



Informatica® PowerExchange for Google
BigQuery
10.2 HotFix 1

User Guide for PowerCenter

Informatica PowerExchange for Google BigQuery User Guide for PowerCenter
10.2 HotFix 1
August 2018

© Copyright Informatica LLC 2018

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

Informatica, the Informatica logo, PowerCenter, and PowerExchange are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners.

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

Portions of this software and/or documentation are subject to copyright held by third parties. Required third party notices are included with the product.

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

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

Publication Date: 2018-10-14

Table of Contents

Preface	5
Informatica Resources.	5
Informatica Network.	5
Informatica Knowledge Base.	5
Informatica Documentation.	5
Informatica Product Availability Matrixes.	6
Informatica Velocity.	6
Informatica Marketplace.	6
Informatica Global Customer Support.	6
 Chapter 1: Introduction to PowerExchange for Google BigQuery.....	 7
PowerExchange for Google BigQuery Overview.	7
Introduction to Google BigQuery.	7
Administration of Google BigQuery.	8
 Chapter 2: PowerExchange for Google BigQuery Configuration.....	 9
PowerExchange for Google BigQuery Configuration Overview.	9
Registering the PowerExchange for Google BigQuery Plug-in.	9
Registering the PowerExchange for Google BigQuery Plug-in from the Administrator Tool. . . .	10
Registering the PowerExchange for Google BigQuery Plug-in from the Command Line Program.	10
 Chapter 3: Google BigQuery Sources and Targets.....	 11
Google BigQuery Sources and Targets Overview.	11
Import Google BigQuery Source and Target Definitions.	11
 Chapter 4: Google BigQuery Mappings.....	 15
Google BigQuery Mappings Overview.	15
Source Filter.	15
Configuring a Source Filter.	16
 Chapter 5: Google BigQuery Sessions.....	 18
Google BigQuery Sessions Overview.	18
Google BigQuery Connections.	18
Connection Modes.	18
Connection Mode Example.	19
Rules and Guidelines for Google BigQuery Connection Modes.	21
Google BigQuery Connection Properties.	22
Configuring a Google BigQuery Connection.	23
Pre SQL and Post SQL Commands	23

Configure Google BigQuery Source Session Properties.	24
Configure Google BigQuery Target Session Properties.	26
Configure the Java Heap Memory.	30
Rules and Guidelines for Google BigQuery Sessions.	30
Appendix A: Google BigQuery Data Type Reference.....	31
Data Type Reference Overview.	31
Google BigQuery and Transformation Data Types.	31
Index.....	33

Preface

The *Informatica PowerExchange for Google BigQuery User Guide for PowerCenter* provides information about extracting data from and loading data to Google BigQuery. This guide is written for database administrators and developers who are responsible for developing mappings, sessions, and workflows that extract data from and load data to Google BigQuery. This guide assumes you have knowledge of Google BigQuery, Google Cloud Storage, and PowerCenter.

Informatica Resources

Informatica Network

Informatica Network hosts Informatica Global Customer Support, the Informatica Knowledge Base, and other product resources. To access Informatica Network, visit <https://network.informatica.com>.

As a member, you can:

- Access all of your Informatica resources in one place.
- Search the Knowledge Base for product resources, including documentation, FAQs, and best practices.
- View product availability information.
- Review your support cases.
- Find your local Informatica User Group Network and collaborate with your peers.

Informatica Knowledge Base

Use the Informatica Knowledge Base to search Informatica Network for product resources such as documentation, how-to articles, best practices, and PAMs.

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

Informatica Documentation

To get the latest documentation for your product, browse the Informatica Knowledge Base at https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx.

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

Informatica Product Availability Matrixes

Product Availability Matrixes (PAMs) indicate the versions of operating systems, databases, and other types of data sources and targets that a product release supports. If you are an Informatica Network member, you can access PAMs at

<https://network.informatica.com/community/informatica-network/product-availability-matrices>.

Informatica Velocity

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

If you are an Informatica Network member, you can access Informatica Velocity resources at

<http://velocity.informatica.com>.

If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at ips@informatica.com.

Informatica Marketplace

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

Informatica Global Customer Support

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

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

<http://www.informatica.com/us/services-and-training/support-services/global-support-centers>.

If you are an Informatica Network member, you can use Online Support at <http://network.informatica.com>.

CHAPTER 1

Introduction to PowerExchange for Google BigQuery

This chapter includes the following topics:

- [PowerExchange for Google BigQuery Overview, 7](#)
- [Introduction to Google BigQuery, 7](#)
- [Administration of Google BigQuery, 8](#)

PowerExchange for Google BigQuery Overview

You can use PowerExchange for Google BigQuery for connectivity between PowerCenter and Google BigQuery.

You can use Google BigQuery objects as sources and targets in mappings. When you use Google BigQuery objects in mappings, you must configure properties specific to Google BigQuery.

Example

Your organization is an open source log data collector, which collects log data from multiple sources and unifies them.

Logs help you understand how systems and applications perform. As the scale and complexity of the system increases, it is difficult to manage multiple logs from different sources.

To overcome this problem, you can use PowerExchange for Google BigQuery to write data to a Google BigQuery target and query terabytes of logs in seconds. You can then use the data to fix and improve the system performance in near real time.

Introduction to Google BigQuery

Google BigQuery is a fast, highly scalable, cost-effective and fully managed enterprise data warehouse that the Google Cloud Platform provides. You can store and analyze massive amounts of data using ANSI:2011 compliant SQL queries.

Administration of Google BigQuery

Google BigQuery is a RESTful web service that the Google Cloud Platform provides.

Before you use PowerExchange for Google BigQuery, you must complete the following prerequisite tasks:

1. Ensure you have a service account in your Google account to access Google BigQuery.
2. Ensure you have the `client_email`, `project_id`, and `private_key` values for the service account. You will need to enter these details when you create a Google BigQuery connection in Developer tool.
3. Ensure you have the project ID, dataset ID, source table name, and target table name when you create mappings in PowerCenter.
4. Verify that you have read and write access to the Google BigQuery dataset that contains the source table and target table.
5. If you use bulk mode, verify that you have write access to the Google Cloud Storage path where the PowerCenter Integration Service creates the staging file.
6. If you use staging mode, verify that you have read access to the Google Cloud Storage path where the PowerCenter Integration Service creates the staging file to store the data from the Google BigQuery source.

CHAPTER 2

PowerExchange for Google BigQuery Configuration

This chapter includes the following topics:

- [PowerExchange for Google BigQuery Configuration Overview, 9](#)
- [Registering the PowerExchange for Google BigQuery Plug-in, 9](#)

PowerExchange for Google BigQuery Configuration Overview

PowerExchange for Google BigQuery installs with Informatica services.

If you upgrade from a previous version, you must register the PowerExchange for Google BigQuery plug-in with the PowerCenter repository.

Registering the PowerExchange for Google BigQuery Plug-in

To register the plug-in, the repository must be running in exclusive mode. Use the Administrator tool or the `pmrep RegisterPlugin` command line program to register the plug-in. If you do not have the correct privileges to register the plug-in, contact the user who manages the PowerCenter Repository Service.

The plug-in file is an `.xml` file that defines the functionality of the adapter. When you install the server component, the installer copies the plug-in file to the following directory:

```
<Informatica installation directory>/server/bin/Plugin
```

The name of the plug-in file for PowerExchange for Google BigQuery is `bigqueryPlugin.xml`.

Registering the PowerExchange for Google BigQuery Plug-in from the Administrator Tool

Register a repository plug-in to add its functionality to the repository.

1. Run the PowerCenter Repository Service in exclusive mode.
2. In the **Navigator**, select the PowerCenter Repository Service to which you want to add the plug-in.
3. In the **Contents** panel, click the **Plug-ins** view.
4. In the **Actions** menu of the **Domain** tab, select **Register Plug-in**.
5. On the **Register Plug-in** page, click the **Browse** button to locate the plug-in file.
6. Enter your user name and password.
7. Click **OK**.

The PowerCenter Repository Service registers the plug-in with the repository. The results of the registration operation appear in the activity log.

8. Run the PowerCenter Repository Service in normal mode.

Registering the PowerExchange for Google BigQuery Plug-in from the Command Line Program

You can use the pmrep RegisterPlugin command to register the plug-in from the command line program.

1. Run the PowerCenter Repository Service in exclusive mode.
2. Run the `pmrep Connect` command to connect to the Repository Service using a user account with Administrator Repository privilege.

The RegisterPlugin command uses the following syntax:

```
pmrep connect -r <repository name> -d <domain_name> -n <domain user name> -x  
               <domain_password>
```

3. Find `bigqueryPlugin.xml` in the following directory:
`<Informatica installation directory>\server\bin\Plugin`
4. Run the `pmrep RegisterPlugin` command to update the repository.

The RegisterPlugin command uses the following syntax:

```
pmrep registerplugin -i <Informatica installation directory>\server\bin\Plugin  
                      \bigqueryPlugin.xml -e -N
```

CHAPTER 3

Google BigQuery Sources and Targets

This chapter includes the following topics:

- [Google BigQuery Sources and Targets Overview, 11](#)
- [Import Google BigQuery Source and Target Definitions, 11](#)

Google BigQuery Sources and Targets Overview

You can create a mapping with a Google BigQuery source to extract data from Google BigQuery. You can create a mapping with any source and a Google BigQuery target to load data to Google BigQuery.

When the PowerCenter Integration Service extracts data from the source or loads data to the target, it converts the data based on the data types associated with the source or the target.

Import Google BigQuery Source and Target Definitions

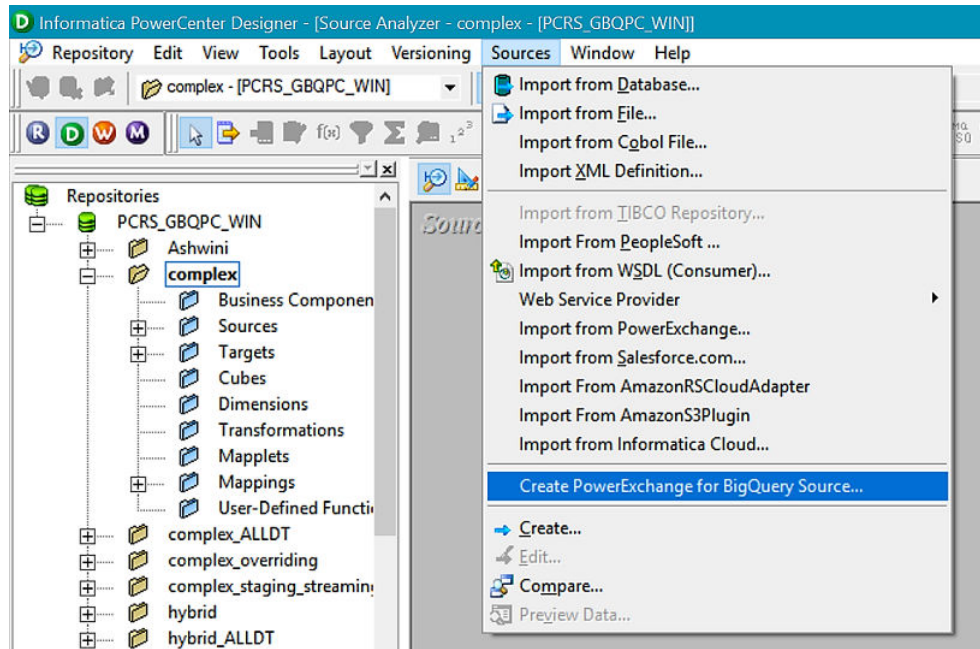
Use the **Create PowerExchange for BigQuery Source** or **Create PowerExchange for BigQuery Target** wizard to import Google BigQuery source and target definitions into the PowerCenter repository.

You must import Google BigQuery source and target objects before you create a mapping.

1. Start PowerCenter Designer, and connect to a PowerCenter repository configured with a Google BigQuery instance.
2. Open a source or target folder.

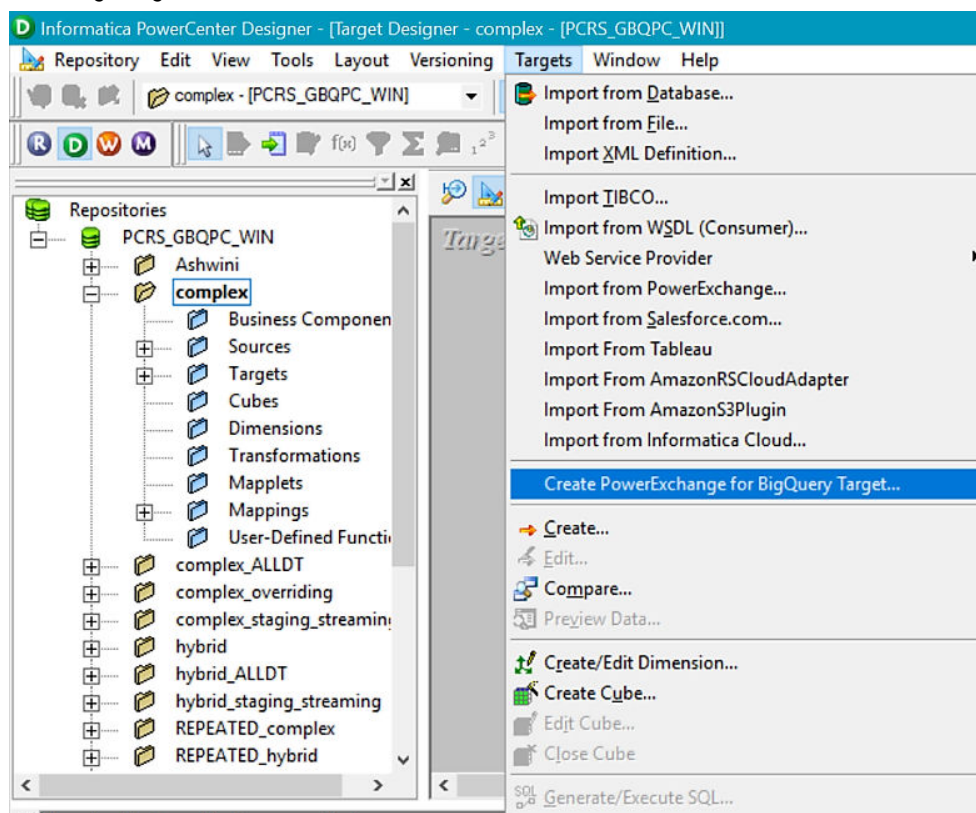
3. Select **Source Analyzer** or **Target Designer**.

- In the Source Analyzer, click **Sources** > **Create PowerExchange for BigQuery Source** as shown in the following image:



The **Google BigQuery Connection** wizard appears.

- In the Target Analyzer, click **Targets > Create PowerExchange for BigQuery Target** as shown in the following image:



The **Google BigQuery Connection** dialog box appears.

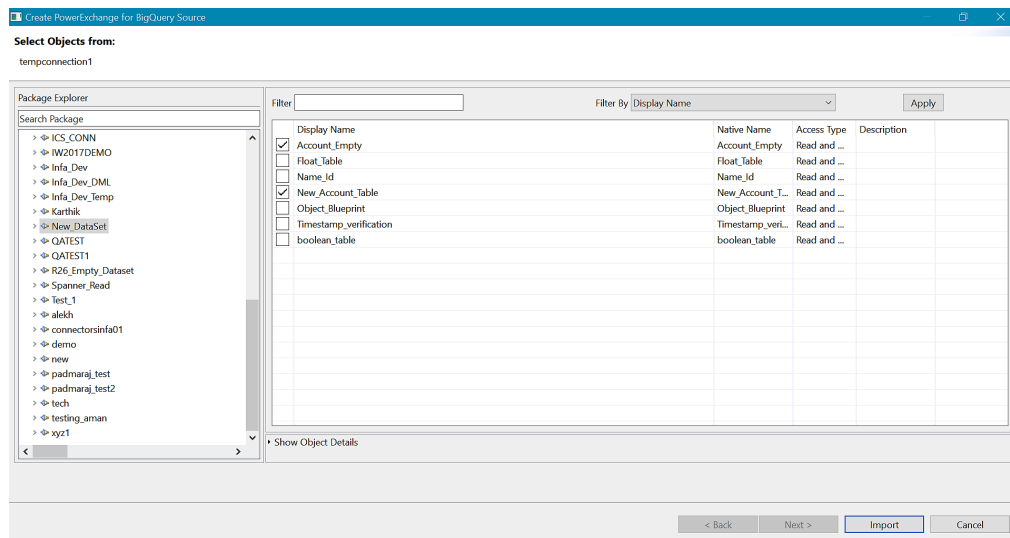
4. Configure the following connection parameters:

Connection Parameter	Description
Service Account ID	Specifies the client_email value present in the JSON file that you download after you create a service account.
Service Account Key	Specifies the private_key value present in the JSON file that you download after you create a service account.
Connection mode	<p>The mode that you want to use to read data from or write data to Google BigQuery.</p> <p>Select one of the following connection modes:</p> <ul style="list-style-type: none"> - Simple. Flattens each field within the Record data type field as a separate field in the mapping. - Hybrid. Displays all the top-level fields in the Google BigQuery table including Record data type fields. PowerExchange for Google BigQuery displays the top-level Record data type field as a single field of the String data type in the mapping. - Complex. Displays all the columns in the Google BigQuery table as a single field of the String data type in the mapping. <p>Default is Simple.</p>

Connection Parameter	Description
Schema Definition File Path	<p>Specifies a directory on the client machine where the PowerCenter Integration Service must create a JSON file with the sample schema of the Google BigQuery table. The JSON file name is the same as the Google BigQuery table name.</p> <p>Alternatively, you can specify a storage path in Google Cloud Storage where the PowerCenter Integration Service must create a JSON file with the sample schema of the Google BigQuery table. You can download the JSON file from the specified storage path in Google Cloud Storage to a local machine.</p>
Project ID	<p>Specifies the project_id value present in the JSON file that you download after you create a service account.</p> <p>If you have created multiple projects with the same service account, enter the ID of the project that contains the dataset that you want to connect to.</p>
Storage Path	<p>This property applies when you read or write large volumes of data.</p> <p>Path in Google Cloud Storage where the PowerCenter Integration Service creates a local stage file to store the data temporarily.</p> <p>You can either enter the bucket name or the bucket name and folder name.</p> <p>For example, enter <code>gs://<bucket_name></code> or <code>gs://<bucket_name>/<folder_name></code></p>

- Click **Test** to test the connection.
- Click **Finish** to add the connection.
- The **Select Objects from** tab appears.
- Select the dataset in **Package Explorer**.

A list of table appears as shown in the following image:



- Select the table that you want to import, and then click **Import**. You can import multiple tables from a Google BigQuery dataset.

To view the table metadata, select the table, and double-click the table name.

CHAPTER 4

Google BigQuery Mappings

This chapter includes the following topics:

- [Google BigQuery Mappings Overview, 15](#)
- [Source Filter, 15](#)

Google BigQuery Mappings Overview

After you import a Google BigQuery source or target definition into the PowerCenter repository, you can create a mapping to extract data from a Google BigQuery source or load data to a Google BigQuery target.

You can extract data from one or more Google BigQuery sources, and load data to one or more Google BigQuery targets.

You can enter a filter condition to reduce the number of source rows the PowerCenter Integration Service returns from Google BigQuery sources. You can enter a single filter condition or a series of conditions.

Note: You cannot preview data of a Google BigQuery source or target definition.

Source Filter

You can enter a filter condition to reduce the number of source rows the PowerCenter Integration Service returns from Google BigQuery sources. You can enter a single filter condition or a series of conditions.

Use the source filter in the Application Source Qualifier to retrieve rows from an entity that meet a condition.

You can provide a source filter to improve the performance when you read data from Google BigQuery.

You can use the Native or Platform expression to apply a filter condition on Google BigQuery columns of the following data types:

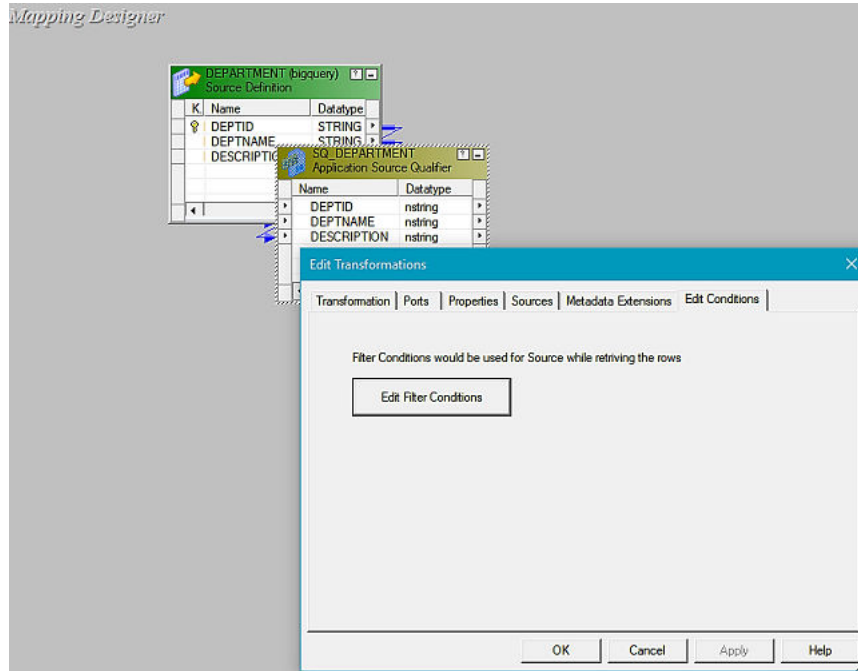
- Integer
- Float
- String
- Timestamp

Configuring a Source Filter

Configure a source filter in the Application Source Qualifier.

1. In the **Mapping Designer**, double-click the Application Source Qualifier.

The **Edit Transformations** dialog box appears as shown in the following image:



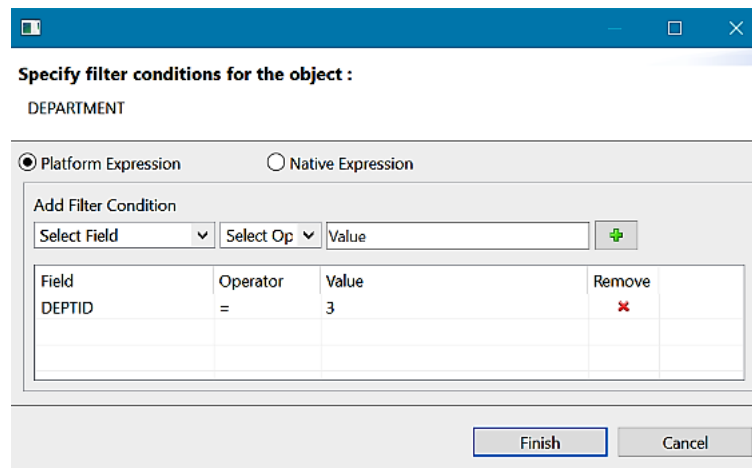
2. Click the **Edit Conditions** tab.
3. Click **Edit Filter Conditions**.

The **Add Filter Condition** dialog box appears.

4. Select **Platform Expression** or **Native Expression**.

- If you configure a platform expression, select the filter field and operator that you want to specify in the condition, enter a value for the condition, and click **Add Condition**

The following image shows a platform filter expression configured for a Google BigQuery source:



- If you configure a native expression, specify a filter expression in the following format:

<Attribute><Operator><Value>

You can use AND, OR, or nested conditions in the filter expression. The expression that you enter becomes the WHERE clause in the query used to retrieve records from the source.

The following image shows a native filter expression configured for a Google BigQuery source:

The screenshot shows a dialog box titled "Specify filter conditions for the object :". Below the title, the text "ALLDT_PRIMITIVE" is displayed. There are two radio buttons: "Platform Expression" (unselected) and "Native Expression" (selected). Below the radio buttons is a text area labeled "Enter your query here :". The text area contains the SQL filter expression: `((COL_INTEGER > 10 AND COL_FLOAT >= 387.8765) OR (COL_STRING = 'MICHAEL'))`. At the bottom right of the dialog are two buttons: "Finish" and "Cancel".

5. Click **Finish** to add the filter condition.
6. Click **OK**.

CHAPTER 5

Google BigQuery Sessions

This chapter includes the following topics:

- [Google BigQuery Sessions Overview, 18](#)
- [Google BigQuery Connections, 18](#)
- [Pre SQL and Post SQL Commands , 23](#)
- [Configure Google BigQuery Source Session Properties, 24](#)
- [Configure Google BigQuery Target Session Properties, 26](#)
- [Configure the Java Heap Memory, 30](#)
- [Rules and Guidelines for Google BigQuery Sessions, 30](#)

Google BigQuery Sessions Overview

After you create mappings, you can create a session to extract and load data.

You must configure a Google BigQuery connection in the **Workflow Manager** to extract data from or load data to a Google BigQuery table. You can define properties in a session to determine how the PowerCenter Integration Service must extract data from a Google BigQuery source or load data to a Google BigQuery target.

Google BigQuery Connections

Create a Google BigQuery connection to read data from a Google BigQuery source and write data to a Google BigQuery target. You must create a connection for each dataset that you want to connect to. You can use Google BigQuery connections in mappings. When you create a Google BigQuery connection, you can configure a connection mode based on how you want to read and write the data.

Connection Modes

You can configure a Google BigQuery connection to use one of the following connection modes:

Simple mode

If you use simple mode, PowerExchange for Google BigQuery flattens each field within the Record data type field as a separate field in the source or target definition.

Hybrid mode

If you use hybrid mode, PowerExchange for Google BigQuery displays all the top-level fields in the Google BigQuery table including Record data type fields. PowerExchange for Google BigQuery displays the top-level Record data type field as a single field of the String data type in the source or target definition.

Complex mode

If you use complex mode, PowerExchange for Google BigQuery displays all the columns in the Google BigQuery table as a single field of the String data type in the source or target definition.

Connection Mode Example

PowerExchange for Google BigQuery reads and writes the Google BigQuery data based on the connection mode that you configure for the Google BigQuery connection.

You have a Customers table in Google BigQuery that contains primitive fields and the **Address** field of the Record data type. The Address field contains two primitive sub-fields, **City** and **State**, of the String data type.

The following image shows the schema of the Customers table in Google BigQuery:

ID	INTEGER	NULLABLE
Name	STRING	NULLABLE
Address	RECORD	NULLABLE
Address.City	STRING	NULLABLE
Address.State	STRING	NULLABLE
Mobile	STRING	REPEATED
Totalpayments	FLOAT	NULLABLE
age	INTEGER	REPEATED

The following table shows the Customers table data in Google BigQuery:

ID	Name	Address.City	Address.State	Mobile	Totalpayments
14	John	LOS ANGELES	CALIFORNIA	+1-9744884744	18433.90
				+1-8267389993	
29	Jane	BOSTON	MANHATTAN	+1-8789390309	28397.33
				+1-9876553784	
				+1-8456437848	

Simple Mode

If you use simple connection mode, PowerExchange for Google BigQuery flattens each field within the Record data type field as a separate field in the **Source Analyzer**.

The following table shows two separate fields, Address_City and Address_State, for the respective sub-fields within the Address Record field in the Customers table:

ID	Name	Address_City	Address_State	Mobile	Totalpayments
14	John	LOS ANGELES	CALIFORNIA	+1-9744884744	18433.90
14	John	LOS ANGELES	CALIFORNIA	+1-8267389993	18433.90
29	Jane	BOSTON	MANHATTAN	+1-8789390309	28397.33
29	Jane	BOSTON	MANHATTAN	+1-9876553784	28397.33
29	Jane	BOSTON	MANHATTAN	+1-8456437848	28397.33

The following image shows the Address_State and Address_City fields in the **Source Analyzer**:

Key Types	Name	Datatype	Length/Precision
NOT A KEY	ID	INTEGER	19
NOT A KEY	Name	STRING	255
NOT A KEY	Address_State	STRING	255
NOT A KEY	Address_City	STRING	255
NOT A KEY	Mobile	STRING	255
NOT A KEY	TotalPayments	FLOAT	16
NOT A KEY	age	INTEGER	19

Hybrid Mode

If you use hybrid connection mode, PowerExchange for Google BigQuery displays all the top-level fields in the Google BigQuery table including Record data type fields. PowerExchange for Google BigQuery displays the top-level Record data type field as a single field of the String data type in the **Source Analyzer**.

The following image shows the Address field in the **Source Analyzer**:

Customers (bigquery)			
Key Types	Name	Datatype	Length/Precision
NOT A KEY	ID	INTEGER	19
NOT A KEY	Name	STRING	255
NOT A KEY	Address	STRING	65535
NOT A KEY	Mobile	STRING	255
NOT A KEY	TotalPayments	FLOAT	16
NOT A KEY	age	STRING	255

Complex Mode

If you use complex connection mode, PowerExchange for Google BigQuery displays all the columns in the Google BigQuery table as a single field of the String data type in the **Source Analyzer**.

The following image shows the STRING_DATA field in the **Source Analyzer**:

Customers (bigquery)			
Key Types	Name	Datatype	Length/Precision
NOT A KEY	STRING_DATA	STRING	65535

Rules and Guidelines for Google BigQuery Connection Modes

Simple Mode

Consider the following rules and guidelines when you configure a Google BigQuery connection to use simple connection mode:

- If the Google BigQuery source table contains repeated columns, you cannot configure source filters for these columns.
- If the Google BigQuery table contains more than one repeated column, you cannot preview data.
- If the Google BigQuery target table contains repeated columns, you cannot configure update and delete operations for these columns.
- When you read data from a Google BigQuery source, you must not map more than one repeated column in a single mapping. You must create multiple mappings for each repeated column.

Hybrid Mode

Consider the following rules and guidelines when you configure a Google BigQuery connection to use hybrid connection mode:

- If the Google BigQuery source table contains columns of the Record data type and repeated columns, you cannot configure source filters for these columns.
- You cannot configure update and delete operations for columns of the Record data type and repeated columns.
- You must select JSON (Newline Delimited) format as the data format of the staging file under the target session properties.

- You cannot use CSV format as the data format of the staging file. The following CSV formatting options in the target session properties are not applicable:
 - Allow Quoted Newlines
 - Field Delimiter
 - Allow Jagged Rows

Complex Mode

Consider the following rules and guidelines when you configure a Google BigQuery connection to use complex connection mode:

- When you configure a Google BigQuery source connection to use complex connection mode, you cannot configure data filters for the source.
- You cannot configure update and delete operations.
- You must select JSON (Newline Delimited) format as the data format of the staging file under the target session properties.
- You cannot use CSV format as the data format of the staging file. The following CSV formatting options in the target session properties are not applicable:
 - Allow Quoted Newlines
 - Field Delimiter
 - Allow Jagged Rows

Google BigQuery Connection Properties

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

The following table describes the Google BigQuery connection properties:

Property	Description
Service Account ID	Specifies the client_email value present in the JSON file that you download after you create a service account.
Service Account Key	Specifies the private_key value present in the JSON file that you download after you create a service account.
Connection mode	<p>The mode that you want to use to read data from or write data to Google BigQuery.</p> <p>Select one of the following connection modes:</p> <ul style="list-style-type: none"> - Simple. Flattens each field within the Record data type field as a separate field in the mapping. - Hybrid. Displays all the top-level fields in the Google BigQuery table including Record data type fields. PowerExchange for Google BigQuery displays the top-level Record data type field as a single field of the String data type in the mapping. - Complex. Displays all the columns in the Google BigQuery table as a single field of the String data type in the mapping. <p>Default is Simple.</p>

Property	Description
Schema Definition File Path	Specifies a directory on the client machine where the PowerCenter Integration Service must create a JSON file with the sample schema of the Google BigQuery table. The JSON file name is the same as the Google BigQuery table name. Alternatively, you can specify a storage path in Google Cloud Storage where the PowerCenter Integration Service must create a JSON file with the sample schema of the Google BigQuery table. You can download the JSON file from the specified storage path in Google Cloud Storage to a local machine.
Project ID	Specifies the project_id value present in the JSON file that you download after you create a service account. If you have created multiple projects with the same service account, enter the ID of the project that contains the dataset that you want to connect to.
Storage Path	This property applies when you read or write large volumes of data. Path in Google Cloud Storage where the PowerCenter Integration Service creates a local stage file to store the data temporarily. You can either enter the bucket name or the bucket name and folder name. For example, enter <code>gs://<bucket_name></code> or <code>gs://<bucket_name>/<folder_name></code>
Dataset ID	Name of the dataset that contains the source table and target table that you want to connect to.

Configuring a Google BigQuery Connection

Configure a Google BigQuery connection in the **Workflow Manager** to define the connection attributes that the PowerCenter Integration Service uses to connect to the Google BigQuery database.

1. In the Workflow Manager, click **Connections > Application**.
The **Application Connection Browser** dialog box appears.
2. Click **New**.
The **Select Subtype** dialog box appears.
3. Select **bigquery** and click **OK**.
The **Application Connection Editor** dialog box appears.
4. Enter a name for the Google BigQuery connection.
5. Enter the Google BigQuery connection attributes.
6. Click **OK** to create a Google BigQuery connection.

Pre SQL and Post SQL Commands

You can specify **pre SQL** and **post SQL** session properties for Google BigQuery sources and targets. When you create a Session task in the **Workflow Manager**, you can specify SQL commands on the **Mapping** tab.

You can perform the following operations by using pre SQL and post SQL commands:

- SELECT
- UPDATE

- DELETE

You can configure the options in Google BigQuery with a pre SQL or post SQL statement in the **pre SQL Configuration** or **post SQL Configuration** session properties for Google BigQuery sources and targets.

You must use the following format to specify a pre SQL configuration or a post SQL configuration:

```
<Option1:Value1,Option2:Value2,...OptionN:ValueN>
```

The following table shows the configuration options and supported values that you can specify in a pre SQL configuration or post SQL configuration:

Options	Supported Values
DestinationDataset	Dataset ID in Google BigQuery
DestinationTable	Table name in Google BigQuery
FlattenResults	True and False
UseLegacySQL	True and False
WriteDisposition	WRITE_TRUNCATE, WRITE_APPEND, and WRITE_EMPTY

Configure Google BigQuery Source Session Properties

You can configure the session properties for a Google BigQuery source on the **Mapping** tab. Define the properties for the source instance in the session.

The following table describes the session properties that you can configure for a Google BigQuery source session:

Property	Description
Source Dataset ID	Optional. Overrides the Google BigQuery dataset name that you specified in the connection.
Source Table Name	Overrides the Google BigQuery table name that you specified in the source.
Number of Rows to Read	Specifies the number of rows to read from the Google BigQuery source.
Allow Large Results	Determines whether PowerExchange for Google BigQuery must produce arbitrarily large result tables to query large source tables. If you select this option, you must specify a destination table to store the query results.
Query Results Table Name	Required if you select the Allow Large Results option. Specifies the destination table name to store the query results. If the table is not present in the dataset, PowerExchange for Google BigQuery creates the destination table with the name that you specify.

Property	Description
Job Poll Interval in Seconds	The number of seconds after which PowerExchange for Google BigQuery polls the status of the read job operation. Default is 10.
Read Mode	Specifies the read mode to read data from the Google BigQuery source. You can select one the following read modes: <ul style="list-style-type: none"> - Direct. In direct mode, PowerExchange for Google BigQuery reads data directly from the Google BigQuery source table. <p>Note: When you use hybrid and complex connection mode, you cannot use direct mode to read data from the Google BigQuery source.</p> <ul style="list-style-type: none"> - Staging. In staging mode, PowerExchange for Google BigQuery exports the data from the Google BigQuery source table into Google Cloud Storage. <p>After the download is complete, PowerExchange for Google BigQuery downloads the data from Google Cloud Storage into the local stage file that you specify into the local stage file and then reads data from the local stage file.</p> <p>Default is Direct mode.</p>
Number of Threads for Downloading Staging Files	Specifies the number of files that PowerExchange for Google BigQuery downloads at a time to enable parallel download. This property applies to staging mode.
Data Format of the staging file	Specifies the data format of the staging file. You can select one of the following data formats: <ul style="list-style-type: none"> - JSON (Newline Delimited). Supports flat and record data with nested and repeated fields. - CSV. Supports flat data. <p>Note: If you use hybrid and complex connection mode, you cannot use CSV format as the data format of the staging file.</p>
Local Stage File Directory	Specifies the directory on your local machine where PowerExchange for Google BigQuery stores Google BigQuery source data temporarily before it reads the data. This property applies to staging mode.
Staging File Name	Name of the staging file where data from the Google BigQuery source table is exported to Google Cloud Storage. This property applies to staging mode.
Enable Staging File Compression	Indicates whether to compress the size of the staging file in Google Cloud Storage before PowerExchange for Google BigQuery reads data from the staging file. You can enable staging file compression to reduce cost and transfer time. This property applies to staging mode.
Persist Destination Table	Indicates whether PowerExchange for Google BigQuery must persist the query results table after it reads data from the query results table. By default, PowerExchange for Google BigQuery deletes the query results table
pre SQL	SQL statement that you want to run before reading data from the source. For example, if you want to select records in the database before you read the records from the table, specify the following pre SQL statement: <pre>SELECT * FROM [api-project-80697026669:EMPLOYEE.DEPARTMENT] LIMIT 1000;</pre>

Property	Description
pre SQL Configuration	Specify a pre SQL configuration. For example, <code>DestinationTable:PRESQL_SRC, DestinationDataset:EMPLOYEE, FlattenResults:False, WriteDisposition:WRITE_TRUNCATE, UseLegacySql:False</code>
post SQL	SQL statement that you want to run after reading data from the source. For example, if you want to update records in a table after you read the records from a source table, specify the following post SQL statement: <code>UPDATE [api-project-80697026669.EMPLOYEE.PERSONS_TGT_DEL] SET phoneNumber.number=1000011, phoneNumber.areaCode=100 where fullname='John Doe' SET phoneNumber.number =1000011, phoneNumber.areaCode=100 where fullname='John Doe'</code>
post SQL Configuration	Specify a post SQL configuration. For example, <code>DestinationTable:POSTSQL_SRC, DestinationDataset:EMPLOYEE, FlattenResults:True, WriteDisposition:WRITE_TRUNCATE, UseLegacySql:False</code>
post SQL Configuration	Specify a post SQL configuration. For example, <code>DestinationTable:POSTSQL_SRC, DestinationDataset:CUSTOMERS, FlattenResults:True, WriteDisposition:WRITE_TRUNCATE, UseLegacySql:False</code>

Configure Google BigQuery Target Session Properties

You can configure the session properties for a Google BigQuery target on the **Mapping** tab. Define the properties for the target instance in the session.

The following table describes the session properties that you can configure for a Google BigQuery target session:

Property	Description
UpdateMode	<p>Determines the mode that PowerExchange for Google BigQuery uses to update rows in the Google BigQuery target.</p> <p>If you select an update mode, you need to select Update for the Treat Source Rows As session property in the Properties page.</p> <p>You can select one of the following modes:</p> <ul style="list-style-type: none"> - Update As Update. PowerExchange for Google BigQuery updates all rows flagged for update if the entries exist. - Update Else Insert. PowerExchange for Google BigQuery first updates all rows flagged for update if the entries exist in the target. If the entries do not exist, PowerExchange for Google BigQuery inserts the entries. <p>Default is Update As Update.</p>
Target Dataset ID	Optional. Overrides the Google BigQuery dataset name that you specified in the target definition.
Target Table Name	Optional. Overrides the Google BigQuery target table name that you specified in the target definition.
Create Disposition	<p>Specifies whether PowerExchange for Google BigQuery must create the target table if it does not exist.</p> <p>You can select one of the following values:</p> <ul style="list-style-type: none"> - Create if needed. If the table does not exist, PowerExchange for Google BigQuery creates the table. - Create never. If the table does not exist, PowerExchange for Google BigQuery does not create the table.
Write Disposition	<p>Specifies how PowerExchange for Google BigQuery must write data in bulk mode if the target table already exists.</p> <p>You can select one of the following values:</p> <ul style="list-style-type: none"> - Write append. If the target table exists, PowerExchange for Google BigQuery appends the data to the existing data in the table. - Write truncate. If the target table exists, PowerExchange for Google BigQuery overwrites the existing data in the table. - Write empty. If the target table exists and contains data, PowerExchange for Google BigQuery displays an error and does not write the data to the target. <p>PowerExchange for Google BigQuery writes the data to the target only if the target table does not contain any data.</p> <p>Note: Write disposition is applicable for bulk mode. Write disposition is applicable only when you perform an insert operation on a Google BigQuery target.</p>
	<p>Specifies the mode to write data to the Google BigQuery target.</p> <p>You can select one of the following modes:</p> <ul style="list-style-type: none"> - Bulk. In bulk mode, PowerExchange for Google BigQuery first writes the data to a staging file in Google Cloud Storage. <p>When the staging file contains all the data, PowerExchange for Google BigQuery loads the data from the staging file to the BigQuery target. Google BigQuery then deletes the staging file unless you configure the mapping to persist the staging file.</p> <ul style="list-style-type: none"> - Streaming. In streaming mode, PowerExchange for Google BigQuery directly writes data to the BigQuery target. PowerExchange for Google BigQuery writes the data into the target row by row. <p>Default is Bulk mode.</p>

Property	Description
Streaming Template Table Suffix	Specifies the suffix that PowerExchange for Google BigQuery adds to the individual target tables that it creates based on the template target table. This property applies to streaming mode.
Rows per Streaming Request	Specifies the number of rows that PowerExchange for Google BigQuery streams to the BigQuery target for each request. Default is 500 rows. The maximum row size that PowerExchange for Google BigQuery can stream to the BigQuery target for each request is 10 MB. This property applies to streaming mode.
Staging File Name	Name of the staging file that PowerExchange for Google BigQuery creates in the Google Cloud Storage before it loads the data to the Google BigQuery target. This property applies to bulk mode.
Data Format of the staging file	Specifies the data format of the staging file. You can select one of the following data formats: <ul style="list-style-type: none"> - JSON (Newline Delimited). Supports flat and record data with nested and repeated fields. - CSV. Supports flat data. Note: If you use hybrid and complex connection mode, you cannot use CSV format as the data format of the staging file.
Persist Staging File After Loading	Indicates whether PowerExchange for Google BigQuery must persist the staging file in the Google Cloud Storage after it writes the data to the Google BigQuery target. You can persist the staging file if you want to archive the data for future reference. By default, PowerExchange for Google BigQuery deletes the staging file in Google Cloud Storage. This property applies to bulk mode.
Enable Staging File Compression	Select this option to compress the size of the staging file before PowerExchange for Google BigQuery writes the data to the Google Cloud Storage and decompress the staging file before it loads the data to the Google BigQuery target. You can enable staging file compression to reduce cost and transfer time.
Job Poll Interval in Seconds	The number of seconds after which PowerExchange for Google BigQuery polls the status of the write job operation. Default is 10.
Number of Threads for Uploading Staging file	The number of files that PowerExchange for Google BigQuery must upload to Google Cloud Storage in bulk mode.
Local Stage File Directory	Specifies the directory on your local machine where PowerExchange for Google BigQuery stores the files temporarily before writing the data to the staging file in Google Cloud Storage. This property applies to bulk mode.
Allow Quoted Newlines	Indicates whether PowerExchange for Google BigQuery must allow the quoted data sections with newline character in a .csv file.
Field Delimiter	Delimiter character for the fields in a .csv file.

Property	Description
Allow Jagged Rows	Indicates whether PowerExchange for Google BigQuery must accept the rows without trailing columns in a .csv file.
Pre SQL	<p>SQL statement that you want to run before writing data to the target.</p> <p>For example, if you want to select records from the database before you write the records into the table, specify the following pre SQL statement:</p> <pre>SELECT * FROM `api-project-80697026669.EMPLOYEE.RegionNation` LIMIT 1000</pre>
Pre SQL Configuration	<p>Specify a pre SQL configuration.</p> <p>For example,</p> <pre>DestinationTable:PRESQL_TGT2, DestinationDataset:EMPLOYEE, FlattenResults:False, WriteDisposition:WRITE_TRUNCATE, UseLegacySql:False</pre>
Post SQL	<p>SQL statement that you want to run after writing the data into the target.</p> <p>For example, if you want to update records in a table after you write the records into the target table, specify the following post SQL statement:</p> <pre>UPDATE [api-project-80697026669.EMPLOYEE.PERSONS_TGT_DEL] SET phoneNumber.number =1000011, phoneNumber.areaCode=100 where fullname='John Doe'</pre>
Post SQL Configuration	<p>Specify a post SQL configuration.</p> <p>For example,</p> <pre>DestinationTable:POSTSQL_SRC, DestinationDataset:EMPLOYEE, FlattenResults:True, UseLegacySQL:False</pre>
INSERT	Inserts all rows to the Google BigQuery target.
DELETE	<p>Deletes rows from the Google BigQuery target.</p> <p>If you select DELETE, you need to select Delete for the Treat Source Rows As session property in the Properties page.</p>
UPDATE	<p>Not applicable for PowerExchange for Google BigQuery.</p> <p>Note: Configure the update strategy for a target object in the UpdateMode session property.</p>
Success File Directory	Not applicable for PowerExchange for Google BigQuery.
Error File Directory	Not applicable for PowerExchange for Google BigQuery.
Forward Rejected Rows	Not applicable for PowerExchange for Google BigQuery.

Configure the Java Heap Memory

For the Google BigQuery session to run successfully, configure the memory for the Java heap size in the node that runs the PowerCenter Integration Service.

1. In the Administrator tool, navigate to the PowerCenter Integration Service for which you want to change the Java heap size.
2. Click the **Processes** tab.
3. Edit the **General Properties** section.
4. Specify the maximum heap size in **Java SDK Maximum Memory** limit based on the amount data you want to process.
5. Click **Ok**.
6. Restart the PowerCenter Integration Service.

Rules and Guidelines for Google BigQuery Sessions

Use the following rules and guidelines when you create a session:

- You cannot configure a Lookup transformation to look up data in a Google BigQuery source.
- You cannot configure key range partitioning to read data from a Google BigQuery source.
- You cannot read hierarchical data from a Google BigQuery source, convert the hierarchical input to relational output, and write data to relational targets.
- You cannot read data from relational sources, convert the relational input to hierarchical output, and write data to a Google BigQuery target.
- You cannot create a target at run time.
- You must define a primary key in the target table for update and delete operations.
If you do not define a primary key in the target table, the mapping fails to update records in or delete records from the target table.
- You cannot perform more than one operation with a pre SQL or post SQL command.
- If you perform an UPDATE or DELETE operation with a pre SQL or post SQL command, you must specify the following parameter in the pre SQL configuration or post SQL configuration:
`UseLegacySQL:False`
- When you use staging mode to read data from Google BigQuery or bulk mode to write data to Google BigQuery, you must increase the maximum heap size in the **Java SDK Maximum Memory** property to 2048 MB or 4096 MB based on the amount data you want to process.

APPENDIX A

Google BigQuery Data Type Reference

This appendix includes the following topics:

- [Data Type Reference Overview, 31](#)
- [Google BigQuery and Transformation Data Types, 31](#)

Data Type Reference Overview

PowerCenter uses the following data types in Google BigQuery mappings:

- Google BigQuery native data types. Google BigQuery data types appear in Google BigQuery definitions in a mapping.
- 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 PowerCenter Integration Service uses to move data across platforms. They appear in all transformations in a mapping.

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

Google BigQuery and Transformation Data Types

The following table lists the Google BigQuery data types that PowerCenter supports and the corresponding transformation data types:

Google BigQuery Data Type	Transformation Data Type	Range and Description for the Transformation Data Type
BOOLEAN	Nstring	1 to 104,857,600 characters
BYTE	Byte	1 to 104,857,600 bytes

Google BigQuery Data Type	Transformation Data Type	Range and Description for the Transformation Data Type
DATE	Date/Time	Jan 1, 0001 A.D. to Dec 31, 9999 A.D. (precision to the nanosecond)
DATETIME	Date/Time	Jan 1, 0001 A.D. to Dec 31, 9999 A.D. (precision to the nanosecond)
FLOAT	Double	Precision 15
INTEGER	Bigint	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision 19, scale 0
RECORD	Nstring	1 to 104,857,600 characters
STRING	Nstring	1 to 104,857,600 characters
TIME	Date/Time	Jan 1, 0001 A.D. to Dec 31, 9999 A.D. (precision to the nanosecond)
TIMESTAMP	Date/Time	Jan 1, 0001 A.D. to Dec 31, 9999 A.D. (precision to the nanosecond)

INDEX

C

- configuration
 - overview [9](#)
- configuring source filter [16](#)

F

- filter
 - application source qualifier [15](#)
 - source filter [15](#)
- filter conditions [16](#)

G

- Google BigQuery
 - mappings [15](#)
 - overview [7](#)
 - source [11](#)
 - target [11](#)
- Google BigQuery connections
 - properties [22](#)
 - connection configuration [23](#)
 - connection modes [18](#)
 - connection modes example [19](#)
 - overview [18](#)
- Google BigQuery data types
 - transformation data types [31](#)

I

- Import
 - connection properties [11](#)
 - source definition [11](#)
 - target definition [11](#)

J

- java heap size [30](#)

M

- mapping
 - filter [15](#)
 - source [15](#)
 - target [15](#)

P

- plug-in registration
 - administrator tool [10](#)
- post SQL commands
 - entering [23](#)
- PowerExchange for Google BigQuery
 - administration [8](#)
 - configuration [9](#)
 - overview [7](#)
 - plug-in registration [9](#)
- pre SQL and post SQL
 - entering [23](#)
- pre SQL commands
 - entering [23](#)

S

- session
 - overview [18](#)
 - source properties [24](#)
 - target properties [26](#)