify a schema name, the gpload utility creates the error table in the public schema. The error table format is predefined in the Greenplum database.

- If the table does not exist, the gpload utility creates the table based on the predefined format.
- If the specified table exists in the schema, but the table is not in the prescribed format, the session fails.
- If a session fails, see the error table for more information about the errors.
- If you run the session again, the gpload utility appends the discarded rows to the error table.

For more information about the error tables, see the Greenplum documentation.

You can view load statistics in the session log. The gpload utility writes the error messages to the gpload log. The PowerCenter Integration Service reads the gpload log and writes the errors to the session log. The gpload utility writes the error messages to the gpload log at the following location:

```
$INFA_HOME/server/infa_shared/temp
```

## APPENDIX A

# Greenplum Datatype Reference

This appendix includes the following topic:

- [Greenplum and ODBC Data Types, 20](#)

## Greenplum and ODBC Data Types

When you import Greenplum tables as target definitions, the ODBC data types corresponding to the Greenplum data types appear in the Designer. The PowerCenter Integration Service writes the data to the Greenplum writer as ODBC data type. The Greenplum writer writes the data into the gpload utility and the gpload utility converts the data type to the Greenplum data type before it writes to the Greenplum database.

The following table lists the Greenplum data types and the corresponding ODBC data types:

Greenplum Data Type	ODBC Data Type	Transformation Data Type
Bigint	Bigint	Bigint
Bigserial	Bigint	Bigint
Boolean	Boolean	String
Character	Default is Char. For multibyte character, use Nchar.	String
Character varying	Default is Varchar. For multibyte character, use Nvarchar.	String
Date	Date	Date/Time
Double precision	Double or Float	Double
Integer	Integer	Integer
Numeric	Numeric	Decimal
Real	Real	Double
Serial	Integer	Integer
Smallint	Tinyint	Integer
Text	Text	Text

Greenplum Data Type	ODBC Data Type	Transformation Data Type
Time	Time. Precision is 6.	Date/Time
Timestamp	Timestamp. Precision is 6.	Date/Time

For more information on transformation data types, see the *PowerCenter Designer Guide*.

# INDEX

## C

- configuration
  - session [14](#)
- configuring
  - connection [13](#)
- connection
  - configuring [13](#)

## E

- error handling [19](#)

## G

- Greenplum target definitions
  - importing [11](#)

## P

- plug-ins
  - registering PowerExchange for Greenplum [9](#)
- prerequisites [8](#)

## S

- session
  - configuration [14](#)
- SSL authentication
  - overview [12](#)