



Informatica® PowerExchange for Google
BigQuery
10.2 HotFix 1

User Guide

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

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

Portions of this software and/or documentation are subject to copyright held by third parties. 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: 2018-09-25

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. | 8 |
| Administration of Google BigQuery. | 8 |
| Chapter 2: PowerExchange for Google BigQuery Installation and Configuration..... | 9 |
| PowerExchange for Google BigQuery Installation and Configuration Overview. | 9 |
| Prerequisites. | 9 |
| Installing the Server Component on Linux. | 10 |
| Installing the Client Component. | 10 |
| Chapter 3: Google BigQuery Connections..... | 12 |
| Google BigQuery Connections Overview. | 12 |
| Connection Modes. | 12 |
| Connection Mode Example. | 13 |
| Rules and Guidelines for Google BigQuery Connection Modes. | 15 |
| Google BigQuery Connection Properties. | 16 |
| Creating a Google BigQuery Connection. | 17 |
| Chapter 4: PowerExchange for Google BigQuery Data Objects..... | 18 |
| Google BigQuery Data Object Overview. | 18 |
| Google BigQuery Data Object Properties. | 18 |
| Pre SQL and Post SQL Commands. | 19 |
| Google BigQuery Data Object Read Operation. | 20 |
| Output Properties of the Data Object Read Operation. | 20 |
| Google BigQuery Data Object Write Operation. | 22 |
| Input Properties of the Data Object Write Operation. | 23 |
| Creating an Google BigQuery Data Object. | 26 |
| Creating a Google BigQuery Data Object Operation. | 26 |

| | |
|--|-----------|
| Rules and Guidelines for PowerExchange for Google BigQuery Data Object Operations. | 27 |
| Chapter 5: PowerExchange for Google BigQuery Mappings..... | 28 |
| PowerExchange for Google BigQuery Mappings Overview. | 28 |
| Mapping Validation and Run-time Environments. | 28 |
| Chapter 6: Google BigQuery Run-Time Processing..... | 30 |
| Google BigQuery Run-Time Processing Overview. | 30 |
| Filter Expression. | 30 |
| Native Expression. | 30 |
| Platform Expression. | 31 |
| Partitioning. | 31 |
| Parameterization for Google BigQuery Sources. | 32 |
| Parameterization for Google BigQuery Targets. | 32 |
| Appendix A: Google BigQuery Data Type Reference..... | 34 |
| Data Type Reference Overview. | 34 |
| Google BigQuery and Transformation Data Types. | 34 |
| Index. | 36 |

Preface

The *Informatica PowerExchange for Google BigQuery User Guide* 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 that extract data from and load data to Google BigQuery. This guide assumes you have knowledge of Google BigQuery and Informatica Data Services.

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, 8](#)
- [Administration of Google BigQuery, 8](#)

PowerExchange for Google BigQuery Overview

You can use PowerExchange for Google BigQuery to extract data from and load data to 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. PowerExchange for Google BigQuery uses the Google APIs to integrate with Google BigQuery and Google Cloud Storage.

You can validate and run Google BigQuery mappings in the native environment or on the Spark engine in the Hadoop environment.

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 Developer Tool.
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 Data 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 Data Integration Service creates the staging file to store the data from the Google BigQuery source.

CHAPTER 2

PowerExchange for Google BigQuery Installation and Configuration

This chapter includes the following topics:

- [PowerExchange for Google BigQuery Installation and Configuration Overview, 9](#)
- [Prerequisites, 9](#)
- [Installing the Server Component on Linux, 10](#)
- [Installing the Client Component, 10](#)

PowerExchange for Google BigQuery Installation and Configuration Overview

The PowerExchange for Google BigQuery installation consists of a server installation and a client installation. You can install PowerExchange for Google BigQuery server component on a Linux 64-bit machine.

Prerequisites

The PowerExchange for Google BigQuery installation consists of a server installation and a client installation.

To use PowerExchange for Google BigQuery, perform the following steps:

1. Install or upgrade Informatica services. Create and configure a Model Repository Service and a Data Integration Service.
2. Verify that you have write permissions on all the directories within the `<Informatica installation directory> directory`.
3. Download the PowerExchange for Google BigQuery server component and client component from the TSFTP location to a directory on your machine.
For more information, contact Informatica Global Customer Support.

Installing the Server Component on Linux

Install the PowerExchange for Google BigQuery server component after you install the Informatica services. The server binaries are copied to the Informatica installation directory.

If multiple nodes exist in your environment, you must first install the server component on the master gateway node. You can then install the server component on the other nodes in the domain. Before you install, shut down the Informatica domain.

1. Delete the contents from the following directories:
 - `<Informatica installation directory>/services/work_dir`
 - `<Informatica installation directory>/tomcat/bin/workspace`
2. Extract the server installation file to a directory on the Linux machine.
3. Open command prompt and navigate to the root directory of the extracted installer files.
4. Enter `./install.sh` at the command prompt.

Note: The `install.sh` file must have executable permissions.
5. Enter the path to the Informatica installation directory.

By default, the server components are installed in the following directory:

`<User Home Directory>/Informatica/<version folder>`
6. Review the installation information and press **Enter** to begin the installation.
7. After you complete installation, open command prompt and navigate to the following directory:

`<Informatica installation directory>/isp/bin/`
8. Enter the following command in the command prompt:

```
infasetup.sh validateandregisterfeature -ff <Informatica installation directory>/plugins/
conf/google_bigquery.xml -up false
```

Installing the Client Component

Install the PowerExchange for Google BigQuery client component after you install the Informatica clients. The client binaries are copied to the Informatica installation directory. Install the client component on every Informatica Developer client machine that connects to the domain.

1. Delete the contents from the following directory:

`<Informatica installation directory>\clients\DeveloperClient\workspace`
2. Delete the configuration files and retain the `config.ini` file from the following directory:

`<Informatica installation directory>\clients\DeveloperClient\configuration`
3. Extract the client installation file to a directory on your machine and navigate to the root directory of the extracted installer files.
4. Run the `install.bat` script file.

The **Welcome** page appears.
5. Click **Next**.

The **Installation Directory** page appears.

6. Enter the absolute path to the Informatica installation directory. Click **Browse** to find the directory or use the default directory.
7. Click **Next**.
The **Pre-Installation Summary** page appears.
8. Verify that all installation requirements are met and click **Install**.
The installer shows the progress of the installation. When the installation is complete, the **Post-Installation Summary** page displays the status of the installation.
9. Click **Done** to close the installer.
For more information about the tasks performed by the installer, view the installation log files.

CHAPTER 3

Google BigQuery Connections

This chapter includes the following topics:

- [Google BigQuery Connections Overview, 12](#)
- [Connection Modes, 12](#)
- [Connection Mode Example, 13](#)
- [Rules and Guidelines for Google BigQuery Connection Modes, 15](#)
- [Google BigQuery Connection Properties, 16](#)
- [Creating a Google BigQuery Connection, 17](#)

Google BigQuery Connections Overview

Use a Google BigQuery connection to access a Google BigQuery database.

Use the Google BigQuery connection to import Google BigQuery metadata, create data objects, preview data, and run mappings. When you create a Google BigQuery connection, you define the connection attributes that the Developer tool uses to connect to the Google BigQuery dataset.

Use the Developer tool, Administrator tool, or infacmd to create a Google BigQuery connection.

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 Google BigQuery data object.

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 Google BigQuery data object.

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 Google BigQuery data object.

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 **Columns** tab of the Google BigQuery data object.

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 **Columns** tab of the Google BigQuery data object:

Object Details:

| General Columns Advanced | | | | | | | | | |
|--------------------------|---------------|---------------|---------|-----------|-------|-------------------------------------|--------------------------|----------------|-------------|
| | Name | Native Name | Type | Precision | Scale | Repeated | isRecord | Access Type | Description |
| 1 | ID | ID | INTEGER | 19 | 0 | <input type="checkbox"/> | <input type="checkbox"/> | Read and Write | |
| 2 | Name | Name | STRING | 255 | 0 | <input type="checkbox"/> | <input type="checkbox"/> | Read and Write | |
| 3 | Address_State | Address.State | STRING | 255 | 0 | <input type="checkbox"/> | <input type="checkbox"/> | Read and Write | |
| 4 | Address_City | Address.City | STRING | 255 | 0 | <input type="checkbox"/> | <input type="checkbox"/> | Read and Write | |
| 5 | Mobile | Mobile | STRING | 255 | 0 | <input checked="" type="checkbox"/> | <input type="checkbox"/> | Read and Write | |
| 6 | TotalPayments | TotalPayments | FLOAT | 16 | 0 | <input type="checkbox"/> | <input type="checkbox"/> | Read and Write | |
| 7 | age | age | INTEGER | 19 | 0 | <input checked="" type="checkbox"/> | <input type="checkbox"/> | Read and Write | |

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 **Columns** tab of the Google BigQuery data object.

The following image shows the Address field in the **Columns** tab of the Google BigQuery data object:

Object Details:

| General Columns Advanced | | | | | | | | | |
|--------------------------|---------------|---------------|---------|-----------|-------|-------------------------------------|-------------------------------------|----------------|-------------|
| | Name | Native Name | Type | Precision | Scale | Repeated | isRecord | Access Type | Description |
| 1 | ID | ID | INTEGER | 19 | 0 | <input type="checkbox"/> | <input type="checkbox"/> | Read and Write | |
| 2 | Name | Name | STRING | 255 | 0 | <input type="checkbox"/> | <input type="checkbox"/> | Read and Write | |
| 3 | Address | Address | STRING | 65535 | 0 | <input type="checkbox"/> | <input checked="" type="checkbox"/> | Read and Write | |
| 4 | Mobile | Mobile | STRING | 255 | 0 | <input checked="" type="checkbox"/> | <input type="checkbox"/> | Read and Write | |
| 5 | TotalPayments | TotalPayments | FLOAT | 16 | 0 | <input type="checkbox"/> | <input type="checkbox"/> | Read and Write | |
| 6 | age | age | STRING | 255 | 0 | <input checked="" type="checkbox"/> | <input type="checkbox"/> | Read and Write | |

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 **Columns** tab of the Google BigQuery data object.

The following image shows the STRING_DATA field in the **Columns** tab of the Google BigQuery data object:

Object Details:

| General Columns Advanced | | | | | | | | | |
|--------------------------|-------------|-------------|--------|-----------|-------|--------------------------|--------------------------|----------------|-------------|
| | Name | Native Name | Type | Precision | Scale | Repeated | isRecord | Access Type | Description |
| 1 | STRING_DATA | STRING DATA | STRING | 65535 | 0 | <input type="checkbox"/> | <input type="checkbox"/> | Read and Write | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |

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 data object contains repeated columns, you cannot configure filters for these columns in a data object read operation.
- If the Google BigQuery table contains more than one repeated column, you cannot preview data.
- If the Google BigQuery data object contains repeated columns, you cannot configure update and delete operations for these columns in a data object write operation.
- 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 data object contains columns of the Record data type and repeated columns, you cannot configure filters for these columns in a data object read operation.
- 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 data object write operation advanced properties.
- You cannot use CSV format as the data format of the staging file. The following CSV formatting options in the data object write operation advanced 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 connection to use complex connection mode, you cannot configure filters in a data object read operation.

- You cannot configure update and delete operations.
- You must select JSON (Newline Delimited) format as the data format of the staging file under the data object write operation advanced properties.
- You cannot use CSV format as the data format of the staging file. The following CSV formatting options in the data object write operation advanced properties are not applicable:
 - Allow Quoted Newlines
 - Field Delimiter
 - Allow Jagged Rows

Google BigQuery Connection Properties

When you set up a Google BigQuery connection, you must configure the connection properties.

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

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 in Google BigQuery. |
| Service Account Key | Specifies the private_key value present in the JSON file that you download after you create a service account in Google BigQuery. |
| 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> |
| Schema Definition File Path | <p>Specifies a directory on the client machine where the Data 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 Data 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> |

| Property | Description |
|--------------|---|
| Project ID | Specifies the project_id value present in the JSON file that you download after you create a service account in Google BigQuery. 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 Data 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> |

Creating a Google BigQuery Connection

Create a Google BigQuery connection before you create a Google BigQuery data object.

1. In the Developer tool, click **Window > Preferences**.
2. Select **Informatica > Connections**.
3. Expand the domain in the **Available Connections**.
4. Select the connection type **Enterprise Application > Google Big Query**, and click **Add**.
5. Enter a connection name and an optional description.
6. Select **Google Big Query** as the connection type.
7. Click **Next**.
8. Configure the connection properties.
9. Click **Test Connection** to verify the connection to Google BigQuery.
10. Click **Finish**.

CHAPTER 4

PowerExchange for Google BigQuery Data Objects

This chapter includes the following topics:

- [Google BigQuery Data Object Overview, 18](#)
- [Google BigQuery Data Object Properties, 18](#)
- [Pre SQL and Post SQL Commands, 19](#)
- [Google BigQuery Data Object Read Operation, 20](#)
- [Google BigQuery Data Object Write Operation, 22](#)
- [Creating an Google BigQuery Data Object, 26](#)
- [Creating a Google BigQuery Data Object Operation, 26](#)
- [Rules and Guidelines for PowerExchange for Google BigQuery Data Object Operations, 27](#)

Google BigQuery Data Object Overview

A Google BigQuery data object is a physical data object that uses Google BigQuery as a source or target. A Google BigQuery data object is a physical data object that represents data based on a Google BigQuery resource.

You can configure the data object read and write operation properties that determine how you can read data from Google BigQuery sources and load data to Google BigQuery targets.

Create a Google BigQuery data object from the Developer tool. PowerExchange for Google BigQuery creates the data object read operation and data object write operation for the Google BigQuery data object. You can edit the advanced properties of the data object read or write operation and run a mapping.

Google BigQuery Data Object Properties

Specify the data object properties when you create the data object.

The following table describes the properties that you configure for the Google BigQuery data objects:

| Property | Description |
|------------|--|
| Name | Name of the Google BigQuery data object. |
| Location | The project or folder in the Model Repository Service where you want to store the Google BigQuery data object. |
| Connection | Name of the Google BigQuery connection. |

Pre SQL and Post SQL Commands

You can specify **pre SQL** and **post SQL** data object read and write operation properties for Google BigQuery data objects. When you create a data object read or write operation in the Developer tool, you can specify SQL commands on the **Advanced** view.

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** data object read and write operation properties for Google BigQuery data objects.

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 |

Google BigQuery Data Object Read Operation

Create a mapping with a Google BigQuery data object read operation to read data from Google BigQuery.

Output Properties of the Data Object Read Operation

The output properties represent data that the Data Integration Service passes into the mapping pipeline. Select the output properties to configure advanced properties of the data object read operation.

The output properties of the data object read operation include general properties that apply to the data object operation. The output properties also include source, query, run-time, and advanced properties that apply to the Google BigQuery data object.

You can view and change the output properties of the data object read operation from the **General**, **Sources**, **Query**, **Run-time**, and **Advanced** tabs.

General Properties

The general properties display the name and description of the data object read operation.

Sources Properties

The sources properties list the Google BigQuery objects used in the data object read operation. You cannot join data from multiple sources of the Google BigQuery data object in a read operation.

Query Properties

Use the **Query** tab to select specific records from a Google BigQuery table.

The following table describes the query properties that you configure for a data object read operation:

| Property | Description |
|----------|--|
| Query | Filter value in a read operation. The filter specifies the where clause of select statement. Use a filter to reduce the number of rows that the Data Integration Service reads from the source. When you enter a source filter, the Developer tool adds a WHERE clause to the default query. You can use the Native or Platform expression to select specific records. |

Run-time Properties

The run-time properties displays the name of the connection that the Data Integration Service uses to read data from the Google BigQuery table.

You can define the partition type as key range partitioning to read data from Google BigQuery data object that use simple or hybrid connection mode. When you configure key range partitioning, the Data Integration Service distributes rows of data based on a port or set of ports that you define as the partition key. You can define a range of values for each port. The Data Integration Service uses the key and ranges to send rows to the appropriate partition.

Advanced Properties

The Data Integration Service reads data from Google BigQuery based on the data object read operation.

The Developer tool displays advanced properties for the Google BigQuery data object operation in the Advanced view.

The following table describes the advanced properties for a Google BigQuery data object read operation:

| Property | Description |
|---|--|
| Source Dataset ID | Optional. Overrides the Google BigQuery dataset name specified in the Google BigQuery data object read operation. |
| Source Table Name | Overrides the Google BigQuery table name specified in the Google BigQuery data object read operation. |
| 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. |
| 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. |

| Property | Description |
|---------------------------------|--|
| 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: <code>SELECT * FROM [api-project-80697026669:EMPLOYEE.DEPARTMENT] LIMIT 1000;</code> |
| 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> |

Google BigQuery Data Object Write Operation

Create a mapping to write data to Google BigQuery. Use the Google BigQuery connection, and define the write operation properties to write data to Google BigQuery.

You can perform insert, update, delete, and upsert operations on a Google BigQuery target.

To perform a update, delete, or upsert operation, you must add an Update Strategy transformation to a mapping and flag the records in an update strategy expression.

For more information about configuring an Update Strategy transformation, see "Update Strategy Transformation" in the *Informatica Developer Transformation Guide*.

Input Properties of the Data Object Write Operation

Input properties represent data that the Data Integration Service writes to a Google BigQuery table. Select the input properties to configure run-time properties of the data object write operation. You can also specify advanced data object write operation properties to write data to Google BigQuery table.

The input properties of the data object write operation include general properties that apply to the data object write operation. Input properties also include source, run-time, and advanced properties that apply to the data object write operation.

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

General Properties

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

Target Properties

The target properties list the Google BigQuery table in the data object write operation.

Run-time Properties

The run-time properties displays the name of the connection that the Data Integration Service uses to write data to the Google BigQuery table.

You can also use the **Run-time** tab to configure partitioning. You can configure dynamic partitioning for Google BigQuery data object write operation.

Advanced Properties

Google BigQuery data object write operation properties include advanced properties that apply to the Google BigQuery data object.

The Developer tool displays advanced properties for the Google BigQuery data object operation in the **Advanced** tab.

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

| Property | Description |
|-------------------|--|
| UpdateMode | <p>Determines the mode that the Data Integration Service uses to update rows in the Google BigQuery target.</p> <p>You can select one of the following modes:</p> <ul style="list-style-type: none">- Update As Update. The Data Integration Service updates all rows flagged for update in the Update Strategy transformation if the entries exist.- Update Else Insert. The Data Integration Service first updates all rows flagged for update in the Update Strategy transformation if the entries exist in the target. If the entries do not exist, the Data Integration Service inserts the entries. <p>Default is Update As Update.</p> |
| Target Dataset ID | <p>Optional. Overrides the Google BigQuery dataset name specified in the Google BigQuery data object write operation.</p> |

| Property | Description |
|---------------------------------|--|
| Target Table Name | Optional. Overrides the Google BigQuery target table name specified in the Google BigQuery data object write operation. |
| 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> |
| Streaming Template Table Suffix | <p>Specifies the suffix that PowerExchange for Google BigQuery adds to the individual target tables that it creates based on the template target table.</p> <p>This property applies to streaming mode.</p> |
| Rows per Streaming Request | <p>Specifies the number of rows that PowerExchange for Google BigQuery streams to the BigQuery target for each request.</p> <p>Default is 500 rows.</p> <p>The maximum row size that PowerExchange for Google BigQuery can stream to the BigQuery target for each request is 10 MB.</p> <p>This property applies to streaming mode.</p> |
| Staging File Name | <p>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.</p> <p>This property applies to bulk mode.</p> |

| Property | Description |
|--|--|
| Data Format of the staging file | <p>Specifies the data format of the staging file.</p> <p>You can select one of the following data formats:</p> <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> |
| Persist Staging File After Loading | <p>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.</p> <p>By default, PowerExchange for Google BigQuery deletes the staging file in Google Cloud Storage. This property applies to bulk mode.</p> |
| Enable Staging File Compression | <p>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.</p> <p>You can enable staging file compression to reduce cost and transfer time.</p> |
| Job Poll Interval in Seconds | <p>The number of seconds after which PowerExchange for Google BigQuery polls the status of the write job operation.</p> <p>Default is 10.</p> |
| Number of Threads for Uploading Staging file | <p>The number of files that PowerExchange for Google BigQuery must upload to Google Cloud Storage in bulk mode.</p> |
| Local Stage File Directory | <p>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.</p> <p>This property applies to bulk mode.</p> <p>Note: This property is not applicable if you run mappings with the Spark engine in the Hadoop environment.</p> |
| Allow Quoted Newlines | <p>Indicates whether PowerExchange for Google BigQuery must allow the quoted data sections with newline character in a .csv file.</p> |
| Field Delimiter | <p>Delimiter character for the fields in a .csv file.</p> |
| Allow Jagged Rows | <p>Indicates whether PowerExchange for Google BigQuery must accept the rows without trailing columns in a .csv file.</p> |
| 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> |

| Property | Description |
|------------------------|--|
| 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> |

Creating an Google BigQuery Data Object

Create a Google BigQuery data object to add to a mapping.

1. Select a project or folder in the **Object Explorer** view.
2. Click **File > New > Data Object**.
3. Select **Bigquery Data Object** and click **Next**.
The **Bigquery Data Object** dialog box appears.
4. Enter a name for the data object.
5. Click **Browse** next to the **Location** option and select the target project or folder.
6. Click **Browse** next to the **Connection** option and select the Google BigQuery connection from which you want to import the Google BigQuery object.
7. To add a resource, click **Add** next to the **Selected Resources** option.
The **Add Resource** dialog box appears.
8. Select the checkbox next to the Google BigQuery object you want to add and click **OK**.
9. Click **Finish**.
The data object appears under Data Objects in the project or folder in the **Object Explorer** view.

Creating a Google BigQuery Data Object Operation

You can create the data object read, write, or lookup operation for Google BigQuery data objects. You can the add the Google BigQuery data object operation to a mapping.

1. Select the data object in the **Object Explorer** view.
2. Right-click and select **New > Data Object Operation**.
The **Data Object Operation** dialog box appears.
3. Enter a name for the data object operation.

4. Select the type of data object operation. You can choose to create a read or write operation.
5. Click **Add**.
The **Select Resources** dialog box appears.
6. Select the Google BigQuery data object for which you want to create the data object operation and click **OK**.
7. Click **Finish**.

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

Rules and Guidelines for PowerExchange for Google BigQuery Data Object Operations

Use the following rules and guidelines when you configure a PowerExchange for Google BigQuery data object read and write operation:

- You cannot configure a Google BigQuery data object lookup operation to lookup data from a Google BigQuery source based on a lookup condition.
- You must define a primary key in the source table and 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 read hierarchical data from a Google BigQuery source, convert the hierarchical input to relational output, and write data to relational targets.
- 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`

CHAPTER 5

PowerExchange for Google BigQuery Mappings

This chapter includes the following topics:

- [PowerExchange for Google BigQuery Mappings Overview, 28](#)
- [Mapping Validation and Run-time Environments, 28](#)

PowerExchange for Google BigQuery Mappings Overview

After you create a Google BigQuery data object read or write operation, you can create a mapping to extract data from a Google BigQuery source or load data to a Google BigQuery target.

You can define properties in an operation to determine how the Data Integration Service must 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. When the Data 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.

Mapping Validation and Run-time Environments

You can validate and run mappings in the native environment or with the Spark engine in the Hadoop environment.

The Data Integration Service validates whether the mapping can run in the selected environment. You must validate the mapping for an environment before you run the mapping in that environment.

Native environment

You can configure the mappings to run in the native environment. When you run mappings in the native environment, the Data Integration Service processes the mapping and runs the mapping from the Developer tool.

Spark Engine

When you run mappings on the Spark engine, the Data Integration Service pushes the mapping to a Hadoop cluster and processes the mapping on a Spark engine. The Data Integration Service generates an execution plan to run mappings on the Spark engine.

You can view the plan in the Developer tool before you run the mapping and in the Administrator tool after you run the mapping.

For more information about the Hadoop environment and Spark engines, see the *Informatica Big Data Management™ User Guide*.

CHAPTER 6

Google BigQuery Run-Time Processing

This chapter includes the following topics:

- [Google BigQuery Run-Time Processing Overview, 30](#)
- [Filter Expression, 30](#)
- [Partitioning, 31](#)
- [Parameterization for Google BigQuery Sources, 32](#)
- [Parameterization for Google BigQuery Targets, 32](#)

Google BigQuery Run-Time Processing Overview

When you create a Google BigQuery data object read or write operation, you define properties that determine how the Data Integration Service reads data from or writes data to a Google BigQuery database.

You can configure lookup caching, partitioning, and parameterization in the run-time properties.

Filter Expression

To read specific data from a Google BigQuery table, you can configure a filter condition to query the Google BigQuery tables. You can use the Native or Platform expression to query specific columns in a Google BigQuery table.

Native Expression

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

An Google BigQuery filter consists of one or more Boolean expressions. The Boolean expressions uses the following format:

```
<Tablename.columnname><Operator><Value>
```

If you use logical operators, add the operators as a prefix to the expression list. Default is blank.

To filter records from an Google BigQuery source, set the native expression in the data object read operation.

Note: When you configure a Native expression and select a column that you want to filter, you must delete the `<Tablename.>` value in the native expression.

Platform Expression

You can use the platform expression to select specific records from a Google BigQuery table based on the filter condition you specify.

The following table describes the properties you specify when you filter records from a Google BigQuery table when you use the platform expression filter:

| Property | Description |
|-----------------|--|
| Expression Type | The type of filter expression that you want to use to filter records. |
| Left Field | The Google BigQuery column on which you want to apply the filter condition. You cannot specify a platform expression for columns of the following data types: <ul style="list-style-type: none">- Date- Time- Date/Time- Timestamp |
| Operator | Simple operators you can use to filter records. You can select one of the following operators: <code>=</code> , <code>!=</code> , <code><</code> , <code><=</code> , <code>></code> , and <code>>=</code> |
| Right Field | The value you specify to filter Google BigQuery columns. |

Partitioning

When you read data from or write data to Google BigQuery, you can configure partitioning to optimize the mapping performance at run time. You can configure partitioning for Google BigQuery mappings that you run in the native or Spark engine. The partition type controls how the Data Integration Service distributes data among partitions at partition points.

You can define the partition type as key range partitioning. To configure key range partitioning, open the Google BigQuery data object read operation, and select the **Key Range** partition type option on the **Run-time** tab.

When you configure key range partitioning, the Data Integration Service distributes rows of data based on a port or set of ports that you define as the partition key. You can define a range of values for each port. The Data Integration Service uses the key and ranges to send rows to the appropriate partition.

You can configure a partition key for fields of the following data types:

- Integer
- Float

Note: You cannot configure a partition key for Record data type columns and repeated columns.

You can configure dynamic partitioning for Google BigQuery data object write operation. To configure dynamic partitioning, open the Google BigQuery data object write operation, and select the **Dynamic** partition type option on the **Runtime** tab.

Parameterization for Google BigQuery Sources

You can parameterize the Google BigQuery connection and data object read operation properties to override the mapping properties at run time.

You can parameterize the following read operation properties for a Google BigQuery source:

- Source Dataset ID
- Source Table name
- Number of Rows to Read
- Query Results Table Name
- Job Poll Interval In Seconds
- Read Mode
- Number of Threads for Downloading Staging Files
- Data format of the staging file
- Local Stage File Directory
- Staging File Name
- pre SQL
- post SQL
- pre SQL Configuration
- post SQL Configuration

Parameterization for Google BigQuery Targets

You can parameterize the Google BigQuery connection and data object write operation properties to override the mapping properties at run time.

You can parameterize the following write operation properties for a Google BigQuery target:

- Target Dataset ID
- Target Table name
- Create Disposition
- Write Disposition
- Write Mode
- Streaming Template Table Suffix
- Rows per Streaming Request
- Staging File Name
- Number of Threads for Uploading Staging Files

- Data format of the staging file
- Job Poll Interval In Seconds
- Local Stage File Directory
- Staging File Name
- Field Delimiter
- pre SQL
- post SQL
- pre SQL Configuration
- post SQL Configuration

APPENDIX A

Google BigQuery Data Type Reference

This appendix includes the following topics:

- [Data Type Reference Overview, 34](#)
- [Google BigQuery and Transformation Data Types, 34](#)

Data Type Reference Overview

Developer Tool 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 Data Integration Service uses to move data across platforms. They appear in all transformations in a mapping.

When the Data Integration Service reads source data, it converts the native data types to the comparable transformation data types before transforming the data. When the Data 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 Developer Tool supports and the corresponding transformation data types:

| Google BigQuery Data Type | Transformation Data Type | Range and Description for the Transformation Data Type |
|---------------------------|--------------------------|--|
| BOOLEAN | String | 1 to 104,857,600 characters. |
| BYTE | Binary | 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 | String | 1 to 104,857,600 characters Precision 65535, scale 0 |
| STRING | String | 1 to 104,857,600 characters. Precision 255, scale 0 |
| 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

A

advanced properties
input [23](#)

C

client component
installation [10](#)
create
 data object operation
 create [26](#)
 Google BigQuery data object [26](#)
creating
 Google BigQuery connection [17](#)

D

data types [34](#)

G

general properties
input [23](#)
Google BigQuery
 data object properties [18](#)
 data object read operation [20](#)
 data object write operation [22](#)
 introduction [8](#)
Google BigQuery connection
 properties [16](#)
Google BigQuery connections
 connection modes [12](#)
 connection modes example [13](#)
 creating [17](#)
 overview [12](#)
Google BigQuery data object
 create [26](#)
 overview [18](#)
Google BigQuery data types [34](#)
Google BigQuery parameterization
 for sources [32](#)
 for targets [32](#)
Google BigQuery run-time processing
 parameterization [32](#)
 partitioning [31](#)

I

input properties [23](#)

installation
 client component [10](#)
installation on Linux
 server component [10](#)

N

native environment
mappings [28](#)

P

post SQL commands
 entering [19](#)
PowerExchange for Google BigQuery
 administration [8](#)
 configuration [9](#)
 example [7](#)
 installation [9](#)
 overview [7](#)
PowerExchange for Google BigQuery mappings
 overview [28](#)
pre SQL commands
 entering [19](#)
pre SQL configuration and post SQL configuration
 entering [19](#)

Q

query
 native expression [30](#)
 platform expression [30, 31](#)

S

server component
 installation on Linux [10](#)
Spark engine
 mappings [28](#)

T

transformation data types [34](#)