



---

PowerExchange for Greenplum

Version 3.0.1.1

User Guide

Copyright © 2011 EMC Corporation

All rights reserved

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED "AS IS." EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com

All other trademarks used herein are the property of their respective owners.

# Table of Contents

<b>Preface</b>	<b>4</b>
<i>About This Guide</i>	4
<i>Getting Support</i>	4
Product information	4
Technical support	4
<b>Understanding PowerExchange for Greenplum</b>	<b>5</b>
<i>Overview</i>	5
<i>Integrating PowerCenter with Greenplum</i>	5
<b>Configuring Greenplum Sessions</b>	<b>7</b>
<i>Overview</i>	7
<i>Configuring Relational Connections</i>	7
<i>Configuring Sessions for Greenplum Targets</i>	9
Guidelines for Configuring Session Properties	14
<i>Error Handling for Greenplum Targets</i>	15
Session Log	15
<i>Troubleshooting</i>	16
<b>Loading Data Into Greenplum Tables – An Example</b>	<b>17</b>
<i>Overview</i>	17
<i>Loading Data into the Greenplum Database</i>	17
Importing Source and Target Definitions	18
Creating Mappings	18
Creating and Configuring Sessions	19
Running Workflows	19
<b>Data Types Appendix</b>	<b>21</b>
<i>Greenplum and ODBC Data Types</i>	21
<i>Unsupported Greenplum Data Types</i>	23
Time with Time Zone or Timestamp with Time Zone	23
<i>Deprecated Greenplum Data Types</i>	24
Money	24
<b>Error Messages Appendix</b>	<b>25</b>
<i>GPWRT Error Messages</i>	25

# Preface

## About This Guide

---

PowerExchange for Greenplum User Guide provides information about installing, configuring, and using PowerExchange for Greenplum to load data into the Greenplum database. This guide is written for database administrators and developers who are responsible for loading data into the Greenplum database. This guide assumes knowledge of Greenplum database and `gpload` utility, PowerCenter, database management systems, and structured query language (SQL).

For information about supported platforms, installation tasks, and upgrade steps, see the *PowerExchange for Greenplum Release Notes*.

## Getting Support

---

EMC support, product, and licensing information can be obtained as follows.

### Product information

For documentation, release notes, software updates, or for information about EMC products, licensing, and service, go to the EMC Powerlink website (registration required) at:

<http://Powerlink.EMC.com>

### Technical support

For technical support, go to [Powerlink](#) and choose **Support**. On the Support page, you will see several options, including one for making a service request. Note that to open a service request, you must have a valid support agreement. Please contact your EMC sales representative for details about obtaining a valid support agreement or with questions about your account.

# Understanding PowerExchange for Greenplum

# 1

## Overview

---

PowerExchange for Greenplum integrates PowerCenter with the Greenplum database by leveraging the bulk loading capabilities of the Greenplum `gpload` utility. The Integration Service invokes the `gpload` utility to load data into Greenplum targets. Greenplum targets represent tables in the Greenplum database system.

## Integrating PowerCenter with Greenplum

---

Before loading data into a Greenplum table, you must import a target definition. Target definitions represent metadata or schema of the Greenplum tables. When you import a target definition, you define the structure of the Greenplum tables in the PowerCenter repository. The PowerCenter repository stores the metadata that describes the schema of the Greenplum tables and not the actual data that is stored in the tables. When you import a target definition, the Designer imports details such as the name of the target, name of the columns, and data types. After you import a target definition, you can use this target to create a mapping.

To import target definitions, the Designer connects to the Greenplum database system through ODBC. When you import Greenplum target definitions, you import metadata about Greenplum tables. For more information about importing target definitions, see *“Working with Targets”* in the *PowerCenter Designer Guide*.

In the Source Analyzer, import source definitions to import the schema for the source. When you import source definitions, the Designer imports details such as the name of the file or table, name of the columns, and data types. For example, if the source data to be loaded into a Greenplum table is in a relational table, import the relational table as a source definition. For more information about importing source definitions, see *“Working with Sources”* in the *PowerCenter Designer Guide*.

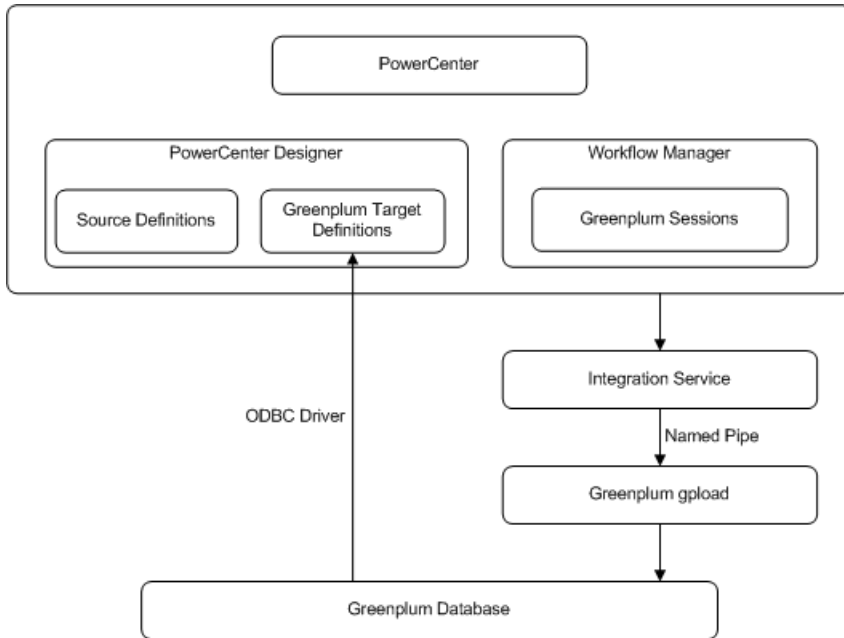
Use these source and target definitions to create mappings in the Mapping Designer. You can use PowerCenter transformations in the mappings to perform operations on the data for implementing business logic and processes. After you create a mapping, create a session in the Workflow Manager. When you create a session, you select the mapping that you want to use in the session.

After you create a session, create Greenplum Connection relational objects to define connections to the Greenplum database. For more information about creating Greenplum Connection relational objects, see [Configuring Relational Connections](#).

The connection object stores information such as the Greenplum user ID, password, and database server name. When you run a session, the `gpload` utility uses these Greenplum Connection relational objects to connect to the Greenplum database. After processing the data, the Integration Service writes data to a named pipe and invokes the `gpload` bulk loading utility. The `gpload` utility reads data from the named

pipe and loads data into Greenplum database tables. For more information about the Greenplum gpload utility, see the *Greenplum Database Administrator Guide*.

Figure 1 shows how PowerCenter integrates with Greenplum:



**Figure 1 PowerCenter Integration with Greenplum**

# Configuring Greenplum Sessions

# 2

## Overview

After you create mappings in the Designer, create sessions in the Workflow Manager. See “Sessions” in the *PowerCenter Workflow Basics Guide* for more information on creating a session. This section only describes the additional session configurations for Greenplum Database.

When you configure a session, you create relational connections to access the Greenplum database. You can also define session properties for Greenplum targets to determine how the `gpload` utility writes data into the Greenplum database.

## Configuring Relational Connections

Before you run a Greenplum session, configure a relational connection for the Greenplum targets in the Workflow Manager. When you configure a relational connection, you specify the connection attributes that the `gpload` utility uses to connect to a Greenplum database.

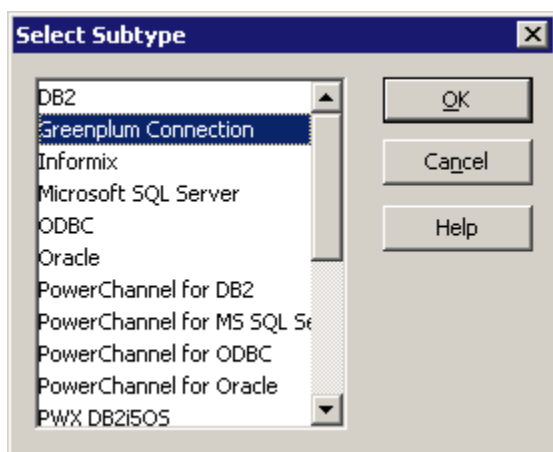
**To configure a Greenplum relational connection:**

1. In the Workflow Manager, click **Connections > Relational**.

The Relational Connection Browser dialog box appears.

2. Click **New**.

The Select Subtype dialog box appears.



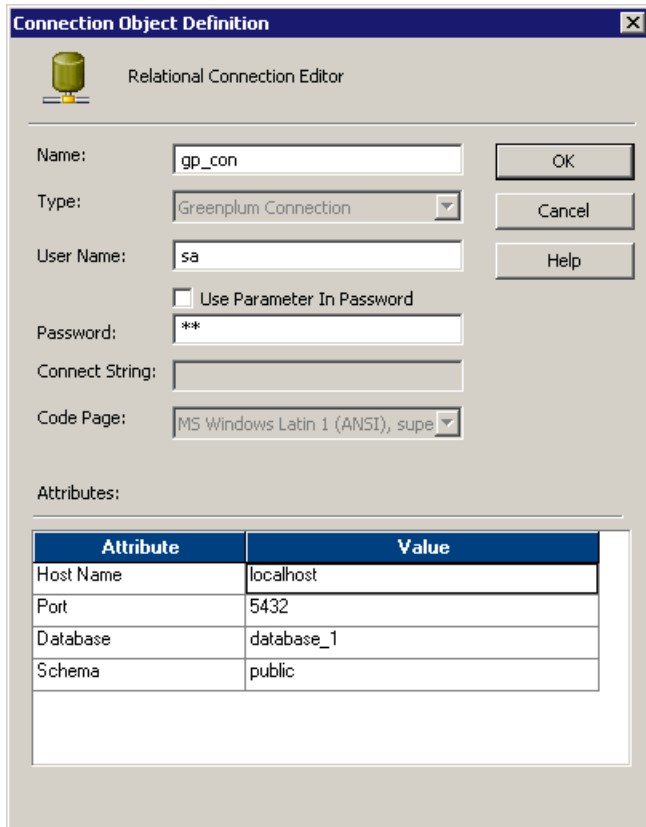
3. Select **Greenplum Connection** from the list.

**Note:** The Greenplum Connection entry is added when you register the plug-in for PowerExchange for Greenplum.



- Click **OK**.

The Connection Object Definition dialog box appears.



**Connection Object Definition**

Relational Connection Editor

Name:  OK

Type:  Cancel

User Name:  Help

☐ Use Parameter In Password

Password:

Connect String:

Code Page:

Attributes:

Attribute	Value
Host Name	localhost
Port	5432
Database	database_1
Schema	public

- Enter the following information:

Connection Attribute	Required/Optional	Description
Name	Required	Name for the Greenplum relational connection.
User Name	Required	User name with appropriate permissions to access the Greenplum database.
Password	Required	Password to connect to the Greenplum database. You cannot use a parameter to specify the password.
Use Parameter in Password*	Not Applicable	Indicates that the password is parameterized.
Host Name	Optional	Host name or IP address of the computer configured to run the Greenplum server. If you do not specify the host name, the <code>gpload</code> utility reads from the environment variable <code>\$PGHOST</code> . Default is <b>localhost</b> .



Port	Optional	Greenplum server port number. If you enter 0, the gpload utility reads from the environment variable \$PGPORT. Default is <b>5432</b> .
Database	Required	Name of the database.
Schema	Required	Name of the schema that contains metadata for Greenplum targets. Default is <b>public</b> .
* This option is currently not supported by PowerExchange for Greenplum.		

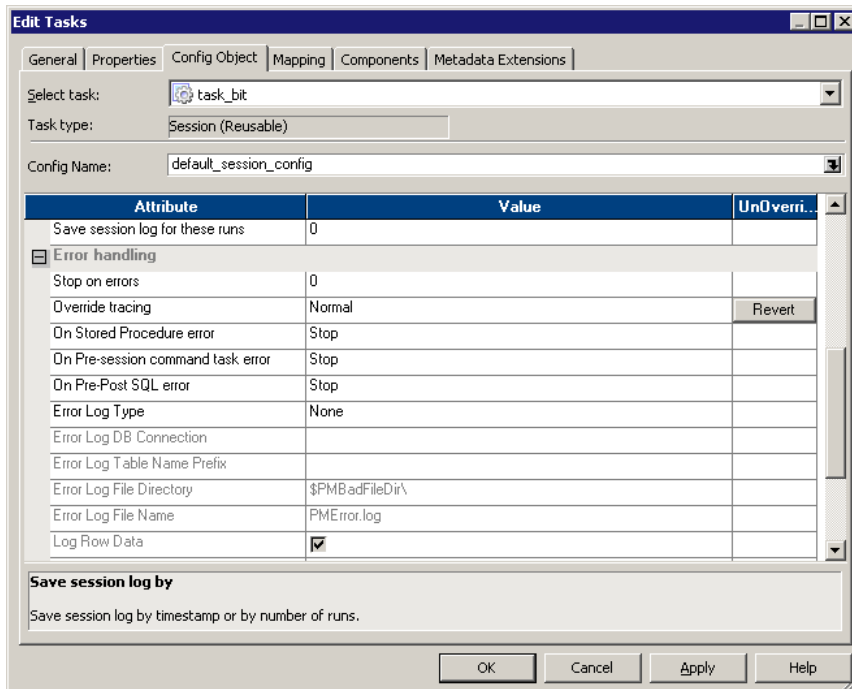
6. Click **OK**.

## Configuring Sessions for Greenplum Targets

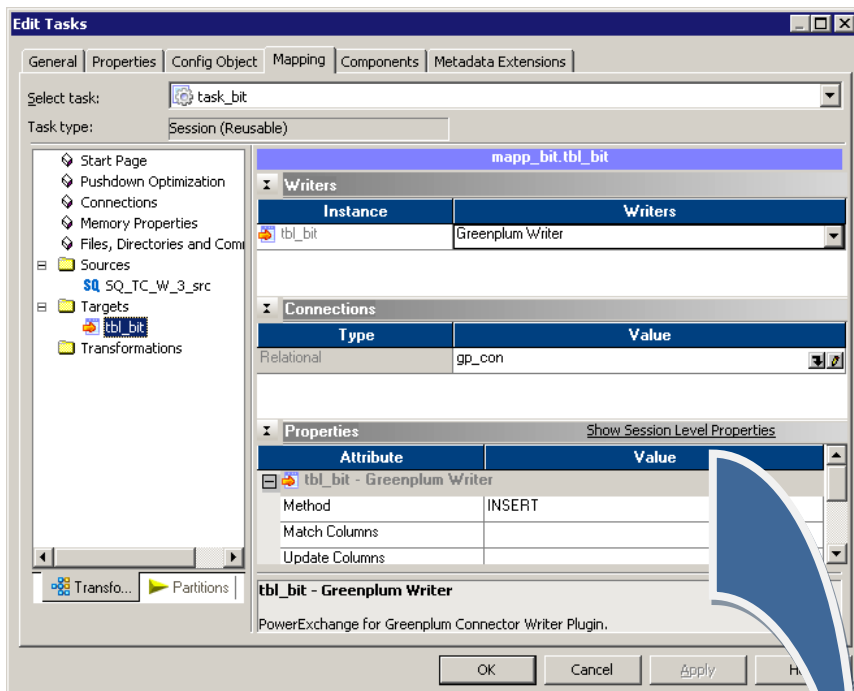
You can configure a session to bulk load into the Greenplum database through the gpload utility. The gpload utility uses the connection attributes that you configured in the Greenplum Connection object to connect to the Greenplum database. For more information about configuring the Greenplum Connection object, see [Configuring Relational Connections](#). You can also define session properties for Greenplum targets to determine how the gpload utility writes data into the Greenplum database. When you run a session, the Integration Service invokes the gpload utility to load data into the Greenplum database.

### To configure a session for a Greenplum target:

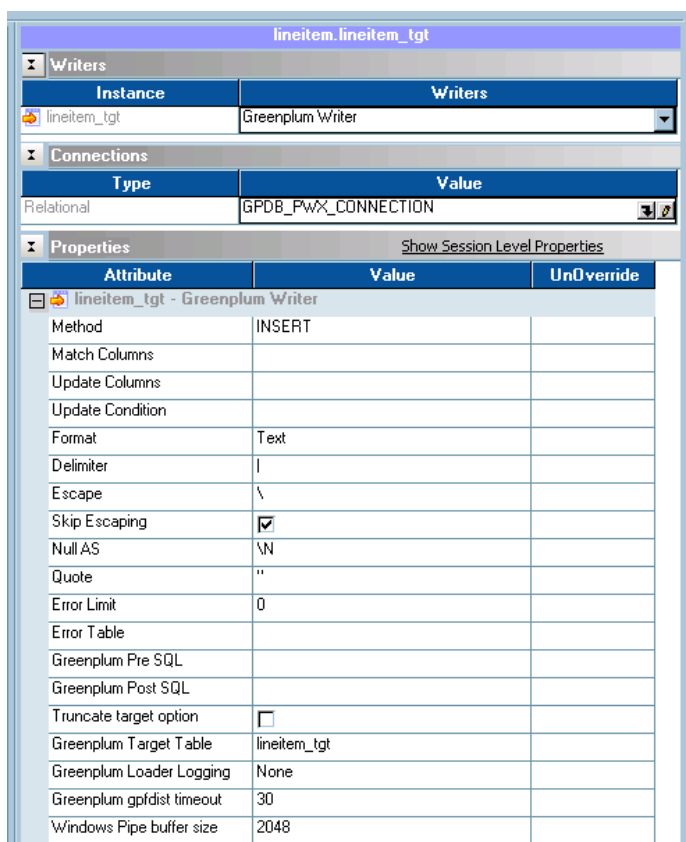
1. In the Workflow Manager, from the Repository Navigator connect to the repository.
2. Select the folder that contains the PowerCenter objects stored in the repository.
3. Click Tools > Task Developer.
4. To open the session properties, double-click the session.  
The Edit Tasks dialog box appears.
5. Click the Config Object tab.



6. To set the error threshold in the Error Handling section, configure the `Stop on errors` property. The value for the `Stop on errors` property must be in the range of 0 to 2,147,483,647. For more information about configuring `Stop on errors` property, see [Error Handling for Greenplum Targets](#).
7. Optionally, to set the amount of detail written to the session log, in the Error Handling section, set the `Override Tracing` property.  
You can set the tracing level to None, Terse, Normal, Verbose Initialization, or Verbose Data. By default, the tracing level is set to None. You can change the tracing level to a verbose setting when you need to log additional details for troubleshooting. For more information about the tracing level, see "Session and Workflow Logs" in the *PowerCenter Workflow Basics Guide*.
8. Click the Mapping tab.



## Edit Tasks – Mapping Tab



## Target Properties

9. Expand the Targets node, and select the target instance that you imported.
10. Under the Writers area, change the Writers field to the Greenplum Writer.  
**Note:** The Greenplum Writer entry is added when you register the plug-in for PowerExchange for Greenplum.
11. Under the Connections area, specify the Greenplum Connection object that you created in the [Configuring Relational Connections](#) section.
12. Under the Properties area, configure the following session property attributes for the Greenplum target:

Property Attribute	Description
Method	<p>Indicates the operations that the <code>gpload</code> utility performs on the source data that is passed by the Integration Service:</p> <ul style="list-style-type: none"> <li>▪ INSERT. Inserts rows into the target.</li> <li>▪ UPDATE. Updates rows into the target.</li> <li>▪ MERGE. If the rows exist in the target, updates the existing rows. If the rows do not exist in the target, inserts rows in the target.</li> </ul>
Match Columns*	Matches rows based on the comma-separated list of column names. Enclose the column names in double quotes.
Update Columns*	Updates the columns specified in the comma-separated list of column names. Enclose the column names in double quotes.
Update Condition*	Updates a row based on the specified condition. The <code>gpload</code> utility determines whether to perform the update or merge operation based on the condition specified in the Update Condition property.
Format	<p>Writes data in a format that is compatible with the <code>gpload</code> utility. The Integration Service writes data in the following format:</p> <ul style="list-style-type: none"> <li>▪ Text. Writes data in text format. In text format, the Integration Service separates data using the character specified in the Delimiter session property. If the data contains characters which match the values specified in the Delimiter or Escape session property: <ul style="list-style-type: none"> <li>➤ Clear the Skip Escaping check box or</li> <li>➤ Specify values for the delimiter and escape characters that are not a part of the data.</li> </ul> </li> <li>▪ CSV. Writes data in CSV format. In CSV format, the Integration Service encloses data with the character specified in the Quote session property. The Integration Service also separates the data using the character specified in the Delimiter session property. If the data contains characters which match the values specified in the Quote or Escape session property: <ul style="list-style-type: none"> <li>➤ Clear the Skip Escaping check box or</li> <li>➤ Specify values for the quote and escape characters that are not a part of the data.</li> </ul> </li> </ul>

	<p>Default is Text.</p> <p><b>Note:</b> If the data contains newline characters, you must use CSV format. If you use Text format and the data contains newline characters, the data after the newline character is treated as a new record. In such situations, the <code>gpload</code> utility might reject or insert incorrect data into the tables.</p>
Delimiter	<p>Character that separates columns of data in a row.</p> <p>For data in text format, use any 7-bit ASCII value except a-z, A-Z, and 0-9. For data in CSV format, use any 7-bit ASCII value except \n, \r, and \\. Default is pipe ( ). You can also specify a non-printable ASCII character via an escape sequence using the decimal representation of the ASCII character. For example, \014 represents the shift out character.</p>
Escape	<p>Character that treats special characters in the data as regular characters. In text format, special characters include delimiter characters and escape characters. In CSV format, special characters include quote characters and escape characters.</p> <p>Use any 7-bit ASCII value as an escape character. Default is backslash (\).</p> <p><b>Note:</b> The session performance improves if the data does not contain escape characters.</p>
Skip Escaping	<p>Skips escaping special characters in data. Clear the Skip Escaping check box to treat special characters in data as regular characters.</p>
Null As	<p>String that represent 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).</p>
Quote	<p>Character that encloses data in CSV format. The Integration Service encloses data by the specified character and passes the data to the <code>gpload</code> utility. The quote character is ignored for data in text format. Use any 7-bit ASCII value that is not equal to the delimiter or null value. Default is double quotes (").</p>
Error Table	<p>Name of the error table where the <code>gpload</code> utility logs rejected rows when reading data that is processed by the Integration Service. The naming convention for the table name is &lt;schema name&gt;.&lt;table name&gt;, where schema name is the name of the schema that contains the table. For more information about the error table, see <a href="#">Error Handling for Greenplum Targets</a>.</p>
Error Limit	<p>For each Greenplum segment, the number of rows that the <code>gpload</code> utility discards or logs in the error table because of format errors. The <code>gpload</code> 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. For more information about the error limit, see <a href="#">Error</a></p>

	<a href="#">Handling for Greenplum Targets.</a>
Greenplum Pre SQL	An SQL command to execute pre-load. For example: INSERT INTO audit VALUES('start', current_timestamp)
Greenplum Post SQL	An SQL command to execute post-load. For example: INSERT INTO audit VALUES('end', current_timestamp)
Truncate Target Option	Select this option to truncate the Greenplum target table.
Greenplum Target Table	Overrides the Target Table name described in the mappings. The default is to use the target table defined in the session's mapping.
Greenplum Loader Logging	Sets the logging level for the gpload utility. The allowed values are: None - Minimum logging. This is the default. Verbose - Moderate logging. Very Verbose - Maximum logging.
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 KBs) that will be allocated by the PowerExchange Connector to buffer data before writing to the Greenplum bulk loader. Enter a value between 1 and 131072. The default value is 2048, and is the recommended value to use for most applications. Adjusting this value to can sometimes improve system performance.  Note that this attribute only applies to Informatica servers running on Windows-based systems.
* Set these attributes when you perform the update or merge operations. These attributes are ignored when you perform the insert operation.	

#### Session Properties for Greenplum Targets

## Guidelines for Configuring Session Properties

Use the following guidelines when you configure the session properties:

- The delimiter, escape, and quote characters must have different values for a Greenplum target.
- Use the recommended values to specify the delimiter, escape, and quote characters. If you specify invalid values, gpload logs an error and the session fails. For more information about the recommended values for the delimiter, escape, and quote characters, see Session Properties for Greenplum Targets.
- To improve session performance, use the default values for the delimiter, escape, and quote characters. When you use the default values, the Integration Service only needs to scan the character and unichar data types for special characters, and can ignore the other data types.

For example, consider a Greenplum table that contains columns of type char, varchar, decimal, and integer. If the delimiter, escape, and quote characters are set to default values, the Integration Service scans the data only in char and varchar columns. The range of data values allowed in decimal and integer columns does not include the default values for the delimiter, quote, and escape characters. To improve session performance, the Integration Service ignores data in decimal and integer columns.

- If the column names specified in the Match Columns and Update Columns properties have leading or trailing spaces, the `gpload` utility logs an error and the session fails. When you import target definitions, the Designer strips leading and trailing spaces from the column names. The `gpload` utility cannot load data into the Greenplum database because the column names in the target definition do not match the column names in the Greenplum table.

## Error Handling for Greenplum Targets

You can set the error threshold for a session by configuring the *Stop on Errors* session property. In the *Stop on Errors* session property, enter the number of non-fatal errors that the Integration Service can encounter before it fails the session. The error threshold does not include database errors that the `gpload` utility might encounter when loading data into the Greenplum database tables. The default value is 0. By default, the Integration Service does not stop a session when it encounters non-fatal errors.

For a Greenplum segment, you can set the error limit 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 only 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.

The error limit does not include rows with constraint errors, such as primary key violations, not null constraints, or unique constraints. If a row contains constraint errors, the `gpload` utility fails the load operation and no rows are written to the target. If you specify an error table, the error limit must be set to a number greater than one.

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 with the predefined format. If the specified table exists in the schema, but the table is not in the expected format, the session fails. When 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 Database Administrator Guide*.

## Session Log

When a session fails, the Integration Service logs the session details and the error statistics for the Integration Service and the `gpload` utility in the session log. In the session properties, configure the *Override Tracing* property to define the level of information written to the session log. For more information about configuring the *Override Tracing* property, see “Session and Workflow Logs” in the *PowerCenter Workflow Basics Guide*.

In the Greenplum Target Load Summary section of the session log, PowerExchange for Greenplum displays statistics for the requested, applied, rejected, and affected rows. The number of rejected rows



includes the number of rows discarded by the Integration Service and the `gpload` utility. The Greenplum Target Load Summary section also includes details such as table name, target, and partition number.

Severity	Timestamp	Node	Thread	Message Code	Message
INFO	3/16/2009 6:18:08 PM	node01_ps4220	WRITER_1_1	GPWRT_34050	[INFO] Deinitializing the partition driver.
INFO	3/16/2009 6:18:11 PM	node01_ps4220	WRITER_1_1	GPWRT_34021	[INFO] Gpload succeeded.
INFO	3/16/2009 6:18:11 PM	node01_ps4220	WRITER_1_1	GPWRT_34034	[INFO] For more information, see the gpload log file: ../infra_shared/Temp/gpload_0_0_0_1237207682_14000.log
INFO	3/16/2009 6:18:11 PM	node01_ps4220	WRITER_1_1	GPWRT_34070	[INFO] Greenplum Target Load Summary
INFO	3/16/2009 6:18:11 PM	node01_ps4220	WRITER_1_1	GPWRT_34072	[INFO] Table: [tbl_bit] Target: [1] Group: [1] Partition: [1] Requested Rows [1], Affected Rows [1], Applied Rows [1], Rejected Rows [0]
INFO	3/16/2009 6:18:11 PM	node01_ps4220	WRITER_1_1	GPWRT_34071	[INFO] =====
INFO	3/16/2009 6:18:11 PM	node01_ps4220	WRITER_1_1	GPWRT_34003	[INFO] Deinitializing the partition driver.
INFO	3/16/2009 6:18:11 PM	node01_ps4220	WRITER_1_1	WRT_8035	Load complete time: Mon Mar 16 18:18:10 2009
					LOAD SUMMARY
					=====
					WRT_8036 Target: tbl_bit: Partition 1 (Instance Name: [tbl_bit])
					WRT_8038 Inserted rows - Requested: 1    Applied: 1    Rejected: 0    Affected: 1
INFO	3/16/2009 6:18:11 PM	node01_ps4220	WRITER_1_1	WRT_8043	*****END LOAD SESSION*****
INFO	3/16/2009 6:18:11 PM	node01_ps4220	WRITER_1_1	WRT_8006	Writer run completed.
INFO	3/16/2009 6:18:11 PM	node01_ps4220	MANAGER	PETL_24031	***** RUN INFO FOR TGT LOAD ORDER GROUP [1], CONCURRENT SET [1] ***** Thread [READER_1_1] created for [the read stage] of partition point [SQ_TC_W_3_src] has completed. The total run time was insufficient for any meaningful statistics. Thread [WRITER_1_1] created for [the write stage] of partition point [tbl_bit] has completed. Total Run Time = [2.484440] secs Total Idle Time = [0.000000] secs Busy Percentage = [100.000000]
INFO	3/16/2009 6:18:11 PM	node01_ps4220	MANAGER	PETL_24005	PETL_24005 Starting post-session tasks. : (Mon Mar 16 18:18:11 2009)
INFO	3/16/2009 6:18:11 PM	node01_ps4220	MANAGER	PETL_24029	PETL_24029 Post-session task completed successfully. : (Mon Mar 16 18:18:11 2009)
INFO	3/16/2009 6:18:11 PM	node01_ps4220	MAPPING	GPWRT_33002	[INFO] Deinitialized the group driver.
INFO	3/16/2009 6:18:11 PM	node01_ps4220	MAPPING	GPWRT_32003	[INFO] Deinitialized the target driver.
INFO	3/16/2009 6:18:11 PM	node01_ps4220	MAPPING	SDKS_38510	SDK target and group deinitialized with status [0].
INFO	3/16/2009 6:18:11 PM	node01_ps4220	MAPPING	GPWRT_31003	[INFO] Deinitialized the PowerExchange for Greenplum plug-in.

**Note:** The Integration Service displays statistics for requested, applied, rejected, and affected rows in the load summary and session load summary sections of the session log. However, these statistics might not be correct across different versions of PowerCenter.

## Troubleshooting

### The session fails when I insert default values in a Greenplum table.

To insert default values for a column in a Greenplum table, ensure that the column is not linked in the mapping. However, you must link the column in the mapping if you want to insert any value other than the default.

For more information about linking columns, see “Mappings” in the *PowerCenter Designer Guide*.

### I do not see the Greenplum Connection in the Select Subtype dialog box.

Verify that the `GreenplumConnector.xml` file is located in the following location:

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

If the XML file is not present in this location, install PowerExchange for Greenplum.

Also, verify that the plug-in is registered with the repository. For more information about registering the plug-in, see the *PowerCenter Administrator Guide* or the *PowerCenter Command Reference*.

For more information about troubleshooting tips, see the [Error Messages Appendix](#).

# Loading Data Into Greenplum Tables – An Example

# 3

## Overview

This chapter describes how to load data into a Greenplum table using PowerExchange for Greenplum. For example, the Greenplum database table where you want to load the data into is the employee table with the following columns:

Column Name	Data Type
EmpID	integer
Joining_Date	date
Contact_Details	varchar

The emp\_src.txt file contains the source data to be loaded into the Greenplum table, employee. The emp\_src is a flat file that contains details for each new employee who recently joined the company.

## Loading Data into the Greenplum Database

**To load the employee details into the employee table, complete the following steps:**

1. In the Source Analyzer, import the flat file as a source definition.
2. In the Target Designer, import the employee table as a target definition.
3. In the Mapping Designer, create a mapping.
4. In the Workflow Manager, create a session.
5. In the Task Developer, configure the session properties.
6. In the Workflow Manager, create and run the workflow.

The following table shows sample data that is to be inserted into the employee table:

EmpID	Joining_Date	Contact_Details
103	01/01/2009	408-734-1833
104	01/02/2009	408-412-1362

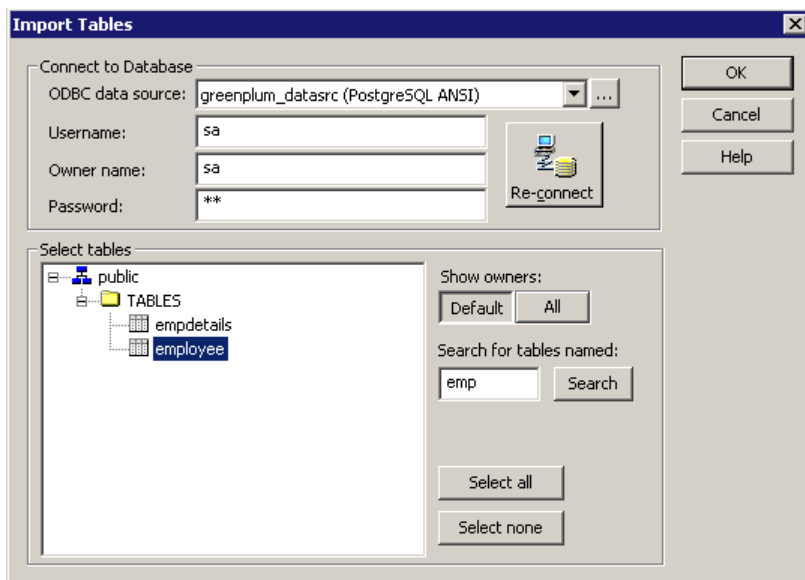
## Importing Source and Target Definitions

To create a source definition, you must import the source data. This source can be a flat file (CSV or XML format) or another database. In the PowerCenter Designer, click Tools > Source Analyzer. Click Sources > Import from File to import the flat file as a source definition.

For more information about importing flat files as source definitions, see “Working with Flat Files” in the *PowerCenter Designer Guide*.

To import the employee table schema as the target definition, you need to specify an ODBC data source to connect to the Greenplum database. From the Control Panel, use the Administrative Tools > Data Sources (ODBC) to create a data source name (DSN) on the computer where the PowerCenter Client is installed. Before creating the DSN for Greenplum, install the Greenplum ODBC driver available in the connectivity package, `greenplum-connectivity-3.2-build-<version>-WinXP-x86_32.msi`.

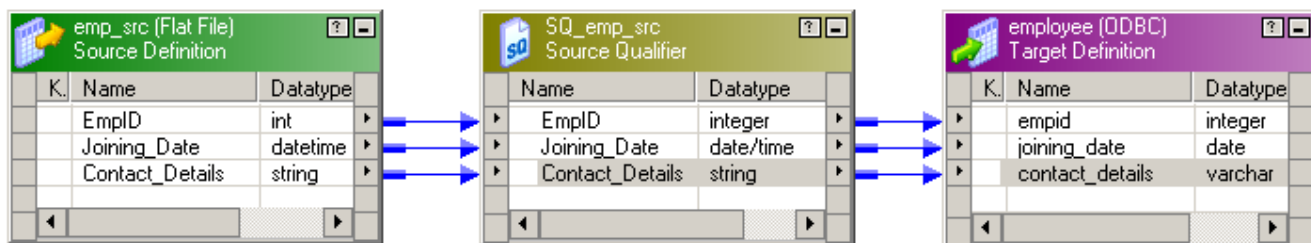
After you create the ODBC data source for Greenplum database, use the Designer to import the target definition. In the Target Designer, click Import from Database and specify the data source that you created using the Greenplum ODBC driver to connect to the Greenplum database. Import the employee table as the target definition.



For more information about importing source and target definitions, see the *PowerCenter Designer Guide*.

## Creating Mappings

Create a mapping by dragging the imported source and target definitions in the Mapping Designer.



This mapping contains the following objects:

- emp\_src - The source definition that defines the schema from the flat file, emp\_src.txt. To insert an employee record, provide source data for the EmpID, Joining\_Date, and Contact\_Details columns.
- SQ\_emp\_src - The Designer creates the Source Qualifier transformation when you add a flat file source definition to a mapping. For more information about the Source Qualifier transformation, see the *PowerCenter Transformation Guide*.
- Employee - The target definition that represents the employee table where the data is to be loaded.

For more information about creating mappings, see the *PowerCenter Designer Guide*.

## Creating and Configuring Sessions

In the Workflow Manager, click Tools > Task Developer to open the Task Developer. In the Task Developer, click Tasks > Create to create a session. For more information about creating sessions, see "Sessions" in the *PowerCenter Workflow Basics Guide*.

By default, the session is configured to use the Relational Writer and ODBC connection object. In the Task Developer, configure the session to use the Greenplum Writer and Greenplum Connection object. You can also configure session properties for the Greenplum target to determine how the gpload utility loads data into Employee table.

For more information about configuring sessions, see [Configuring Sessions for Greenplum Targets](#).

## Running Workflows

In the Workflow Manager, click Tools > Workflow Designer to open the Workflow Designer. In the Workflow Designer, click Workflows > Create to create a workflow. For more information about creating and running workflows, see the *PowerCenter Workflow Basics Guide*.

Add the session that you created to this workflow. Run the workflow to load data into the employee table. In the Workflow Monitor, right-click the session and select Get Session Log to view the session log.

Severity	Timestamp	Node	Thread	Message Co...	Message
INFO	3/17/2009 6:06:33 PM	node01_ps42	READER_1_1_1	DBG_21430	DBG_21430 Reading data from input source file [D:\Informatica\PowerCenter8.6.1\server\infra_shared\SrcFiles\emp_src.txt] : (Tue Mar 17 18:06:33 2009) Read [2] rows, read [0] error rows for source table [emp_src] instance name [emp_src]
INFO	3/17/2009 6:06:33 PM	node01_ps42	READER_1_1_1	BLKR_16019	Header run completed.
INFO	3/17/2009 6:06:33 PM	node01_ps42	WRITER_1_1_1	WRT_8005	Writer run started.
INFO	3/17/2009 6:06:33 PM	node01_ps42	WRITER_1_1_1	WRT_8158	*****START LOAD SESSION*****  Load Start Time: Tue Mar 17 18:06:33 2009  Target tables:  employee: Partition 1
INFO	3/17/2009 6:06:33 PM	node01_ps42	WRITER_1_1_1	WRT_8167	Start loading table [employee: Partition 1] at: Tue Mar 17 18:06:33 2009
INFO	3/17/2009 6:06:33 PM	node01_ps42	WRITER_1_1_1	WRT_8168	End loading table [employee: Partition 1] at: Tue Mar 17 18:06:33 2009
INFO	3/17/2009 6:06:33 PM	node01_ps42	WRITER_1_1_1	WRT_8141	Commit on end-of-data Tue Mar 17 18:06:33 2009 =====
					WRT_8036 Target: employee: Partition 1 (Instance Name: [employee]) WRT_8038 Inserted rows - Requested: 2      Applied: 2      Rejected: 0      Affected: 0
INFO	3/17/2009 6:06:33 PM	node01_ps42	WRITER_1_1_1	WRT_8143	Commit at end of Load Order Group Tue Mar 17 18:06:33 2009 =====
					WRT_8036 Target: employee: Partition 1 (Instance Name: [employee]) WRT_8038 Inserted rows - Requested: 2      Applied: 2      Rejected: 0      Affected: 0
INFO	3/17/2009 6:06:33 PM	node01_ps42	WRITER_1_1_1	GPWRT_340	[INFO] Deinitializing the partition driver.
INFO	3/17/2009 6:06:35 PM	node01_ps42	WRITER_1_1_1	GPWRT_340	[INFO] Gpload succeeded.
INFO	3/17/2009 6:06:35 PM	node01_ps42	WRITER_1_1_1	GPWRT_340	[INFO] For more information, see the gpload log file: .../infra_shared/Temp/gpload_0_0_0_1237293389_330000.log
INFO	3/17/2009 6:06:35 PM	node01_ps42	WRITER_1_1_1	GPWRT_340	[INFO] Greenplum Target Load Summary =====
INFO	3/17/2009 6:06:35 PM	node01_ps42	WRITER_1_1_1	GPWRT_340	[INFO] Table: [employee] Target: [1] Group: [1] Partition: [1] Requested Rows [2], Affected Rows [2], Applied Rows [2], Rejected Rows [0]
INFO	3/17/2009 6:06:35 PM	node01_ps42	WRITER_1_1_1	GPWRT_340	[INFO] =====

# Data Types Appendix



## Greenplum and ODBC Data Types

When you import Greenplum tables as target definitions, the ODBC data types corresponding to Greenplum data types appear in the Designer.

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

Greenplum Data Type	ODBC Data Type	Description
bigint or int8	Bigint	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision 19, scale 0
bigserial or serial8	Bigint	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision 19, scale 0
Boolean or bool	Bit	1 to 104,857,600 characters
box	SQL_LONGVARCHAR or SQL_VARCHAR or SQL_WVARCHAR	n/a
bpchar	Wchar	n/a
bytea	Binary or Long Varbinary	1 to 104,857,600 bytes
character or char(L)	Nchar	1 to 104,857,600 characters
character varying or varchar	Nvarchar or Long Varchar or WVarchar or WLongVarchar	n/a
cidr	SQL_LONGVARCHAR or SQL_VARCHAR or SQL_WVARCHAR	n/a

circle	SQL_LONGVARCHAR or SQL_VARCHAR or SQL_WVARCHAR	n/a
date	Date	Jan 1, 0001 A.D. to Dec 31, 9999 A.D. (precision to the nanosecond)
datetime	Timestamp	Jan 1, 0001 A.D. to Dec 31, 9999 A.D. (precision to the nanosecond)
decimal or numeric	Numeric	Precision 1 to 28, scale 0 to 28
double precision or float8	Double or Float	Precision 15
inet	SQL_LONGVARCHAR or SQL_VARCHAR or SQL_WVARCHAR	n/a
integer or int or int4	Integer	-2,147,483,648 to 2,147,483,647 Precision 10, scale 0
interval	SQL_LONGVARCHAR or SQL_VARCHAR or SQL_WVARCHAR	n/a
line	SQL_LONGVARCHAR or SQL_VARCHAR or SQL_WVARCHAR	n/a
lseg	SQL_LONGVARCHAR or SQL_VARCHAR or SQL_WVARCHAR	n/a
macaddr	SQL_LONGVARCHAR or SQL_VARCHAR or SQL_WVARCHAR	n/a
path	SQL_LONGVARCHAR or SQL_VARCHAR or SQL_WVARCHAR	n/a
point	SQL_LONGVARCHAR or SQL_VARCHAR or SQL_WVARCHAR	n/a
polygon	SQL_LONGVARCHAR or SQL_VARCHAR or SQL_WVARCHAR	n/a



real or float4	Real	Precision 7, scale 0
serial or serial4	Integer	-2,147,483,648 to 2,147,483,647 Precision 10, scale 0
smallint or int2	Tinyint	Precision 5, scale 0
text	Text or Ntext	1 to 104,857,600 characters
time	Time	Jan 1, 0001 A.D. to Dec 31, 9999 A.D. (precision to the nanosecond)
UUID	GUID	n/a

For more information about Greenplum data types, see the *Greenplum Database Administrator Guide*.

## Unsupported Greenplum Data Types

PowerExchange for Greenplum does not support the following Greenplum data types:

- bigserial
- bytea
- line
- serial
- time with time zone
- timestamp with time zone

When you import a Greenplum table that contains columns with unsupported data types, the Designer imports the columns. When you run a session, the results are unpredictable, and can include session failures. The `gpload` utility also fails and logs an error in the `gpload` log file.

## Time with Time Zone or Timestamp with Time Zone

When you import target definitions, the Designer imports columns of time with time zone or timestamp with time zone data types as time. To process the data correctly, you must edit the target definition in the Designer. On the Columns tab of the target definition, manually change the time data type to character.

The Greenplum data types, Time with Time Zone and Timestamp with Time zone are represented in a specific format in the Greenplum database. The values for these columns must be in the same format as defined by these data types. For more information about these formats, see the *Greenplum Database Administrator Guide*.

## Deprecated Greenplum Data Types

---

### Money

When you import Greenplum tables with columns of money data type, the Designer imports the columns. When you run a session with a mapping that contains target instances with columns of money data type, the session results are unpredictable.

To process the data correctly, you must edit the target definition in the Designer. On the Columns tab of the target definition, manually change the money data type to numeric or decimal. You can also change the precision and the scale for the column. Ensure that the scale for the money data type is not greater than two. For more information about the precision for the money data type, see the *Greenplum Database Administrator Guide*.

# Error Messages Appendix

# B

## GPWRT Error Messages

---

**GPWRT\_31012 The GPHOME\_LOADERS environment variable is not set.**

### *Solution*

Set the GPHOME\_LOADERS environment variable to the location where the `gpload` utility is installed.

**GPWRT\_32004 The session failed because the mapping contains a multi-group target instance.**

### *Cause*

PowerExchange for Greenplum cannot load data into a target table that is imported as a multi-group target definition.

### *Solution*

Ensure that the mapping does not contain a multi-group target instance.

**GPWRT\_33004 The session failed because multiple partitions are not supported.**

### *Cause*

The Greenplum session is configured for multiple partitions.

### *Solution*

In the session properties, in the Partitions view of the Mapping tab, verify that multiple partitions are not configured for the Greenplum session. By default, the Integration Service creates a single partition when you create Greenplum sessions.

**GPWRT\_34004 Failed to retrieve the field list.**

### *Cause*

Internal error.

### *Solution*

Contact Greenplum Customer Support.

**GPWRT\_34005 The Integration Service failed to retrieve the connection object.**

### *Cause*

The connection object is not created.

### *Solution*

Verify that a Greenplum Connection object is created and configured for each target instance in the mapping.

### *or*

### *Cause*

Internal error.

*Solution*

Contact Greenplum Customer Support.

**GPWRT\_34006 The Integration Service is unable to retrieve the attributes for the connection object.**

*Cause*

An attribute in the connection object is invalid.

*Solution*

Verify that the attributes specified in the connection object are valid.

*or*

Verify that the required attributes are specified in the connection object. For more information about connection attributes, see [Configuring Relational Connections](#).

**GPWRT\_34007 Terminating the session because the Integration Service received an abort or a stop request from the Data Transformation Manager (DTM) process.**

*Cause*

Internal error.

*Solution*

Contact Greenplum Customer Support.

**GPWRT\_34010 Failed to retrieve the name for the target table.**

*Cause*

Internal error.

*Solution*

Contact Greenplum Customer Support.

**GPWRT\_34011 Failed to retrieve fields for the table <table name>.**

*Cause*

Internal error.

*Solution*

Contact Greenplum Customer Support.

**GPWRT\_34014 The Integration Service failed to create a pipe.**

*Cause*

The Integration Service was unable to allocate sufficient memory to create a pipe.

*Solution*

Verify that the computer configured to run the Integration Service has sufficient memory.

*or*

Delete all the files from the <PowerCenter Installation Directory>/server/infa\_shared/Temp directory.

*or*

### *Cause*

The Integration Service was unable to create a pipe because it does not have the required permissions on the <PowerCenter Installation Directory>/server/infa\_shared/Temp directory.

### *Solution*

Verify that the user configured to start the PowerCenter Services has the required permissions on the <PowerCenter Installation Directory>/server/infa\_shared/Temp directory.

*or*

Contact your system administrator.

## **GPWRT\_34015 The Integration Service failed to create a thread.**

### *Cause*

The Integration Service was unable to allocate sufficient system resources to create a thread.

### *Solution*

Verify that the computer configured to run the Integration Service has sufficient system resources to create a thread.

*or*

Contact your system administrator.

## **GPWRT\_34016 The Integration Service failed to open a pipe.**

### *Cause*

The user configured to start the PowerCenter Services does not have permissions to open the pipe for writing data.

### *Solution*

Verify that the user configured to start the PowerCenter Services has permissions to open the pipe for writing data.

*or*

Contact your system administrator.

## **GPWRT\_34017 The Integration Service failed to write to a pipe.**

### *Cause*

The gpload process terminated unexpectedly.

### *Solution*

Verify that the gpload process is running when the Integration Service writes data to a pipe. For more information about the gpload error, see the session log.

*or*

Verify that the user configured to start the PowerCenter Services has permissions to write data to the pipe.

*or*

Contact your system administrator.

## **GPWRT\_34018 Data type not supported.**

### *Cause*

The session failed because PowerExchange for Greenplum does not support the bytea data type.

#### *Solution*

In the mapping, ensure that columns with bytea data type are not linked. When you import a table that contains a column with a bytea data type, the Designer imports the bytea data type as a raw data type. For more information about unsupported data types, see [Unsupported Greenplum Data Types](#).

**GPWRT\_34022 Gpload failed with exit status <error code>.**

#### *Cause*

The `gpload` script was not found in the location configured in the `GPHOME_LOADERS` environment variable.

#### *Solution*

Verify that a valid path for the `gpload` script is set in the `GPHOME_LOADERS` environment variable.

*or*

#### *Cause*

The `pyyaml` Python module is not installed or could not be found.

#### *Solution*

Set the `PYTHONPATH` environment variable to the location where the `pyyaml` Python module is installed. For more information about the path where this module is installed, see the *Greenplum Database Administrator Guide*.

*or*

#### *Cause*

Format errors. The data contains special characters, such as delimiters, escape characters, and quote characters.

#### *Solution*

Examine the data and specify a character that is not a part of the data for the delimiters, escape characters, or quote characters.

*or*

In the session properties, clear the Skip Escaping check box to escape special characters in the data.

*or*

#### *Cause*

Attributes in the connection object are invalid.

#### *Solution*

Ensure that attributes specified in the connection object are correct.

*or*

#### *Cause*

The error limit is set to a number less than two.

#### *Solution*

If the session is configured to log discarded rows in an error table, ensure that the error limit is set to a number greater than or equal to two.

*or*

### *Cause*

The gpload utility encountered an error while loading data into the Greenplum tables.

### *Solution*

Examine the gpload log file or the GPWRT\_34065 error message for the cause of the failure, and eliminate the errors.

*or*

Contact your system administrator.

## **GPWRT\_34023 The session failed because the columns specified in the Match Columns session property are invalid.**

### *Cause*

The specified columns do not exist in the target definition, or the columns are not linked in the mapping.

### *Solution*

Verify that the columns exist in the target definition and are linked in the mapping.

*or*

### *Cause*

The column names are not enclosed in double quotes, or the column names are not separated by a comma.

### *Solution*

Enclose each column name in double quotes, and use a comma to separate the column names.

## **GPWRT\_34024 The session failed because the columns specified in the Update Columns session property are invalid.**

### *Cause*

The specified columns do not exist in the target definition, or the columns are not linked in the mapping.

### *Solution*

Verify that the columns exist in the target definition, and are linked in the mapping.

*or*

### *Cause*

The column names are not enclosed in double quotes, or the column names are not separated by a comma.

### *Solution*

Enclose each column name in double quotes, and use a comma to separate the column names.

## **GPWRT\_34027 The session failed because the delimiter specified in the Delimiter session property is invalid.**

### *Cause*

The specified delimiter is a multibyte string.

### *Solution*

In the session properties, ensure that a single 7-bit ASCII character is specified as a delimiter. For more information about delimiters, see [Configuring Sessions for Greenplum Targets](#).



**GPWRT\_34028 The session failed because the escape character specified in the Escape session property is invalid.**

*Cause*

The specified value is a multibyte string.

*Solution*

In the session properties, ensure that a single 7-bit ASCII character is specified as an escape character. For more information about escape characters, see [Configuring Sessions for Greenplum Targets](#).

**GPWRT\_34030 The session failed because the character specified in the Quote session property is invalid.**

*Cause*

The specified value is a multibyte string.

*Solution*

In the session properties, ensure that a single 7-bit ASCII character is specified as a quote character. For more information about quote characters, see [Configuring Sessions for Greenplum Targets](#).

**GPWRT\_34031 The session failed because the error limit specified in the Error Limit session property is invalid.**

*Cause*

The error limit is set to a number less than zero.

*Solution*

Verify that the error limit is set to a number greater than or equal to zero.

**GPWRT\_34053 Failed to allocate memory for Unicode conversion.**

*Cause*

The Integration Service was unable to allocate sufficient memory for Unicode conversion when running in the Unicode data movement mode.

*Solution*

Verify that the computer configured to run the Integration Service has sufficient memory.

*or*

Contact your system administrator.

**GPWRT\_34055 Error threshold <threshold> reached.**

*Cause*

The session failed because the Integration Service reached the error threshold specified in the session properties.

*Solution*

Eliminate the errors in the data.

*or*

In the session properties, increase the error threshold.

**GPWRT\_34056 The session failed because no columns are specified in the Match Columns session property.**

### Cause

To perform merge and update operations, column names must be specified in the Match Columns session property.

### Solution

In the session properties, specify the column names.

## **GPWRT\_34057 The session failed because no columns are specified in the Update Columns session property.**

### Cause

To perform merge and update operations, column names must be specified in the Update Columns session property.

### Solution

In the session properties, specify the column names.

## **GPWRT\_34058 Unable to create the gpload process.**

### Cause

Python is not installed on the computer configured to run the gpload utility.

### Solution

Install a Python version that is compatible with the gpload utility. For more information about the Python version that is compatible with the gpload utility, see the *Greenplum Database Administrator Guide*.

or

Contact your system administrator.

or

### Cause

The PATH environment variable is not set to the location where Python is installed.

### Solution

Set the PATH environment variable to the location where Python is installed.

## **GPWRT\_34059 An error occurred while parsing the log file <name> created by gpload. The file does not exist or cannot be opened.**

### Cause

The gpload log file was not created in the <PowerCenter Installation Directory>/server/infa\_shared/Temp directory because Python 2.4.4 or later version is not installed on the machine configured to run the gpload utility.

### Solution

Install Python 2.4.4 or later version on the computer configured to run the gpload utility.

or

### Cause

The gpload log file was not created in the <PowerCenter Installation Directory>/server/infa\_shared/Temp directory because the `pyyaml` Python module is not installed or could not be found.

### Solution

Set the PYTHONPATH environment variable to the location where the `pyyaml` Python module is installed. For more information about the location of the module, see the *Greenplum Database Administrator Guide*.

*or*

#### *Cause*

The computer configured to run the Integration Service does not have sufficient system resources to create the file.

#### *Solution*

Delete the files in the <PowerCenter Installation Directory>/server/ infa\_shared/Temp directory, and run the session again.

*or*

Contact your system administrator.

*or*

#### *Cause*

The Integration Service does not have read permissions on the directory.

#### *Solution*

Verify that the user configured to start the PowerCenter Services has read and write permissions on the <PowerCenter Installation Directory>/server/infa\_shared/Temp directory.

### **GPWRT\_34063 Failed to convert <value> to UTF-8.**

#### *Cause*

The Integration Service was unable to allocate sufficient memory for converting data to UTF-8.

#### *Solution*

Verify that the computer configured to run the Integration Service has sufficient memory.

*or*

Contact Greenplum Customer Support.

### **GPWRT\_34064 The Integration Service failed to retrieve the value for the session attribute <attribute name>.**

#### *Cause*

The session attribute is not specified in the plug-in XML file that is registered with the repository.

#### *Solution*

Verify that the correct plug-in XML file is registered with the repository.

*or*

#### *Cause*

Internal error.

#### *Solution*

Contact Greenplum Customer Support.

### **GPWRT\_34065 Gpload error: <Reason for the gpload failure>**

#### *Cause*

See the error message for the cause of the gpload failure.

### *Solution*

For more information about the cause of the gpload error, see the Greenplum documentation. For more information about the gpload execution log, see the gpload log file.

**GPWRT\_34069 Quote, escape, or delimiter character cannot be the same.**

### *Cause*

Among Quote, Escape, and Delimiter session properties, two or more properties have the same value.

### *Solution*

In the session properties, specify different values for the quote, escape, and delimiter characters.