



Informatica® PowerExchange for Greenplum
10.5.9

User Guide for PowerCenter

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

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

Informatica, the Informatica logo, PowerCenter, 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. Required third party notices are included with the product.

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

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

Table of Contents

Preface	5
Informatica Resources.	5
Informatica Network.	5
Informatica Knowledge Base.	5
Informatica Documentation.	5
Informatica Product Availability Matrices.	6
Informatica Velocity.	6
Informatica Marketplace.	6
Informatica Global Customer Support.	6
 Chapter 1: Introduction to PowerExchange for Greenplum.....	 7
PowerExchange for Greenplum Overview.	7
 Chapter 2: PowerExchange for Greenplum Configuration	 8
PowerExchange for Greenplum Configuration Overview.	8
Prerequisites.	8
Configuring PowerExchange for Greenplum on Windows.	8
Configuring PowerExchange for Greenplum on Linux, AIX, and SUSE.	9
Registering the Plug-in.	9
Upgrading PowerExchange for Greenplum from Version 10.0 or Later.	10
 Chapter 3: Greenplum Sources and Targets.....	 11
Greenplum Sources and Targets Overview.	11
Importing Greenplum Source and Target Definitions.	11
 Chapter 4: Greenplum Sessions and Workflows.....	 13
Greenplum Sessions and Workflows.	13
SSL Authentication for Greenplum Targets.	13
Configuring a PowerExchange for Greenplum Connection.	14
PowerExchange for Greenplum Connections.	14
Session Configuration with a Greenplum Source.	15
Session Configuration with a Greenplum Target.	17
Guidelines for Configuring Greenplum Session Properties.	20
Match and Update Columns.	21
Partitioning.	21
Parameterization.	22
Error Handling for Greenplum Targets.	22
 Appendix A: Greenplum Datatype Reference.....	 23
Greenplum and ODBC Data Types.	23

Rules and Guidelines for Greenplum Text Data Type.	24
Index.	25

Preface

Use the *Informatica® PowerExchange® for Greenplum User Guide* to learn how to read from or load data to Greenplum Database by using the PowerCenter Client. Learn to create a Greenplum connection, develop mappings, and run sessions in an Informatica domain.

Informatica Resources

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

Informatica Network

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

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

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

Informatica Knowledge Base

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

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

Informatica Documentation

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

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

Informatica Product Availability Matrices

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

Informatica Velocity

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

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

Informatica Marketplace

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

Informatica Global Customer Support

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

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link:

<https://www.informatica.com/services-and-training/customer-success-services/contact-us.html>.

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

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. You can use PowerExchange for Greenplum to read from or load data into the Greenplum distributed storage system.

When you run a Greenplum session to read data, the PowerCenter Integration Service invokes the Greenplum database parallel file server, gpfdist, which is Greenplum's file distribution program, to read data.

When you run a Greenplum session to write data, 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 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.

CHAPTER 2

PowerExchange for Greenplum Configuration

This chapter includes the following topics:

- [PowerExchange for Greenplum Configuration Overview, 8](#)
- [Prerequisites, 8](#)
- [Registering the Plug-in, 9](#)
- [Upgrading PowerExchange for Greenplum from Version 10.0 or Later, 10](#)

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. Verify that you have read, write, and execute permissions on the following directories:

`<PowerCenter Installation Directory>/server/bin`

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

Configuring PowerExchange for Greenplum on Windows

Before you can use PowerExchange for Greenplum to read or write data from a Windows machine, you must perform the following tasks:

1. Go to the Pivotal Greenplum website: [?>](#)

2. Download and install the supported version of the software on the Windows machine where the PowerCenter Integration Service runs. For example,
 - Python 2.5.4 32-bit.
 - 5.18 Greenplum Clients software for Windows
 - 5.18 Greenplum loaders software for Windows.
3. Set the environment variables:
 - a. Set the GPHOME_LOADERS environment variable to the folder that contains the Greenplum loader libraries:


```
GPHOME_LOADERS = C:\Program Files (x86)\Greenplum\greenplum-loaders-5.18.0
```
 - b. Set the GPHOME_CLIENTS environment variable to the folder that contains the Greenplum clients libraries:


```
GPHOME_CLIENTS=C:\Program Files (x86)\Greenplum\greenplum-clients-5.18.0\
```
 - c. Set the PATH environment variable to the folder that contains the Greenplum clients, Greenplum loaders, Python libraries:


```
Path=C:\Python25;C:\Program Files (x86)\Greenplum\greenplum-loaders-5.18.0\lib;C:\Program Files (x86)\Greenplum\greenplum-loaders-5.18.0\bin;C:\Program Files (x86)\Greenplum\greenplum-clients-5.18.0\lib;C:\Program Files (x86)\Greenplum\greenplum-clients-5.18.0\bin;
```
 - d. Set the PYTHONPATH environment variable to the folder that contains the Greenplum loader libraries: :


```
PYTHONPATH=%GPHOME_LOADERS%\bin\lib
```
4. Restart both the PowerCenter Repository Service and the PowerCenter Integration Service after you update the environment variables.

Configuring PowerExchange for Greenplum on Linux, AIX, and SUSE

Before you can use PowerExchange for Greenplum to read or write data from a Linux, AIX, or SUSE machine, you must perform the following tasks:

1. 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.
2. Verify that you can connect to Greenplum Database with the gpload utility.
3. Verify that the PATH, GPHOME_LOADERS, PYTHONPATH, and LIBRARY_PATH environment variables are set after you install the loaders package.
4. Restart both the PowerCenter Repository Service and the PowerCenter Integration Service after you update the environment variables.

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.

Upgrading PowerExchange for Greenplum from Version 10.0 or Later

When you upgrade PowerExchange for Greenplum from version 10.0 or later, you must update the existing plug-in file registration. If you have an existing Greenplum repository, and you want to upgrade to 10.4.0, you must update the `GreenplumConnector.xml` plugin to use the Greenplum reader.

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 Sources and Targets

This chapter includes the following topics:

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

Greenplum Sources and Targets Overview

Greenplum source and target definitions represent metadata for Greenplum tables.

You can create a mapping with a Greenplum source to extract data from Greenplum. You can create a mapping with any source and a Greenplum target to load data to Greenplum. When the PowerCenter Integration Service extracts data from the source or loads data to the target, it converts the data based on the data types associated with the source or the target.

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 source or target definitions, the ODBC data types corresponding to the Greenplum data types appear in the Designer.

Importing Greenplum Source and Target Definitions

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

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

1. Open a source or target folder.
2. Select Source Analyzer or Target Designer based on what you want to import the object as.
3. In the Source Analyzer, click **Sources > Import from Database**.
4. In the Target Designer, click **Targets > Import from Database**.
5. Select the Greenplum data source used to connect to the source or 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.
6. 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 source or target, specify the owner name.

7. Drill down through the list of database objects to view the available tables as sources or targets.

8. 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.

9. Click **OK**.

The selected source or target definitions appear in the Navigator under the Sources or Targets node.

CHAPTER 4

Greenplum Sessions and Workflows

This chapter includes the following topics:

- [Greenplum Sessions and Workflows, 13](#)
- [SSL Authentication for Greenplum Targets, 13](#)
- [Configuring a PowerExchange for Greenplum Connection, 14](#)
- [Session Configuration with a Greenplum Source, 15](#)
- [Session Configuration with a Greenplum Target, 17](#)
- [Partitioning, 21](#)
- [Parameterization, 22](#)
- [Error Handling for Greenplum Targets, 22](#)

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 reads or writes data to Greenplum. When you run a session, the PowerCenter Integration Service invokes the gpload utility to read from or 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 connection information.
7. Click **OK**.
The database connection appears in the **Connection Browser** list.

Note: If the session contains a CDC real-time source, use an ODBC connection to load data to Greenplum. The Greenplum connection is not supported for CDC sources.

PowerExchange for Greenplum Connections

When you configure a Greenplum connection, you define the connection attributes that the PowerCenter Integration Service uses to connect to Greenplum.

The following table describes the Greenplum connection properties that you must configure:

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.

Connection Attribute	Description
Database	Name of the database.
Enable SSL	Select this option to establish secure communication between the gpload utility and the Greenplum server over SSL. Applicable for Greenplum connections used for loading data to Greenplum. Not applicable for Greenplum reader. Not applicable for Greenplum writer on the Windows platform.
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.

Session Configuration with a Greenplum Source

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

You can configure the following session properties for Greenplum sources:

Delimiter

Separates successive input fields.

Default format is text. For data in the text format, use any 7-bit ASCII value except a-z, A-Z, and 0-9.

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.

Null Value

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).

Escape Character

Character that treats special characters in the data as regular characters.

Default format is text. In the text format, special characters comprise delimiter and escape characters.

You can improve the task performance if the data does not contain escape characters.

GPFDist Port

The HTTP port on which gpfdist serves files.

Default is 39871.

GPFDist Maximum Line Length

The integer that specifies the maximum length of a line in the XML transformation data that is passed to gpload. Default is 32768.

Valid range is 32 K to 256 MB, with the upper limit at 1 MB on Windows operating systems.

Encoding

Character set encoding of the source data. UTF-8 character set encoding is supported.

GPFDist Logging

Sets the logging level for the gpload utility.

You can specify from the following logging options:

- None
- Verbose. Shows the progress and status messages.
- Very Verbose. Shows all the output messages generated by this utility.

Greenplum Pre SQL

The SQL command to run before the Integration Service reads data from the source.

Greenplum Post SQL

The SQL command to run after the Integration Service writes to the target.

Pipe Directory Path

The file system location where the pipes that are used for data transfer are created.

Default is `$PMRootDir/Temp`

GPFDist Log File

The fully qualified path and log file name for the standard output message logs.

Default is `INFAGPFDistLogfile.out`.

User Defined Join

Specifies the condition used to join data from multiple sources represented in the same Source Qualifier transformation.

Number of Sorted Ports

Indicates the number of columns used when sorting rows queried from relational sources. If you select this option, the Integration Service adds an ORDER BY to the default query when it reads source rows. The ORDER BY includes the number of ports specified, starting from the top of the transformation. When selected, the database sort order must match the session sort order.

Tracing Level

Sets the amount of detail included in the session log when you run a session containing this transformation.

You can choose Terse, Normal, Verbose Initialization, or Verbose Data tracing level. Default is Normal.

Select Distinct

Specifies if you want to select only unique rows. The Integration Service includes a SELECT DISTINCT statement if you choose this option.

Pre SQL

The pre-session SQL commands to run against the source database before the Integration Service reads the source.

Post SQL

The post-session SQL commands to run against the source database after the Integration Service writes to the target.

SQL Query

Defines a custom query that replaces the default query the Integration Service uses to read data from sources represented in this Source Qualifier transformation. A custom query overrides entries for a custom join or a source filter.

Source Filter

Specifies the filter condition the Integration Service applies when querying rows.

Owner Name

The schema name in Greenplum.

Source Table Name

Overrides the default Greenplum source table name.

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.

Note: For Greenplum server versions 5.0 and later, LOG_ERRORS are generated instead of an error table. For more information, see the Greenplum documentation.

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 unichar 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

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. Partitioning is not supported for the Greenplum reader.

You can configure pass-through partitioning for Greenplum target 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 topics:

- [Greenplum and ODBC Data Types, 23](#)
- [Rules and Guidelines for Greenplum Text Data Type, 24](#)

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

Greenplum Data Type	ODBC Data Type	Transformation Data Type
Text	Text	Text
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*.

Rules and Guidelines for Greenplum Text Data Type

The Text data type is imported with a default precision of 104857600. If required, you must manually reduce the precision size from 104857600 to 8190 while creating the mapping.

INDEX

C

- configuration
 - session [15](#), [17](#)
- configuring
 - connection [14](#)
- connection
 - configuring [14](#)

E

- error handling [22](#)

G

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

P

- plug-ins
 - registering PowerExchange for Greenplum [9](#)
- PowerExchange for Greenplum connections
 - properties [14](#)
- prerequisites [8](#)

S

- session
 - configuration [15](#), [17](#)
- SSL authentication
 - overview [13](#)