



# Informatica® PowerExchange for Greenplum 10.2

## User Guide

© Copyright Informatica LLC 2014, 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, PowerExchange, and Big Data Management 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.

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

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

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

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

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

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

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

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

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

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

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

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

This product includes software developed by Boost (<http://www.boost.org/>) or under the Boost software license. Permissions and limitations regarding this software are subject to terms available at [http://www.boost.org/LICENSE\\_1\\_0.txt](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/hsqldbLicense.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/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 products are warranted according to the terms and conditions of the agreements under which they are provided. INFORMATICA PROVIDES THE INFORMATION IN THIS DOCUMENT "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

Publication Date: 2018-09-25

# Table of Contents

<b>Preface .....</b>	<b>6</b>
Informatica Resources. ....	6
Informatica Network. ....	6
Informatica Knowledge Base. ....	6
Informatica Documentation. ....	6
Informatica Product Availability Matrixes. ....	7
Informatica Velocity. ....	7
Informatica Marketplace. ....	7
Informatica Global Customer Support. ....	7
 <b>Chapter 1: Introduction to PowerExchange for Greenplum.....</b>	 <b>8</b>
PowerExchange for Greenplum Overview. ....	8
Introduction to the Greenplum Database. ....	8
 <b>Chapter 2: PowerExchange for Greenplum Configuration.....</b>	 <b>10</b>
Prerequisites. ....	10
Prerequisites for Native Environment. ....	10
Prerequisites for Hadoop Environment. ....	11
Environment Variables. ....	11
 <b>Chapter 3: Greenplum Connections.....</b>	 <b>13</b>
Greenplum Connection Overview. ....	13
SSL Authentication for Greenplum Targets. ....	13
Greenplum Connection Properties. ....	14
Creating a Greenplum Connection. ....	15
 <b>Chapter 4: PowerExchange for Greenplum Data Objects.....</b>	 <b>16</b>
Greenplum Data Object Overview. ....	16
Greenplum Data Object Views. ....	16
Greenplum Data Object Overview Properties. ....	17
Greenplum Data Object Write Operation Properties. ....	17
Input Properties of a Data Object Write Operation. ....	17
Target Properties of the Data Object Write Operation. ....	21
Importing a Greenplum Data Object. ....	22
Creating a Data Object Operation Write Operation. ....	23
 <b>Chapter 5: Greenplum Mappings.....</b>	 <b>24</b>
Greenplum Mappings Overview. ....	24
Hadoop Validation and Run-time Environments. ....	24
Greenplum Mapping Example. ....	25

<b>Chapter 6: Greenplum Run-time Processing.....</b>	<b>27</b>
Greenplum Run-time Processing Overview. . . . .	27
Match and Update Columns. . . . .	27
Match Columns. . . . .	28
Update Columns. . . . .	28
Error Handling for Greenplum Targets. . . . .	28
Parameterization. . . . .	29
Partitioning. . . . .	30
 <b>Appendix A: Data Type Reference.....</b>	 <b>31</b>
Data Type Reference Overview. . . . .	31
Greenplum, JDBC, and Transformation Data Types. . . . .	31
 <b>Index. . . . .</b>	 <b>33</b>

# Preface

The *Informatica PowerExchange® for Greenplum User Guide* 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 Informatica.

## 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 topics:

- [PowerExchange for Greenplum Overview, 8](#)
- [Introduction to the Greenplum Database, 8](#)

## PowerExchange for Greenplum Overview

You can use PowerExchange for Greenplum to load data to a Greenplum database.

When you run a Greenplum mapping, the Data 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 pipe and loads data into the Greenplum target.

You can also use PowerExchange for Greenplum to load data to a HAWQ database in bulk.

You can run Greenplum mappings in native or Hadoop environments.

For example, in a native environment, you can store sales information of a store for the past five years. You can use PowerExchange for Greenplum in a native environment to process sales information and then write to Greenplum tables for storage.

For example, in a Hadoop environment, you can gather, store, and analyze large volumes of unstructured data like web logs for an airline. You can process these large amounts of data in a Hadoop environment and use PowerExchange for Greenplum to write meaningful data to Greenplum tables for analysis.

## Introduction to the Greenplum Database

The Greenplum Database is a distributed storage system that you can use to store and analyze terabyte to petabytes of data. You can store data on large clusters of powerful and inexpensive servers and storage systems.

A Greenplum database is an array of PostgreSQL databases that consist of the master and segments. You can connect to the Greenplum Database through the master. The master authenticates the client connections and processes the SQL commands. The segments store data. Additionally, the segments process most of the



queries. Greenplum interconnect or the networking layer that communicates between the PostgreSQL databases allows the system to behave as one logical database.

## CHAPTER 2

# PowerExchange for Greenplum Configuration

This chapter includes the following topics:

- [Prerequisites, 10](#)
- [Environment Variables, 11](#)

## Prerequisites

PowerExchange for Greenplum installs with the Informatica services and clients. Based on the environment type, complete the prerequisites.

### Prerequisites for Native Environment

Before you use PowerExchange for Greenplum in the native environment, you must configure the Informatica services and clients.

#### Configure Informatica Services

1. Install the Informatica services.
2. Create a Data Integration Service and a Model Repository Service in the Informatica domain.
3. Install the Greenplum loaders package on the machine where the Data Integration Service runs. The loaders package contains the gpload utility.
4. Verify that you can connect to the Greenplum Database with the gpload utility.
5. Configure the environment variables.
6. On UNIX platforms, ensure that you have write access to the \$HOME directory.

#### Configure Informatica Clients

1. Install the Informatica clients. When you install the Informatica clients, the Developer tool is installed.
2. On the Windows machine that hosts the Developer tool, copy the Pivotal JDBC driver to the following location:

```
<Informatica Installation Directory>/clients/externaljdbcjars
```

Use the Pivotal JDBC driver to import metadata from Greenplum.

## Prerequisites for Hadoop Environment

Before you use PowerExchange for Greenplum in a Hadoop environment, you must configure the Informatica services and clients.

### Configure Informatica Services

1. Install the Informatica services.
2. Create a Data Integration Service and a Model Repository Service in the Informatica domain.
3. Install the Greenplum loaders package on the machine where the Data Integration Service runs. The loaders package contains the gpload utility.
4. Verify that you can connect to the Greenplum Database with the gpload utility.
5. Additionally, install the Greenplum loaders package on every node of the Hadoop cluster.
6. Install Informatica Big Data Management™ on every node of the Hadoop cluster.
7. Configure the environment variables in the `hadoopEnv.properties` file.

**Note:** For more information about the `hadoopEnv.properties` file see the *Informatica Big Data Management User Guide*.

8. On UNIX platforms, ensure that you have write access to the \$HOME directory.

### Configure Informatica Clients

1. Install the Informatica clients. When you install the Informatica clients, the Developer tool is installed.
2. On the Windows machine that hosts the Developer client, copy the Pivotal JDBC driver to the following location:

`<Informatica Installation Directory>/clients/externaljdbcjars`

Use the Pivotal JDBC driver to import metadata from Greenplum.

## Environment Variables

You must configure Greenplum environment variables before you can use PowerExchange for Greenplum in the native or Hadoop environment.

For information about the environment variables that you need to set, access the `greenplum_loaders_path.sh` file or the `greenplum_loaders_path.bat` file from the location where you installed the Greenplum loaders package.

**Note:** The list of environment variables that you need to set might change. See the `greenplum_loaders_path` file for the latest updates.

Set the shared library environment variable based on the operating system.

The following table lists the shared library variables for each operating system:

Operating System	Variable
Solaris	LIBRARY_PATH
Linux	LIBRARY_PATH

Operating System	Variable
AIX	PATH
All supported operating systems	GPHOME_LOADERS
All supported operating systems	PYTHONPATH

In the native environment, set the environment variables on the node that runs the Data Integration Service.

In a Hadoop environment, set the environment variables in the `hadoopEnv.properties` file. On a Linux operating system, set the environment variables in the following format:

`infapdo.env.entry.ld_library_path`. The entry exists. Append the Greenplum `LIBRARY_PATH`.

`infapdo.env.entry.path`. The entry exists. Append the Greenplum `LIBRARY_PATH`.

`infapdo.env.entry.gphome_loaders=GPHOME_LOADERS=<gp home directory>`

`infapdo.env.entry.pythonpath=PYTHONPATH=<gp python directory>`

## CHAPTER 3

# Greenplum Connections

This chapter includes the following topics:

- [Greenplum Connection Overview, 13](#)
- [SSL Authentication for Greenplum Targets, 13](#)
- [Greenplum Connection Properties, 14](#)
- [Creating a Greenplum Connection, 15](#)

## Greenplum Connection Overview

Use a Greenplum connection to access a Greenplum database.

Create a connection to import Greenplum table metadata to create data objects, preview data, and run mappings. When you create a Greenplum connection, you define the connection attributes that the gpload utility uses to connect to a Greenplum database.

You can also use a Greenplum connection to load data to a HAWQ database in bulk. When you create the Greenplum connection, enter the connection attributes that are specific to the HAWQ database that you want to connect to. The gpload utility uses these attributes to connect to the HAWQ 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.

# Greenplum Connection Properties

Use a Greenplum connection to connect to a Greenplum database. The Greenplum connection is a relational type connection. You can create and manage a Greenplum connection in the Administrator tool or the Developer tool.

**Note:** The order of the connection properties might vary depending on the tool where you view them.

When you create a Greenplum connection, you enter information for metadata and data access.

The following table describes Greenplum connection properties:

Property	Description
Name	Name of the Greenplum relational connection.
ID	String that the Data Integration Service uses to identify the connection. The ID is not case sensitive. It must be 255 characters or fewer and must be unique in the domain. You cannot change this property after you create the connection. Default value is the connection name.
Description	Description of the connection. The description cannot exceed 765 characters.
Location	Domain on which you want to create the connection.
Type	Type of connection.

The user name, password, driver name, and connection string are required to import the metadata. The following table describes the properties for metadata access:

Property	Description
User Name	User name with permissions to access the Greenplum database.
Password	Password to connect to the Greenplum database.
Driver Name	The name of the Greenplum JDBC driver. For example: <code>com.pivotal.jdbc.GreenplumDriver</code> For more information about the driver, see the Greenplum documentation.
Connection String	Use the following connection URL: <code>jdbc:pivotal:greenplum://&lt;hostname&gt;:&lt;port&gt;;DatabaseName=&lt;database_name&gt;</code> For more information about the connection URL, see the Greenplum documentation.

PowerExchange for Greenplum uses the host name, port number, and database name to create a control file to provide load specifications to the Greenplum gpload bulk loading utility. It uses the Enable SSL option and the certificate path to establish secure communication to the Greenplum server over SSL.

The following table describes the connection properties for data access:

Property	Description
Host Name	Host name or IP address of the Greenplum server.
Port Number	Greenplum server port number. If you enter 0, the gpload utility reads from the environment variable \$PGPORT. Default is 5432.
Database Name	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.

## Creating a Greenplum Connection

Before you create a Greenplum data object, create a connection in the Developer tool.

1. Click **Window > Preferences**.
2. Select **Informatica > Connections**.
3. Expand the domain in the **Available Connections**.
4. Select **Databases > Greenplum Loaders** and click **Add**.
5. Enter a connection name.
6. Enter an ID for the connection.
7. Optionally, enter a connection description.
8. Select the domain on which you want to create the connection.
9. Select a Greenplum Loader connection type.
10. Click **Next**.
11. Configure the connection properties for metadata access and data access.
12. Click **Test Connection** to verify the connection to Greenplum.
13. Click **Finish**.

## CHAPTER 4

# PowerExchange for Greenplum Data Objects

This chapter includes the following topics:

- [Greenplum Data Object Overview, 16](#)
- [Greenplum Data Object Views, 16](#)
- [Greenplum Data Object Overview Properties, 17](#)
- [Greenplum Data Object Write Operation Properties, 17](#)
- [Importing a Greenplum Data Object, 22](#)
- [Creating a Data Object Operation Write Operation, 23](#)

## Greenplum Data Object Overview

A Greenplum data object is a physical data object that uses Greenplum as a target. A Greenplum data object is the representation of data that is based on a Greenplum table. You can configure the data object write operation properties that determine how data can be loaded to Greenplum targets.

Import the Greenplum table into the Developer tool to create a Greenplum data object. Create a data object write operation for the Greenplum data object. Then, you can add the data object write operation to a mapping.

## Greenplum Data Object Views

The Greenplum data object contains views to edit the object name and the properties.

After you create a Greenplum data object, you can change the data object and data object operation properties in the following data object views:

- **Overview** view. Use the **Overview** view to edit the Greenplum data object name, description, and tables.
- **Data Object Operation** view. Use the **Data Object Operation** view to view and edit the properties that the Data Integration Service uses when it writes data to Greenplum table.

When you create mappings with a Greenplum target, you can view the data object properties in the **Properties** view.



# Greenplum Data Object Overview Properties

The Greenplum **Overview** section displays general information about the Greenplum data object and detailed information about the Greenplum table that you imported.

You configure the following general properties for a Greenplum data object:

**Name**

Name of the Greenplum data object.

**Description**

Description of the Greenplum data object.

**Connection**

Name of the Greenplum connection. Click **Browse** to select another Greenplum connection.

You can view the properties for the Greenplum table that you import:

**Name**

Name of the Greenplum table.

**Type**

Native datatype of the Greenplum table.

**Description**

Description of the Greenplum table.

## Greenplum Data Object Write Operation Properties

The Data Integration Service writes data to a Greenplum table based on the data object write operation properties. The Developer tool displays the data object write operation properties for the Greenplum data object in the Data Object Operation section.

You can view or configure the data object write operation from the input and target properties.

- **Input properties.** Represents data that the Data Integration Service passes into the mapping pipeline. Select the input properties to edit the port properties and specify the advanced properties of the data object write operation.
- **Target properties.** Represents data that the Data Integration Service writes to the Greenplum table. Select the target properties to view data such as the name and description of the Greenplum table, create key columns to identify each row in a data object write operation, and create key relationships between data object write operations.

### Input Properties of a Data Object Write Operation

The input properties represent data that the Data Integration Service passes into the mapping pipeline. Select the input properties to edit the port properties of the data object write operation. You can also specify advanced data object write operation properties to load data into Greenplum tables.

The input properties of the data object write operation include general properties that apply to the data object write operation. They also include port, source, and advanced properties that apply to the data object write operation.

You can view and change the input properties of the data object write operation from the **General**, **Ports**, **Sources**, and **Advanced** tabs.

## General Properties

The general properties list the name and description of the data object write operation.

## Ports Properties

The input ports properties list the datatypes, precision, and scale of the data object write operation.

You configure the following input ports properties in a data object write operation:

### **Name**

Name of the port.

### **Type**

Datatype of the port.

### **Precision**

Maximum number of significant digits for numeric datatypes or maximum number of characters for string datatypes. For numeric datatypes, precision includes scale.

### **Scale**

Maximum number of digits after the decimal point for numeric values.

### **Description**

Description of the port.

## Sources Properties

The sources properties list the Greenplum tables in the data object write operation.

## Advanced Properties

The advanced properties allow you to specify data object write operation properties to load data into Greenplum tables.

You can configure the following advanced properties in the data object write operation:

### **Load 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 quotes and ensure that there are no leading and 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 specified condition. The gpload utility performs an update or merge operation based on the update condition specified.

### Format

The Data 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 Data Integration Service separates data using the delimiter character specified 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 Data Integration Service encloses the data with the quote character specified in the session properties. The Data Integration Service also separates the data using the delimiter character specified in 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 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 Data 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 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.

**Error Table**

Name of the error table where the gpload utility logs rejected rows when reading data that is processed by the Data 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.

**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. Default is unchecked.

**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 Schema Name**

Overrides the default schema name.

Name of the schema that contains the metadata for Greenplum targets.

**Greenplum Target Table**

Overrides the default target table name.

**Greenplum Loader Logging Level**

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.

**Windows Pipe Buffer Size**

The number of kilobytes that the Data Integration Service allocates to buffer data before writing to the Greenplum bulk loader. Enter a value from 1 through 131072. The default value is 2048 KB. You might need to test different settings for optimal performance. This attribute is applicable for Informatica servers running on Windows.

**Delete Control File**

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

Default is checked.

**Gpload Log File Location**

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

Ensure that the file system location exists on all the nodes.

Default is the location of the temporary directories of the Data Integration Service process node.

**Gpload Control File Location**

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

Ensure that the file system location exists on all the nodes.

Default is the location of the temporary directories of the Data Integration Service process node.

**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 used for data transfer are created. This attribute is not applicable to Informatica servers running on Windows.

Ensure that the file system location exists on all the nodes.

Default is the location of the temporary directories of the Data Integration Service process node.

**Port [Range]**

The port number or port range that the gpfdist uses to read data from the pipe and load data to the Greenplum target.

If you specify a port number, the gpfdist uses the port number. If the port number you specify is not available, gpload fails.

If you specify a port range, the gpfdist uses a port that is available from the port range.

Default is [8000, 9000].

**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.

## Target Properties of the Data Object Write Operation

The target properties represent the data that is populated based on the Greenplum tables that you added when you created the data object. The target properties of the data object write operation include general and column properties that apply to the Greenplum tables. You can view the target properties of the data object write operation from the **General**, **Column**, and **Advanced** tabs.

### General Properties

The general properties display the name and description of the Greenplum table.

## Column Properties

The column properties display the datatypes, precision, and scale of the target property in the data object write operation.

You can view the following target column properties of the data object write operation:

**Name**

Name of the column property.

**Type**

Native datatype of the column property.

**Precision**

Maximum number of significant digits for numeric datatypes, or maximum number of characters for string datatypes. For numeric datatypes, precision includes scale.

**Scale**

Maximum number of digits after the decimal point for numeric values.

**Primary Key**

Determines if the column property is a part of the primary key.

**Description**

Description of the column property.

## Advanced Properties

The advanced property displays the physical name of the Greenplum table.

# Importing a Greenplum Data Object

Import a Greenplum data object to add to a mapping.

1. Select a project or folder in the **Object Explorer** view.
2. Click **File > New > Data Object**.
3. Select **Greenplum Data Object** and click **Next**.  
The **Greenplum Data Object** dialog box appears.
4. Enter a name for the data object.
5. Click **Browse** next to the **Location** option and select the target project or folder.
6. Click **Browse** next to the **Connection** option and select the Greenplum connection from which you want to import the Greenplum table metadata.
7. To add a table, click **Add** next to the **Selected Resources** option.  
The **Add Resource** dialog box appears.
8. Select a table. You can search for it or navigate to it.
  - Navigate to the Greenplum table that you want to import and click **OK**.
  - To search for a table, enter the name or the description of the table you want to add. Click **OK**.
9. If required, add more tables to the Greenplum data object.

You can also add tables to a Greenplum data object after you create it.

10. Click **Finish**.  
The data object appears under Data Objects in the project or folder in the **Object Explorer** view.

## Creating a Data Object Operation Write Operation

You can create the data object write operation for one or more Greenplum data objects. You can add a Greenplum data object write operation to a mapping as a target.

1. Select the data object in the **Object Explorer** view.
2. Right-click and select **New > Data Object Operation**.  
The **Data Object Operation** dialog box appears.
3. Enter a name for the data object operation.
4. Select **Write** as the type of data object operation.
5. Click **Add**.  
The **Select Resources** dialog box appears.
6. Select the Greenplum table for which you want to create the data object write operation and click **OK**.  
You can select only one data object to a data object write operation.
7. Click **Finish**.

The Developer tool creates the data object write operation for the selected data object.

## CHAPTER 5

# Greenplum Mappings

This chapter includes the following topics:

- [Greenplum Mappings Overview, 24](#)
- [Hadoop Validation and Run-time Environments, 24](#)
- [Greenplum Mapping Example, 25](#)

## Greenplum Mappings Overview

After you create a Greenplum data object write operation, you can develop a mapping.

You can create an Informatica mapping that contains objects such as a relational or flat file data object operation as the input. You can add transformations and a Greenplum data object write operation as the output to load data to Greenplum tables.

Validate and run the mapping. You can deploy the mapping and run it, or you can add the mapping to a Mapping task in a workflow.

## Hadoop Validation and Run-time Environments

You can configure the mappings to run in the native or Hadoop run-time environment. When you run mappings in the native environment, the Data Integration Service processes the mapping. When you run a mapping in a Hadoop environment, the Data Integration Service can push mappings to a Hadoop cluster.

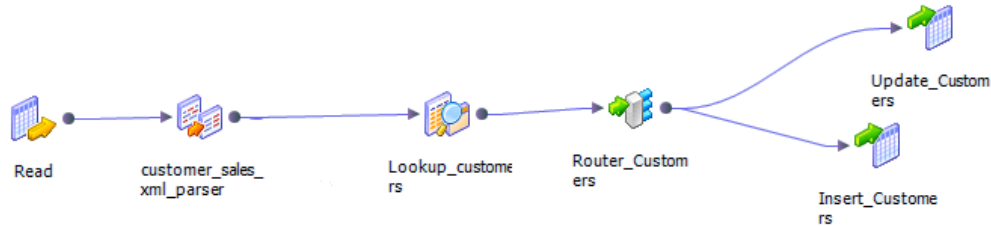
For more information about the Hadoop run-time and validation environment, see the *Informatica Big Data Management User Guide*.



# Greenplum Mapping Example

Your organization needs to sort sales information based on new and existing customers. Create a mapping that reads the sales information, segregates the sales information for new and existing customers, and loads specific data to the Greenplum table.

The following image shows the Greenplum mapping example:



The mapping contains an XML file as a source that contains sales information for your organization. The Data Processor transformation parses the XML file. The Lookup transformation looks up sales information based on the customer ID. The Router transformation routes the sales data based on new customers and existing customers. The mapping output contains two Greenplum data object write operations to write data to the Greenplum table.

## Mapping Input

The source for the mapping is an XML file stored in HDFS. The source contains information for your organization, including customer information, product information, and country information.

Source columns include columns for customer information, such as customer ID, first name, last name, gender, year of birth, email, phone number, and address.

## Transformations

You can add the following transformations to get sales information based on new customers and existing customers:

### Data Processor Transformation

The customer\_sales\_xml\_parser transformation parses the XML file.

### Lookup Transformation

The Lookup\_customer transformation looks up sales information based on the customer ID. The lookup table has the same structure as the source.

You use the following lookup condition:

```
customerid=customerid1
```

### Router Transformation

The Router\_Customers transformation routes the sales data for new customers to the group called Insert\_Customers and routes the sales data for existing customers to the group called Update\_Customers.

## Mapping Output

The mapping output is a Greenplum data object. To write data to the Greenplum table, create a data object write operation called Insert\_Customers for new customers and a data object write operation called Update\_Customers for existing customers.

The target columns include customer ID, first name, last name, gender, year of birth, email, phone number, and address.

When you run the mapping, the Data Integration Service loads the sales information to the Greenplum table. Analysts can run queries and generate reports based on the data in the Greenplum table.

You can configure to run the mapping either in the native or Hadoop environment.

## CHAPTER 6

# Greenplum Run-time Processing

This chapter includes the following topics:

- [Greenplum Run-time Processing Overview, 27](#)
- [Match and Update Columns, 27](#)
- [Error Handling for Greenplum Targets, 28](#)
- [Parameterization, 29](#)
- [Partitioning, 30](#)

## Greenplum Run-time Processing Overview

When you develop a Greenplum mapping, you define data object operation write properties that determine how data can be loaded to Greenplum targets.

The data object operation write properties you can set include configuring the match and update columns and setting the error limit.

## Match and Update Columns

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

The Data Integration Service matches rows based on the list of column names you specify in the advanced properties of the data object write operation. The Data Integration Service updates columns based on the list of column names you specify in the advanced properties of the data object write operation.

You can configure the following data object write operation properties for match and update columns:

### **Match Column**

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

## Match Columns

You can specify the columns to use for the update of rows in the target. 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 you configure update or merge as the method to process data from the named pipe.

## Update Columns

You can specify the columns to update for the rows that meet the criteria of the update column property. You must specify update columns that are not used for the Greenplum distribution key for the table. You must specify the update columns if you configure update or merge as the method to process data from the named pipe.

# 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 mapping. You can set the error limit in the Advanced properties of the data object write operation. 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 mapping 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.

Consider the following behavior for error tables:

- If the table does not exist, the gpload utility creates the table based on the format predefined in the Greenplum database.
- If the specified table exists in the schema, but the table is not in the format predefined in the Greenplum database, the mapping fails.
- If a mapping fails, see the error table for more information about the errors.
- If you run the mapping 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 Data 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 location of the temporary directories of the Data Integration Service process node.

# Parameterization

You can parameterize Greenplum data object operation properties to override the write data object operation properties during run time.

You can parameterize the following data object write operation properties for Greenplum targets:

- Load Method
- Match Columns
- Update Columns
- Update Condition
- Format
- Delimiter
- Escape
- Skip Escaping
- Null AS
- Quote
- Error Limit
- Error Table
- Greenplum Pre SQL
- Greenplum Post SQL
- Truncate Target Option
- Reuse Table
- Greenplum Schema Name
- Greenplum Target Table
- Greenplum Loader Logging Level
- Greenplum gpfdist Timeout
- Windows Pipe Buffer Size
- Delete Control File
- Gpload Log File Location
- Gpload Control File Location
- Encoding
- Pipe Location
- Port[Range]
- Max Line Length
- Treat Null as Empty String

The following attributes support partial parameterization:

- Update Condition
- Greenplum Pre SQL
- Greenplum Post SQL
- Gpload Log File Location

- Gpload Control File Location
- Pipe Location

## Partitioning

When a mapping that is enabled for partitioning contains a Greenplum data object as a target, the Data Integration Service can use multiple threads to write to the target.

You can configure dynamic partitioning for Greenplum data object operations. The Data Integration Service processes partitions concurrently.

To enable partitioning, administrators and developers perform the following tasks:

**Administrators set maximum parallelism for the Data Integration Service to a value greater than 1 in the Administrator tool.**

Maximum parallelism determines the maximum number of parallel threads that process a single pipeline stage. Administrators increase the **Maximum Parallelism** property value based on the number of CPUs available on the nodes where mappings run.

**Developers can set a maximum parallelism value for a mapping in the Developer tool.**

By default, the **Maximum Parallelism** property for each mapping is set to Auto. Each mapping uses the maximum parallelism value defined for the Data Integration Service.

Developers can change the maximum parallelism value in the mapping run-time properties to define a maximum value for a particular mapping. When maximum parallelism is set to different integer values for the Data Integration Service and the mapping, the Data Integration Service uses the minimum value of the two.

When you run a Greenplum data object write operation, the Data 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 writes data into the Greenplum target.

# APPENDIX A

## Data Type Reference

This appendix includes the following topics:

- [Data Type Reference Overview, 31](#)
- [Greenplum, JDBC, and Transformation Data Types, 31](#)

### Data Type Reference Overview

When you import Greenplum tables as a data object and create a data object write operation, the JDBC data types corresponding to the Greenplum data types appear in the Developer tool.

The Data Integration Service writes the data as JDBC data types. PowerExchange for Greenplum 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.

### Greenplum, JDBC, and Transformation Data Types

The following table lists the Greenplum data types that Data Integration Service supports and the corresponding JDBC and transformation data types:

Greenplum Data Type	JDBC Data Type	Transformation Data Type
Bigint	BIGINT	Bigint
Bigserial	BIGINT	Bigint
Boolean	BOOLEAN	String
Character	CHARACTER	String
Character varying	CHARACTER VARYING	String
Date	DATE	Date/Time
Numeric	NUMERIC	Decimal

Greenplum Data Type	JDBC Data Type	Transformation Data Type
Double precision	DOUBLE PRECISION	Double
Integer	INTEGER	Integer
Real	REAL	Double
Serial	INTEGER	Integer
Smallint	INTEGER	Integer
Text	TEXT	Text
Time	TIME	Date/Time
Timestamp	TIMESTAMP	Date/Time

**Note:** The Developer tool converts the unsupported data types to CHARACTER VARYING data types with a precision of zero.



# INDEX

## A

advanced properties  
input [18](#)

## C

creating  
data object write operation [23](#)  
data object writer operation  
creating [23](#)

## D

data object overview [16](#)  
data object properties  
Greenplum  
data object properties [17](#)  
data types  
Greenplum [31](#)  
ODBC [31](#)  
datatype reference overview  
description [31](#)

## E

environment variables [11](#)

## G

general properties  
input [18](#)  
greenplum  
importing a data object [22](#)  
Greenplum  
prerequisites [10](#)  
Greenplum connections  
properties [14](#)  
Greenplum data object  
overview [16](#)  
Greenplum mapping example [25](#)  
Greenplum mappings [24](#)

## H

Hadoop validation [24](#)

## I

importing  
Greenplum data object [22](#)  
input properties [17](#)  
Introduction to Greenplum [8](#)

## M

mapping example [25](#)  
match columns [27](#)

## O

overview  
Greenplum data object [16](#)

## P

prerequisites  
Greenplum [10](#)  
Hadoop environment [11](#)  
Native environment [10](#)  
prerequisites for Hadoop environment [11](#)  
prerequisites for Native environment [10](#)

## R

run-time environments [24](#)

## S

SSL authentication  
overview [13](#), [24](#)

## U

update columns [27](#)