



Informatica® Cloud Data Integration

Greenplum Connector

© Copyright Informatica LLC 2018, 2024

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, Informatica Cloud, and PowerCenter 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.

See patents at <https://www.informatica.com/legal/patents.html>.

DISCLAIMER: Informatica LLC provides this documentation "as is" without warranty of any kind, either express or implied, including, but not limited to, the implied warranties of noninfringement, merchantability, or use for a particular purpose. Informatica LLC does not warrant that this software or documentation is error free. The information provided in this software or documentation may include technical inaccuracies or typographical errors. The information in this software and documentation is subject to change at any time without notice.

NOTICES

This Informatica product (the "Software") includes certain drivers (the "DataDirect Drivers") from DataDirect Technologies, an operating company of Progress Software Corporation ("DataDirect") which are subject to the following terms and conditions:

1. THE DATADIRECT DRIVERS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT.
2. IN NO EVENT WILL DATADIRECT OR ITS THIRD PARTY SUPPLIERS BE LIABLE TO THE END-USER CUSTOMER FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, CONSEQUENTIAL OR OTHER DAMAGES ARISING OUT OF THE USE OF THE ODBC DRIVERS, WHETHER OR NOT INFORMED OF THE POSSIBILITIES OF DAMAGES IN ADVANCE. THESE LIMITATIONS APPLY TO ALL CAUSES OF ACTION, INCLUDING, WITHOUT LIMITATION, BREACH OF CONTRACT, BREACH OF WARRANTY, NEGLIGENCE, STRICT LIABILITY, MISREPRESENTATION AND OTHER TORTS.

The information in this documentation is subject to change without notice. If you find any problems in this documentation, 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.

Publication Date: 2024-07-15

Table of Contents

Preface	4
Informatica Resources.	4
Informatica Documentation.	4
Informatica Intelligent Cloud Services web site.	4
Informatica Intelligent Cloud Services Communities.	4
Informatica Intelligent Cloud Services Marketplace.	4
Data Integration connector documentation.	5
Informatica Knowledge Base.	5
Informatica Intelligent Cloud Services Trust Center.	5
Informatica Global Customer Support.	5
Chapter 1: Introduction to Greenplum Connector	6
Greenplum Connector assets.	6
Chapter 2: Connections for Greenplum	7
Prerequisites.	7
Configuring JDBC and ODBC drivers.	7
Configuring Kerberos authentication.	9
Connect to Greenplum.	11
Before you begin.	11
Connection details.	11
Authentication types.	11
Chapter 3: Mappings and mapping tasks with Greenplum Connector	14
Greenplum sources in mappings.	14
Greenplum targets in mappings.	16
Match and update columns.	19
Error handling for Greenplum targets.	20
Chapter 4: Data type reference	21
Greenplum and transformation data types.	21
Index	23

Preface

Use *Greenplum Connector* to learn how to read from or write to Greenplum by using Cloud Data Integration. Learn to create a connection, develop mappings, and run mapping tasks and dynamic mapping tasks in Cloud Data Integration.

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 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 Intelligent Cloud Services web site

You can access the Informatica Intelligent Cloud Services web site at <http://www.informatica.com/cloud>. This site contains information about Informatica Cloud integration services.

Informatica Intelligent Cloud Services Communities

Use the Informatica Intelligent Cloud Services Community to discuss and resolve technical issues. You can also find technical tips, documentation updates, and answers to frequently asked questions.

Access the Informatica Intelligent Cloud Services Community at:

<https://network.informatica.com/community/informatica-network/products/cloud-integration>

Developers can learn more and share tips at the Cloud Developer community:

<https://network.informatica.com/community/informatica-network/products/cloud-integration/cloud-developers>

Informatica Intelligent Cloud Services Marketplace

Visit the Informatica Marketplace to try and buy Data Integration Connectors, templates, and mapplets:

<https://marketplace.informatica.com/>

Data Integration connector documentation

You can access documentation for Data Integration Connectors at the Documentation Portal. To explore the Documentation Portal, visit <https://docs.informatica.com>.

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 Intelligent Cloud Services Trust Center

The Informatica Intelligent Cloud Services Trust Center provides information about Informatica security policies and real-time system availability.

You can access the trust center at <https://www.informatica.com/trust-center.html>.

Subscribe to the Informatica Intelligent Cloud Services Trust Center to receive upgrade, maintenance, and incident notifications. The [Informatica Intelligent Cloud Services Status](#) page displays the production status of all the Informatica cloud products. All maintenance updates are posted to this page, and during an outage, it will have the most current information. To ensure you are notified of updates and outages, you can subscribe to receive updates for a single component or all Informatica Intelligent Cloud Services components. Subscribing to all components is the best way to be certain you never miss an update.

To subscribe, on the [Informatica Intelligent Cloud Services Status](#) page, click **SUBSCRIBE TO UPDATES**. You can choose to receive notifications sent as emails, SMS text messages, webhooks, RSS feeds, or any combination of the four.

Informatica Global Customer Support

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

To find online support resources on the Informatica Network, click **Contact Support** in the Informatica Intelligent Cloud Services Help menu to go to the **Cloud Support** page. The **Cloud Support** page includes system status information and community discussions. Log in to Informatica Network and click **Need Help** to find additional resources and to contact Informatica Global Customer Support through email.

The telephone numbers for Informatica Global Customer Support are available from the Informatica web site at <https://www.informatica.com/services-and-training/support-services/contact-us.html>.

CHAPTER 1

Introduction to Greenplum Connector

Greenplum Connector provides connectivity between Cloud Data Integration and the Greenplum database.

You can configure mappings in Cloud Data Integration to read data from or write data to Greenplum.

When you run a Greenplum mapping to read data from Greenplum, the Secure Agent invokes the Greenplum database parallel file server, gpfdist, to read data. You can switch mappings to advanced mode to include transformations and functions that enable advanced functionality.

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

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

Greenplum Connector assets

Create assets in Data Integration to integrate data using Greenplum Connector.

When you use Greenplum Connector, you can include the following Data Integration assets:

- Dynamic mapping task
- Mapping
- Mapping task

For more information about configuring assets and transformations, see *Mappings*, *Transformations*, and *Tasks* in the Data Integration documentation.

CHAPTER 2

Connections for Greenplum

Create a Greenplum connection to securely read data from or write data to Greenplum.

You can use a Greenplum connection to specify sources or targets in mappings and mapping tasks.

Prerequisites

Before you use Greenplum Connector, ensure that you meet the prerequisites.

Perform the following prerequisite tasks:

1. Install the Greenplum loaders package on the Secure Agent machine. The loaders package contains the gload utility. You can download the Greenplum loaders package from the Pivotal Greenplum website.
2. Configure the DataDirect Greenplum ODBC and JDBC drivers on the Secure Agent machine.
3. Configure the authentication prerequisites to connect to a Greenplum database. You can configure Database or Kerberos authentication. Keep the following authentication details handy based on the authentication type that you want to use:
 - To configure Database authentication, you need the user name, password, host name, port, database name from your Greenplum account.
 - To configure Kerberos authentication, you need the service principal name, host name, port, and database name from your Greenplum account.

Configuring JDBC and ODBC drivers

Before you use Greenplum Connector, configure DataDirect Greenplum JDBC and ODBC drivers on Windows and Linux.

Configure the DataDirect Greenplum JDBC Driver on Linux

To use a Greenplum connection from a Linux machine, you must configure the DataDirect Greenplum JDBC driver on Linux.

1. Download the DataDirect Greenplum JDBC driver version 6.x from the Pivotal Greenplum website.
2. Copy the Greenplum JDBC driver to the following directory on the Secure Agent machine: <Secure Agent installation directory>/apps/Data_Integration_Server/ext/deploy_to_main/bin/rdtm-extra/Greenplum
Note: You must create the `deploy_to_main/bin/rdtm-extra` directory manually.
3. Restart the Secure Agent.

Configure the DataDirect Greenplum JDBC driver on Windows

To use a Greenplum connection from a Windows machine, you must configure the DataDirect Greenplum JDBC driver on Windows.

1. Download the DataDirect Greenplum JDBC driver version 6.x from the Pivotal Greenplum website.
2. Copy the Greenplum JDBC driver to the following directory on the Secure Agent machine:
<Secure Agent installation directory>\apps\Data_Integration_Server\ext\deploy_to_main\bin\rdtm-extra\Greenplum
You must create the deploy_to_main\bin\rdtm-extra\Greenplum directory manually.
3. Restart the Secure Agent.

Configure the DataDirect Greenplum ODBC driver on Linux

To use a Greenplum connection from a Linux machine, you must configure the DataDirect Greenplum ODBC driver on Linux.

1. Download the DataDirect Greenplum ODBC driver version 7.1.6 from the Pivotal website and install the DataDirect Greenplum ODBC driver on the Secure Agent machine.
2. Add the following driver entries to the `odbcinst.ini` file in the following directory: <Secure Agent installation directory>/odbcinst.ini

Use the following syntax:

```
[DataDirect 7.1 Greenplum Wire Protocol]
Driver=<ODBC driver path>/lib/ddgplm27.so
Setup=<ODBC driver path>/lib/ddgplm27.so
DriverODBCVer=<ODBC Driver version>
HelpRootDirectory=<ODBC Driver path>/help
GSSClient=libgssapi_krb5.so.2
```

For example:

```
[DataDirect 7.1 Greenplum Wire Protocol]
Driver=/opt/Progress/DataDirect/Connect64_for_ODBC_71/lib/ddgplm27.so
Setup=/opt/Progress/DataDirect/Connect64_for_ODBC_71/lib/ddgplm27.so
APILevel=0
ConnectFunctions=YYY
DriverODBCVer=3.52
FileUsage=0
HelpRootDirectory=/opt/Progress/DataDirect/Connect64_for_ODBC_71/help
SQLLevel=0
GSSClient=libgssapi_krb5.so.2
```

3. Set the `GPHOME_LOADERS`, `PATH`, and `LD_LIBRARY_PATH` environmental variables for the driver. Perform the following tasks:
 - a. Set the `GPHOME_LOADERS` environmental variable to the directory that contains the Greenplum loader libraries. Using a C shell and run the following command:

```
setenv GPHOME_LOADERS /export/qa_adp/thirdparty/greenplum/rhel.64/loaders
```
 - b. Set the `PATH` environmental variable to the directory that contains the Greenplum loader libraries. Using a C shell and run the following command:

```
setenv PATH ${GPHOME_LOADERS}/bin:${PATH}
```
 - c. Set the `LD_LIBRARY_PATH` environmental variable to include the following directories that contain the Greenplum drivers and the DataDirect Greenplum ODBC libraries. Using a C shell and run the following command:

```
setenv LD_LIBRARY_PATH .:${GPHOME_LOADERS}/lib:/export/qa_adp/thirdparty/
greenplum/rhel.64/loaders/ext/python/lib
```



```
setenv LD_LIBRARY_PATH /opt/Progress/DataDirect/Connect64_for_ODBC_71/  
lib:$LD_LIBRARY_PATH
```

4. Restart the Secure Agent after you update the environment variables.

Configure the DataDirect Greenplum ODBC driver on Windows

To use a Greenplum connection from a Windows machine, you must configure the DataDirect Greenplum ODBC driver on Windows.

1. Download the DataDirect Greenplum ODBC driver version 7.1.6 from the Pivotal website and install the DataDirect Greenplum ODBC driver on the Secure Agent machine.
2. Install Python 2.5.4 32-bit.
3. Install the 5.18 Greenplum Clients software for Windows.
4. Install the 5.18 Greenplum loaders software for Windows.
5. Set the following environmental variables for the driver:

- Set the GPHOME_LOADERS environment variable to the folder that contains the Greenplum loader libraries:

For example, Set GPHOME_LOADERS = C:\Program Files (x86)\Greenplum\greenplum-loaders-5.18.0

- Set the GPHOME_CLIENTS environment variable to the folder that contains the Greenplum clients libraries:

For example, set GPHOME_CLIENTS=C:\Program Files (x86)\Greenplum\greenplum-clients-5.18.0\

- Set the PYTHONPATH environment variable:
Set PYTHONPATH=%GPHOME_LOADERS%\bin\lib

- Set the DDCPATH environment variable to the folder containing the DataDirect libraries.

For example, DDCPATH=C:\Program Files\Progress\DataDirect
\Connect64_for_ODBC_71\drivers;C:\Program Files\Progress\DataDirect
\Connect64_for_ODBC_71\jre\bin;C:\Program Files\Progress\DataDirect
\Connect64_for_ODBC_71\jre\bin\server

- Set the Path environment variable to the folder that contains the Greenplum clients, Greenplum loaders, Python and Greenplum ODBC Datadirect driver libraries:

For example, 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;%DDCPATH%

Configuring Kerberos authentication

Before you use Kerberos authentication to connect to Greenplum on Linux or Windows, the organization administrator needs to perform the prerequisite tasks.

1. To configure the Java Authentication and Authorization Service configuration file (JAAS), perform the following tasks:
 - a. Create a JAAS configuration file on the Secure Agent machine.

- b. Add the following entries to the JAAS configuration file:

```
JDBC_DRIVER_01 {
  com.sun.security.auth.module.Krb5LoginModule required useTicketCache=true;
};
```

2. To configure the `krb5.conf` file, perform the following tasks:
 - a. Create a `krb5.conf` file on the Secure Agent machine.
 - b. Add the details of the Key Distribution Center (KDC) and admin server to the `krb5.conf` file in the following format:

```
[libdefaults]
default_realm = <Realm name>
forwardable = true
ticket_lifetime = 24h

[realms]
<REALM NAME> = {
  kdc = <Location where KDC is installed>
  admin_server = <Location where KDC is installed>
}

[domain_realm]
<domain name or host name> = <Domain name or host name of Kerberos>
<domain name or host name> = <Domain name or host name of Kerberos>
```

3. Set the environment variables on the Secure Agent machine.
4. Restart the Secure Agent.
5. To generate the credential cache file on the Secure Agent machine and use Kerberos authentication to connect to Greenplum, perform the following tasks:
 - a. From the command line on the Secure Agent machine, run the following command and specify the Greenplum user name and realm name:

```
Kinit <user name>@<realm_name>
```
 - b. When prompted, enter the password for the Kerberos principal user.

Setting environment variables

To use Kerberos authentication to connect to Greenplum, you need to set the required environment variables on the Secure Agent machine.

Run the following commands to set the environment variables:

- `setenv KRB5CCNAME <Absolute path and file name of the credentials cache file>`
- `setenv KRB5_CONFIG <Absolute path of the Kerberos configuration file>\krb5.conf`
- `setenv JAASCONFIG <Absolute path of the JAAS config file>\<File name>.conf`

After you set the environmental variables, you need to restart the Secure Agent.

Alternatively, you can add the `KRB5_CONFIG` and `JAASCONFIG` environment variables when you create a Greenplum connection.

To add the environment variables when you configure a connection with Kerberos authentication, you need to add the `KRB5_CONFIG` and `JAASCONFIG` properties in the **Kerberos Connection Properties** field in a Greenplum connection.

For example, add the properties in the following format:

```
KRB5_CONFIG=<Absolute path of the Kerberos configuration file>
\krb5.conf;JAASCONFIG=<Absolute path of the JAAS config file>\<File name>.conf
```

Note: Ensure that you separate each key-value pair with a semicolon.

Connect to Greenplum

Let's configure the Greenplum connection properties to connect to Greenplum.

Before you begin

Before you get started, be sure to complete the prerequisites.

Check out ["Prerequisites" on page 7](#) to learn more about the authentication prerequisites and other tasks that you must perform.

Connection details

The following table describes the basic connection properties:

Property	Description
Connection Name	Name of the connection. Each connection name must be unique within the organization. Connection names can contain alphanumeric characters, spaces, and the following special characters: _ . + -, Maximum length is 255 characters.
Description	Description of the connection. Maximum length is 4000 characters.
Type	Greenplum
Runtime Environment	The name of the runtime environment where you want to run tasks. You can specify a Secure Agent, Hosted Agent, or serverless runtime environment.
Host Name	Host name or IP address of the Greenplum server.
Port	Greenplum server port number. If you enter 0, the gpload utility reads from the environment variable \$PGPORT. Default is 5432.
Database	Name of the Greenplum database.
Schema	Name of the schema that contains the metadata for Greenplum sources or targets. Default is public.

Authentication types

You can configure Database or Kerberos authentication types to connect to the Greenplum database.

Select the required authentication type and then configure the authentication-specific parameters.

Database authentication

To configure Database authentication, you need the user name and password from your Greenplum account.

The following table describes the basic connection properties for database authentication:

Property	Description
Username	User name with permissions to access the Greenplum database.
Password	Password to connect to the Greenplum database.

The following table describes the advanced connection properties for Database authentication:

Property	Description
Certificates Path	<p>Path where the SSL certificates for the Greenplum server are stored.</p> <p>Specify the path if you want to establish secure communication between the gpload utility and the Greenplum server over SSL.</p> <p>For information about the files that need to be available in the certificates path, see the gpload documentation.</p> <p>Note: You can use the SSL-based connection only in a Target transformation in a mapping to write to Greenplum.</p>
Metadata Additional Connection Configuration	<p>Additional connection properties that you want to set to fetch the metadata from Greenplum.</p> <p>Enter the properties in the following format:</p> <pre><parameter name1>=<value1>, <parameter name2>=<value2></pre>
Driver Name	<p>The driver name.</p> <p>Specify DataDirect 7.1 Greenplum Wire Protocol.</p>

Kerberos authentication

To configure Kerberos authentication, you need the Kerberos connection properties, service principal name, host name, port, database name, and your Greenplum account details from your Greenplum account.

The following table describes the basic connection properties for Kerberos authentication:

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

The following table describes the advanced connection property for Kerberos authentication:

Property	Description
Kerberos Connection Properties	<p>Additional connection properties to use Kerberos authentication to connect to the Greenplum database.</p> <p>Enter properties in the following format:</p> <p><parameter name>=<parameter value></p> <p>If you specify more than one property, separate each key-value pair with a semicolon.</p>
Service Principal Name	<p>Service principal name that you want to use for Kerberos authentication.</p> <p>Specify the service principal name in the following format:</p> <p><Service Name>/<Fully Qualified Domain Name>@<REALM.COM></p> <ul style="list-style-type: none">- Service Name is the name of the service hosting the instance.- Fully Qualified Domain Name is the fully qualified domain name of the host machine.- REALM.COM is the domain name of the host machine. This value is optional. If you do not specify the realm name, the default realm is used.

CHAPTER 3

Mappings and mapping tasks with Greenplum Connector

When you create a mapping, you can configure a Source or Target transformation to represent a Greenplum object.

Greenplum source and target objects represent metadata for Greenplum tables. You can define properties in the mapping to determine how the gpload utility reads or writes data to Greenplum. When you run a mapping, the Secure Agent invokes the gpload utility to read data from or load data to Greenplum.

In advanced mode, the Mapping Designer updates the mapping canvas to include transformations and functions that enable advanced functionality.

Greenplum sources in mappings

To read data from Greenplum, configure a Greenplum object as a source in the mapping.

Specify the name and description of the Greenplum source. Configure the source and advanced properties for the source object. You can parameterize the source object name.

The following table describes the source properties that you can configure in a Source transformation:

Property	Description
Connection	Name of the source connection.
Source Type	Type of the source object. Select Single, Multiple, or Parameter.
Object	The source object.

The following table describes the Greenplum Query options:

Property	Description
Filter	<p>Filters records based on the filter condition you specify. Specify the source field and configure the operator and value to use in the filter.</p> <p>You can specify the following filter conditions:</p> <ul style="list-style-type: none"> - Not Parameterized. Specify the source field and configure the operator and value to use in the filter. - Completely Parameterized. Use a parameter for a filter expression and define the filter expression in the task. - Advanced. Create a filter expression using the expression editor.
Sort	Not applicable.

The following table describes the Greenplum advanced source properties:

Property	Description
Pipe Directory Path	<p>The file system location where the pipes that are used for data transfer are created.</p> <p>Default is <code>\$PMRootDir/Temp</code>.</p>
Delimiter	<p>Separates successive input fields.</p> <p>For data in the text format, use any 7-bit ASCII value except a-z, A-Z, and 0-9.</p> <p>For data in the CSV format, use any 7-bit ASCII value except \n, \r, and \\. Default is pipe ().</p> <p>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.</p>
Null Value	<p>String that represents a null value.</p> <p>In the source data, any data item that matches the string is treated as a null value.</p> <p>Default is backslash N (\N).</p>
Escape Character	<p>Character that treats special characters in the data as regular characters.</p> <p>In the text format, special characters comprise delimiter and escape characters.</p> <p>In the CSV format, special characters comprise quotes and escape characters.</p> <p>Use any 7-bit ASCII value as an escape character.</p> <p>Default is backslash (\).</p> <p>Note: You can improve the task performance if the data does not contain escape characters.</p>
SQL Query Override	<p>Overrides the schema specified in the Greenplum connection object. If you do not configure this property, the Data Secure Agent uses the schema value specified in the Greenplum connection object.</p>
Owner Name	The schema name in Greenplum.
Source Table Name	Overrides the default Greenplum source table name.
PreSQL	The SQL command to run before reading data from the source.
PostSQL	The SQL command to run after reading data from the source.

Property	Description
GPFDist Port	The HTTP port on which gpfdist serves files. Default is 39871.
GPFDIST Log File	The fully qualified path and log file name for the standard output message logs. Default is INFAGPFDistLogfile.out.
GPFDIST Timeout	The time allowed for Greenplum Database to establish a connection to a gpfdist process. Default is 60 seconds. Allowed values are 2 to 7200 seconds (2 hours).
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. Greenplum Connector supports only the UTF-8 character set encoding.
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.
Tracing Level	Determines the amount of detail that appears in the log for the source. You can choose Terse, Normal, Verbose Initialization, or Verbose Data tracing level. Default is Normal.

Greenplum targets in mappings

To write data to Greenplum, configure a Greenplum object as a target in the mapping.

Specify the name and description of the Greenplum target. Configure the target and advanced properties for the target object.

You can parameterize the target object name.

The following table describes the target properties that you can configure in a Target transformation:

Property	Description
Connection	Name of the target connection.
Target Type	Type of the target object. Select Single Object or Parameter.
Object	Target object for a single target. If you select parameter, define a name and type of data object.
Operation	The target operation.

The following table describes the Greenplum advanced target properties:

Property	Description
Method	<p>Determines how the gpload utility processes the data from the named pipe.</p> <p>You can select from the following methods:</p> <ul style="list-style-type: none"> - 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	<p>Matches rows based on the comma-separated list of column names. Enclose the column names in single quotation marks and ensure that there are no leading or trailing spaces between the column names.</p>
Update Columns	<p>Updates the columns specified in the comma-separated list of column names. Enclose the column names in single quotes and ensure that there are no leading and trailing spaces between the column names.</p>
Update Condition	<p>Updates a row based on the condition that you specify. The gpload utility performs an update or merge operation based on the update condition.</p>
Format	<p>The Secure Agent writes data in a format that is compatible with the gpload utility.</p> <p>Select one of the following values:</p> <ul style="list-style-type: none"> - Text. In the text format, the Secure Agent separates data using the delimiter character that you specify in the task properties. <p>If the data contains the delimiter or escape characters specified in the task properties, you can choose to ignore the escape character or specify delimiter and escape character values that are not a part of the data.</p> <ul style="list-style-type: none"> - CSV. In the CSV format, the Secure Agent encloses the data with the quote character that you specify in the task properties. The Secure Agent also separates the data using the delimiter character specified in the task properties. <p>If the data contains the quote or escape characters specified in the task properties, you can choose to ignore the escape character or specify quote and escape character values that are not a part of the data.</p> <p>Default is Text.</p> <p>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.</p>
Delimiter	<p>Separates successive input fields.</p> <p>For data in the text format, use any 7-bit ASCII value except a-z, A-Z, and 0-9.</p> <p>For data in the CSV format, use any 7-bit ASCII value except \n, \r, and \\. Default is pipe ().</p> <p>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.</p>
Escape	<p>Character that treats special characters in the data as regular characters.</p> <p>In the text format, special characters comprise delimiter and escape characters.</p> <p>In the CSV format, special characters comprise quotes and escape characters.</p> <p>Use any 7-bit ASCII value as an escape character.</p> <p>Default is backslash (\).</p> <p>Note: You can improve the task performance if the data does not contain escape characters.</p>

Property	Description
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 Secure Agent encloses data by the specified character and passes the data to the gpload utility. The quote character is ignored for data in the text format. Use any 7-bit ASCII value that is not equal to the delimiter or null value. Default is double quotes (").
Error Limit	For each Greenplum segment, across all partitions if pass-through partitioning is used, the number of rows that the gpload utility discards or logs in the error table because of format errors. The gpload utility fails the session if the error limit is reached for any Greenplum segment. Default is zero. The maximum error limit is 2,147,483,647.
Error Table	Name of the error table where the gpload utility logs rejected rows when reading data that is processed by the Secure Agent. 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.
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.
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.
Truncate Target Option	Truncates the Greenplum target table before loading data to the target.
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.

Property	Description
Window Pipe Buffer Size	The size (in kilobytes) that the Secure Agent 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 Secure Agent must delete the gpload control file after the task is complete. Default is selected.
Gpload Log File Location	The file system location where the gpload utility generates the gpload log file. Default is <code>\$PMRootDir/Temp</code> .
Gpload Control File Location	The file system location where the Secure Agent generates the gpload control file. Default is <code>\$PMRootDir/Temp</code> .
Encoding	Character set encoding of the source data. Greenplum Connector 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. Default is <code>\$PMRootDir/Temp</code> .
Schema Override	Overrides the schema specified in the Greenplum connection object. If you do not configure this property, the Data Secure Agent uses the schema value specified in the Greenplum connection object.
Max_Line_Length	The integer that specifies the maximum length of a line in the XML transformation data that is passed to gpload.
Success File Directory	Not applicable.
Error File Directory	Not applicable.
Forward Rejected Rows	Not applicable.

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.

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 Secure Agent 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
```

CHAPTER 4

Data type reference

Data Integration uses the following data types in mappings, and mapping tasks with Greenplum:

Greenplum Native Data Types

Greenplum data types appear in the Source and Target transformations when you choose to edit metadata for the fields.

Transformation Data Types

Set of data types that appear in the transformations. They are internal data types based on ANSI SQL-92 generic data types, which the runtime environment uses to move data across platforms. Transformation data types appear in all transformations in mappings, and mapping tasks.

When Data Integration reads source data, it converts the native data types to the comparable transformation data types before transforming the data. When Data Integration writes to a target, it converts the transformation data types to the comparable native data types.

Greenplum and transformation data types

When you import Greenplum tables, the ODBC data types corresponding to the Greenplum data types appear. The Secure Agent 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 that the runtime environment supports and the corresponding ODBC and transformation 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

Greenplum Data Type	ODBC Data Type	Transformation Data Type
Double precision	Double or Float	Double
Integer	Integer	Integer
Numeric	Numeric	Decimal
Real	Real	Double
Serial	Integer	Integer
Smallint	Tinyint	Integer
Text	Text	Text
Time	Time. Precision is 6.	Date/Time
Timestamp	Timestamp. Precision is 6.	Date/Time

INDEX

C

Cloud Application Integration community
URL [4](#)
Cloud Developer community
URL [4](#)
connections
Greenplum [7, 11](#)

D

Data Integration community
URL [4](#)
data type reference
overview [21](#)

E

error handling [20](#)

G

Greenplum
connection properties [11](#)
supported object types [6](#)
supported task types [6](#)
Greenplum Connections
overview [7](#)
Greenplum Connector
overview [6](#)
Greenplum Data Types
data types [21](#)
Greenplum objects
mapping [14](#)
Greenplum sources
mappings [14](#)

Greenplum targets
mappings [16](#)

I

Informatica Global Customer Support
contact information [5](#)
Informatica Intelligent Cloud Services
web site [4](#)

M

maintenance outages [5](#)

S

status
Informatica Intelligent Cloud Services [5](#)
system status [5](#)

T

trust site
description [5](#)

U

upgrade notifications [5](#)

W

web site [4](#)