



Informatica® PowerExchange for Google
Cloud Storage

10.5.6

User Guide for PowerCenter

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

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.

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 Cloud Storage.....	 7
PowerExchange for Google Cloud Storage Overview.	7
Google Cloud Storage File Formats.	7
Introduction to Google Cloud Storage.	8
 Chapter 2: PowerExchange for Google Cloud Storage Configuration.....	 11
PowerExchange for Google Cloud Storage Configuration Overview.	11
Registering the PowerExchange for Google Cloud Storage Plug-in.	11
Registering the PowerExchange for Google Cloud Storage Plug-in from the Administrator Tool.	12
Registering the PowerExchange for Google Cloud Storage Plug-in from the Command Line Program.	12
Configuring Environment Variables.	12
 Chapter 3: Google Cloud Storage Sources and Targets.....	 14
Google Cloud Storage Sources and Targets Overview.	14
Import Google Cloud Storage Source and Target Definitions.	14
Reading File Names for Source Objects.	19
Rules and Guidelines for Using FileName Port.	20
Rules and Guidelines for Google Cloud Storage Sources and Targets.	20
 Chapter 4: Google Cloud Storage Mappings.....	 21
Google Cloud Storage Mappings Overview.	21
Google Cloud Storage Mapping Example.	21
 Chapter 5: Google Cloud Storage Sessions.....	 23
Google Cloud Storage Sessions Overview.	23
Google Cloud Storage Connections.	23
PowerExchange for Google Cloud Storage Connections.	24
Configuring a Google Cloud Storage Connection.	24

Directory Source in Google Cloud Storage Sources.	24
Configure Google Cloud Storage Source Session Properties.	25
Configure Google Cloud Storage Target Session Properties.	25
Java Heap Memory Configuration.	26
Appendix A: Google Cloud Storage Data Type Reference.	27
Data Type Reference Overview.	27
CSV Google Cloud Storage File Data Types and Transformation Data Types.	27
Avro Google Cloud Storage File Data Types and Transformation Data Types.	28
JSON Google Cloud Storage File Data Types and Transformation Data Types.	29
Parquet Google Cloud Storage File Data Types and Transformation Data Types.	30
Index.	31

Preface

Use the *Informatica® PowerExchange® for Google Cloud Storage User Guide for PowerCenter* to learn how to read from and write to Google Cloud Storage by using PowerCenter Client. Learn to create a Google Cloud Storage connection, develop mappings, and run sessions in an Informatica domain.

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 Cloud Storage

This chapter includes the following topics:

- [PowerExchange for Google Cloud Storage Overview, 7](#)
- [Google Cloud Storage File Formats, 7](#)
- [Introduction to Google Cloud Storage, 8](#)

PowerExchange for Google Cloud Storage Overview

You can use PowerExchange for Google Cloud Storage for connectivity between PowerCenter and Google Cloud Storage.

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

Example

You run the IT department of a major bank and are responsible for storing huge volumes of transaction files in a relational database. You want to store the data in another database to avoid data loss if the relational database fails.

You can use PowerExchange for Google Cloud Storage to upload huge volumes of transaction files to Google Cloud storage from any location and at any time. You can back up data in Google Cloud Storage for disaster recovery purposes and retrieve the data later, if needed.

You can configure a mapping in PowerCenter to write data to Google Cloud Storage.

Google Cloud Storage File Formats

PowerExchange for Google Cloud Storage can read data from and write data to Google Cloud Storage in the following formats:

- Comma-separated values (CSV)
- JSON (newline-delimited)

- Avro files
- Parquet

When you import a Google Cloud Storage source or target CSV file, you can edit the schema to change the data type and precision.

When you import a Google Cloud Storage source or target file in Avro, Parquet, or JSON format, you can edit the schema

Introduction to Google Cloud Storage

Google Cloud Storage is a web service that allows global storage and retrieval of large volumes of data at any time.

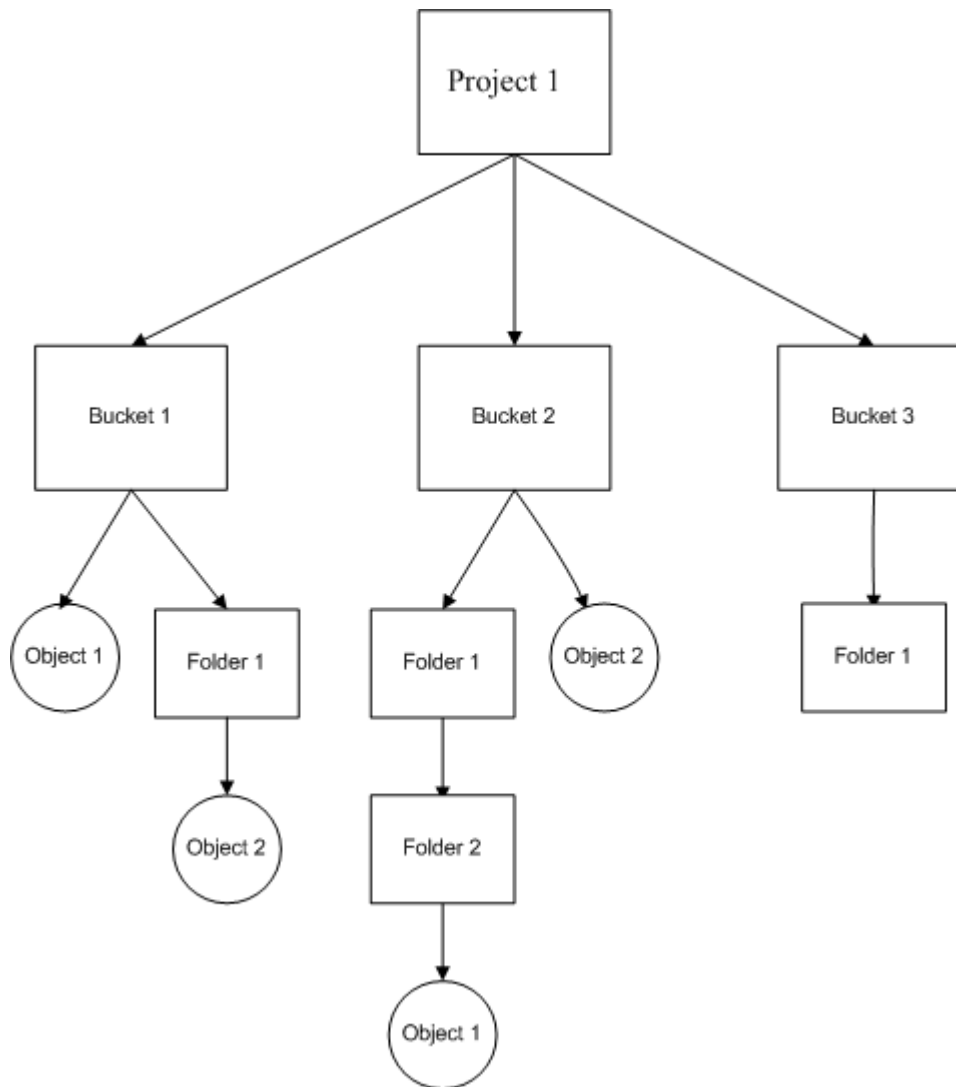
You can use Google Cloud Storage to stream multimedia, store custom data analytics pipelines, or distribute large data objects to users through direct download. You can write data to Google Cloud Storage for data backup. In the event of a database failure, you can read the data from Google Cloud Storage and restore it back to the database.

Google Cloud Storage offers different storage classes based on factors such as data availability, latency, and price.

Google Cloud Storage comprises the following components:

- Projects
- Buckets
- Objects

The following image shows how data can be organized in Google Cloud Storage:



You can use the following components to write to Google Cloud Storage:

Projects

In Google Cloud Storage, all resources are stored within a project. Project is a top-level container that stores billing details and user details. You can create multiple projects. A project has a unique project name, project ID, and project number.

Buckets

Each bucket acts like a container that stores data. You can use buckets to organize and access data. You can create more than one bucket but you cannot nest buckets.

You can create multiple folders within a bucket and you can also nest folders.

You can define access control lists to manage objects and buckets. An access control list consists of permission and scope entries. Permission defines the access to perform a read or write operation. Scope defines a user or a group who can perform the operation.

Objects

Objects comprise the data that you upload to Google Cloud Storage. You can create objects in a bucket. Objects consist of object data and object metadata components. The object data is a file that you store

in Google Cloud Storage. The object metadata is a collection of name-value pairs that describe object qualities.

CHAPTER 2

PowerExchange for Google Cloud Storage Configuration

This chapter includes the following topics:

- [PowerExchange for Google Cloud Storage Configuration Overview, 11](#)
- [Registering the PowerExchange for Google Cloud Storage Plug-in, 11](#)
- [Configuring Environment Variables, 12](#)

PowerExchange for Google Cloud Storage Configuration Overview

PowerExchange for Google Cloud Storage installs with Informatica services.

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

Registering the PowerExchange for Google Cloud Storage 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 Cloud Storage is `googlestorage_Plugin.xml`.

Registering the PowerExchange for Google Cloud Storage 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 Cloud Storage 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 `googlestorage_Plugin.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  
\googlestorage_Plugin.xml -e -N
```

Configuring Environment Variables

After you install PowerExchange for Google Cloud Storage, configure environment variables.

To successfully preview data from the Avro and Parquet files or run a mapping in PowerCenter with the Avro or Parquet files, you must configure the `INFA_HADOOP_DIST_DIR` environment variable for the PowerCenter Integration Service in Informatica Administrator. Perform the following steps to configure the `INFA_HADOOP_DIST_DIR` property:

1. Log in to Informatica Administrator.
2. Click the PowerCenter Integration Service and then click the **Processes** tab on the right pane.
3. Click **Edit** in the **Environment Variables** section.

4. Click **New** to add an environment variable.
5. Enter the name of the environment variable as INFA_HADOOP_DIST_DIR.
6. Set the value of the environment variable to the following absolute path of the Hadoop distribution directory on the machine that runs the PowerCenter Integration Service:

<Informatica installation directory/Informatica/services/shared/hadoop/<Hadoop distribution name>_<version>

For example, enter the following absolute path:

C:/Informatica/services/shared/hadoop/CDH_6.3

CHAPTER 3

Google Cloud Storage Sources and Targets

This chapter includes the following topics:

- [Google Cloud Storage Sources and Targets Overview, 14](#)
- [Import Google Cloud Storage Source and Target Definitions, 14](#)
- [Reading File Names for Source Objects, 19](#)
- [Rules and Guidelines for Google Cloud Storage Sources and Targets, 20](#)

Google Cloud Storage Sources and Targets Overview

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

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 Cloud Storage Source and Target Definitions

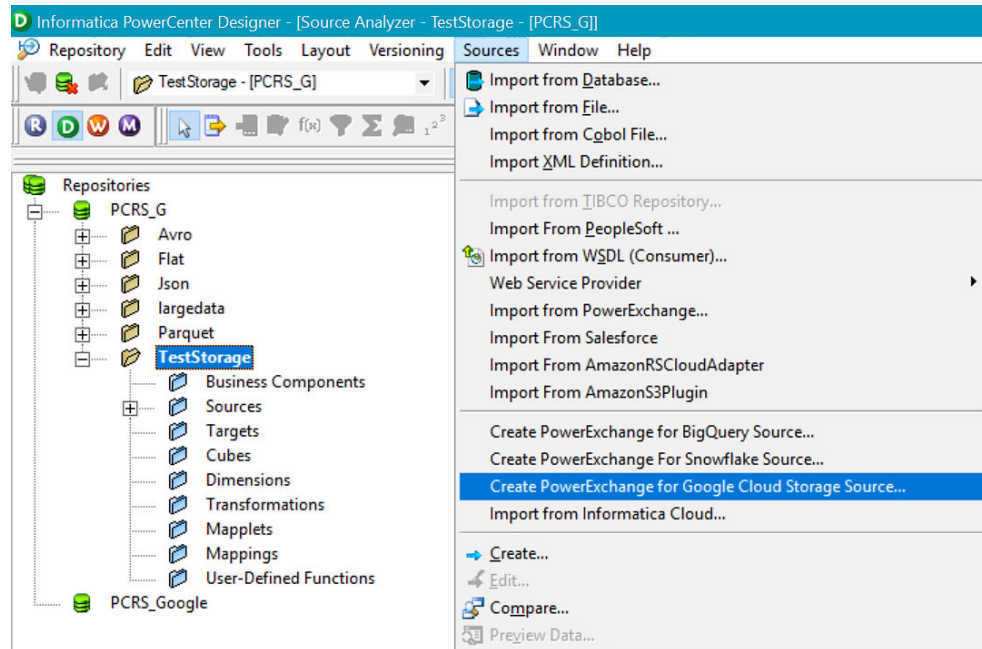
Use the **Create PowerExchange for Google Cloud Storage Source** or **Create PowerExchange for Google Cloud Storage Target** wizard to import Google Cloud Storage source or target definitions into the PowerCenter repository.

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

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

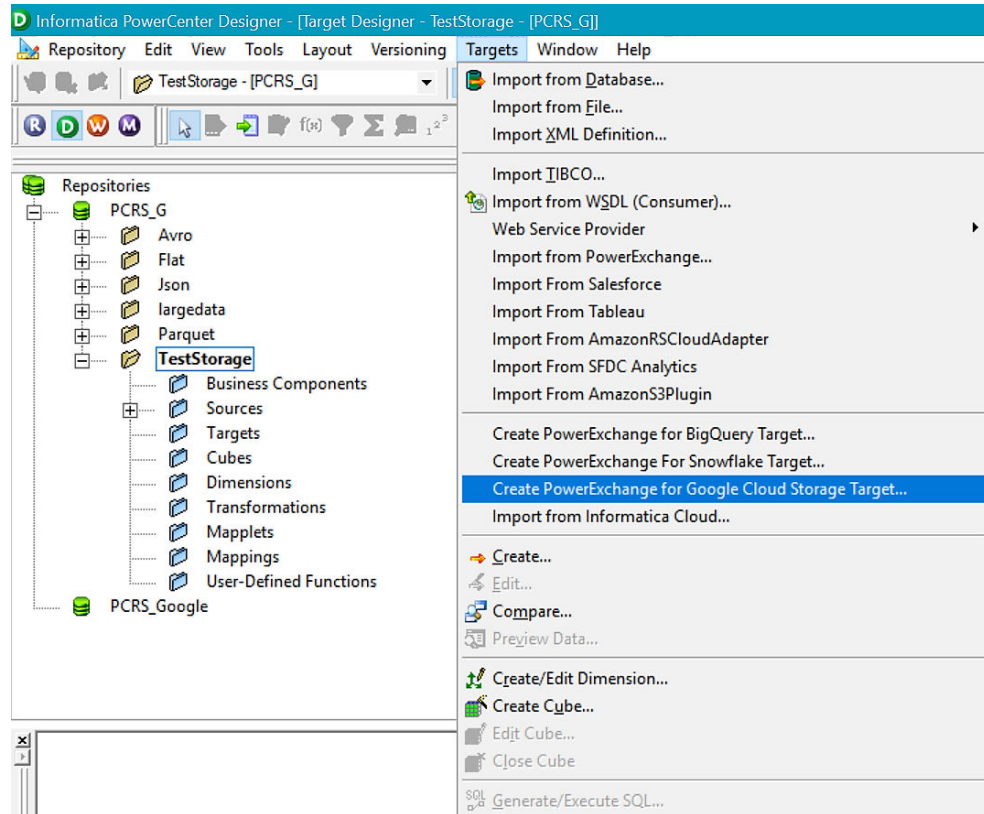
3. Select **Source Analyzer** or **Target Designer**.

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



The **Google Cloud Storage Connection** wizard appears.

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



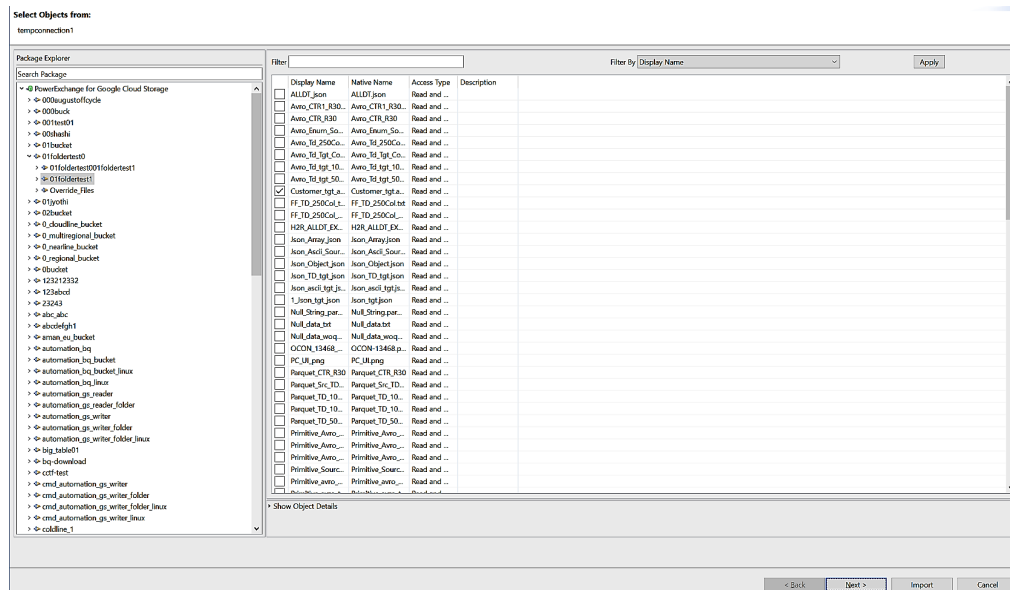
The **Google Cloud Storage Connection** dialog box appears.

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.
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.

5. Click **Test** to test the connection.
6. Click **Finish** to add the connection.
The **Select Objects from** dialog box appears.
7. Select the dataset in **Package Explorer**.

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

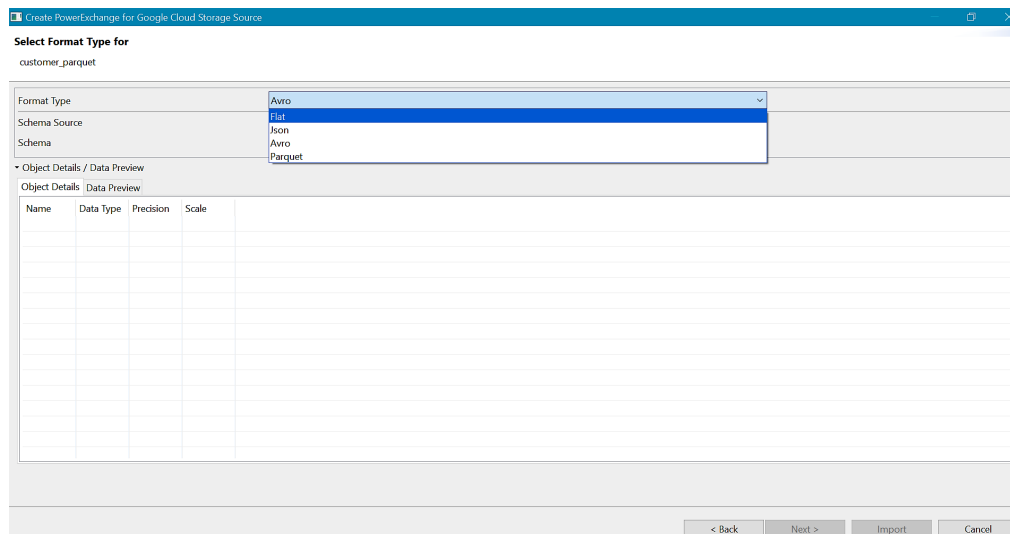


To view the object details, select the file, and double-click the file name.

8. Select the file that you want to import, and then click **Next**.

The **Select Format Type** for dialog box appears.

The following image shows the **Select Format Type** for dialog box:



9. In the **Format Type** list, select one of the following data format to import data from files in Google Cloud Storage:

- Flat. To read a flat resource.
- Avro. To read an Avro resource.
- Json. To read a JSON resource.
- Parquet. To read a Parquet resource.

10. If you select **Flat** as the **Format Type**, you must configure the following format properties:

Property	Description
Schema Source	<p>Specifies the mode to import schema for the Google Cloud Storage file.</p> <p>You can select one of the following options:</p> <ul style="list-style-type: none">- Import Schema from File. Import schema from a schema definition file in your local machine.- Read from Data File. PowerExchange for Google Cloud Storage imports the schema from the file in Google Cloud Storage.
Schema	<p>If you selected the Import Schema from File option in the Schema Source drop-down list, you must choose a schema definition file on your local machine.</p> <p>If you selected the Read from Data File option in the Schema Source drop-down list, you can view or edit the schema of the Google Cloud Storage file.</p>
Delimiter	<p>Character used to separate columns of data. If you enter a delimiter that is the same as the escape character or the text qualifier, you might receive unexpected results. Google Cloud Storage reader and writer support Delimiters.</p> <p>Default is comma.</p>
Escape Character	<p>Character immediately preceding a column delimiter character embedded in an unquoted string, or immediately preceding the quote character in a quoted string. When you specify an escape character, the PowerCenter Integration Service reads the delimiter character as a regular character.</p>
Text Qualifier	<p>Quote character that defines the boundaries of text strings. If you select a quote character, the PowerCenter Integration Service ignores delimiters within pairs of quotes. Google Cloud Storage reader supports Text Qualifier.</p>
Qualifier Mode	<p>Qualifier behavior for the source object.</p> <p>You can select one of the following options:</p> <ul style="list-style-type: none">- Minimal. Default mode. Applies qualifier to data that have a delimiter value or a special character present in the data. Otherwise, the Data Integration Service does not apply the qualifier.- All. Applies qualifier to all data.
Header Line Number	<p>Line number that you want to use as the header.</p> <p>You can also read a data from a file that does not have a header. To read data from a file with no header, specify the value of the Header Line Number field as 0.</p>
First Data Row	<p>Line number from where you want the Secure Agent to read data.</p>
Row Delimiter	<p>Specify a line break character. Select from the list or enter a character. Preface an octal code with a backslash (\). To use a single character, enter the character. The Data Integration Service uses only the first character when the entry is not preceded by a backslash. The character must be a single-byte character, and no other character in the code page can contain that byte. Default is line-feed, \012 LF (\n).</p>
Max Rows to Preview	<p>Specify the maximum number of rows that you want to display in the data preview.</p>

Property	Description
Row Delimiter	Specify a line break character. Select from the list or enter a character. Preface an octal code with a backslash (\). To use a single character, enter the character. The PowerCenter Integration Service uses only the first character when the entry is not preceded by a backslash. The character must be a single-byte character, and no other character in the code page can contain that byte. Default is line-feed, \012 LF (\n).
Target Header	Indicates whether you want to write data with or without a header. Note: This property appears only when you import a Google Cloud Storage target definition.
Distribution Column	Not applicable.

11. If you select **Json**, **Avro**, or **Parquet** as the **Format Type**, you must configure the following format properties:

Property	Description
Schema Source	Specifies the mode to import schema for the Google Cloud Storage file. You can select one of the following options: <ul style="list-style-type: none"> - Import Schema from File. Import schema from a schema definition file in your local machine. - Read from Data File. PowerExchange for Google Cloud Storage imports the schema from the file in Google Cloud Storage.
Schema	If you selected the Import Schema from File option in the Schema Source drop-down list, you must choose a schema definition file on your local machine. If you selected the Read from Data File option in the Schema Source drop-down list, you can view or edit the schema of the Google Cloud Storage file.

12. Click **Import**.

Note: If you click **Import** in the **Select Objects from** dialog box, PowerExchange for Google Cloud Storage imports the data from the Google Cloud Storage file in binary format.

To view the object details or preview data, expand the **Object Details/Data Preview** option.

Reading File Names for Source Objects

You can use the data in the FileName port when you import a Google Cloud Storage source.

When you import a Google Cloud Storage source object, the PowerCenter Integration Service creates a FileName port in the imported source transformation. The FileName port stores the name of the file from which the PowerCenter Integration Service reads the data at run-time. For example, a directory contains a number of files and each file contains multiple records that you want to read. You select the directory as source type in the Google Cloud Storage source session properties. When you run the session, the PowerCenter Integration Service reads each record and stores the respective file name in the FileName port.

When you run a session to read a Google Cloud Storage file using the FileName port, the PowerCenter Integration Service appends an underscore (_) to the directory name.

The FileName port is applicable to the following file formats:

- Flat file
- Avro (excluding hierarchical data types)
- Binary
- JSON (excluding hierarchical data types)
- Parquet (excluding hierarchical data types)

Rules and Guidelines for Using FileName Port

Use the following rules and guidelines when you run a session to read data using the FileName port:

- When you run a session to read a Google Cloud Storage file and if one of the values in the FileName port does not contain any value, the PowerCenter Integration Service creates the file in the following format:

```
<valueoftheNativeNamepropertyorFileNameDataObjectWriteOperation>_<fileextension>=<>
```

However, if you run a session to read the newly created file, the session fails with the following error message:

```
java.lang.AssertionError: assertion failed: Empty partition column value in '< >'
at scala.Predef$.assert(Predef.scala:170)
```

You must ensure that all entries in the FileName port contains a value to read the newly created file successfully.

- Do not use a colon (:) and forward slash (/) character in the file name data of the FileName port of the source object to run a session.
- Do not connect FileName port to a FileName port because the FileName port in the source might contain colon (:) and forward slash (/) characters.

Rules and Guidelines for Google Cloud Storage Sources and Targets

Use the following rules and guidelines when you import a Google Cloud Storage source or target:

- When you import a Google Cloud Storage source or target definition and select **Json** as the **Format Type**, you cannot import the JSON schema using the **Import Schema from File** option.
- When you import a Google Cloud Storage source or target definition and the file contains Unicode characters, the metadata import fails.
- When you import a Google Cloud Storage source or target definition, you can select only one bucket at a time in the **Package Explorer**.
- When you import a Google Cloud Storage flat file, by default comma is the delimiter. If you want to specify a delimiter other than comma, you must first change the value of the delimiter and then select a value for the **Schema** option.

For example, if you want to specify a semicolon (;) as a delimiter, you must first change the value of the delimiter to semicolon and then select **Read Data from File** in the **Schema** option.

CHAPTER 4

Google Cloud Storage Mappings

This chapter includes the following topics:

- [Google Cloud Storage Mappings Overview, 21](#)
- [Google Cloud Storage Mapping Example, 21](#)

Google Cloud Storage Mappings Overview

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

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

Google Cloud Storage Mapping Example

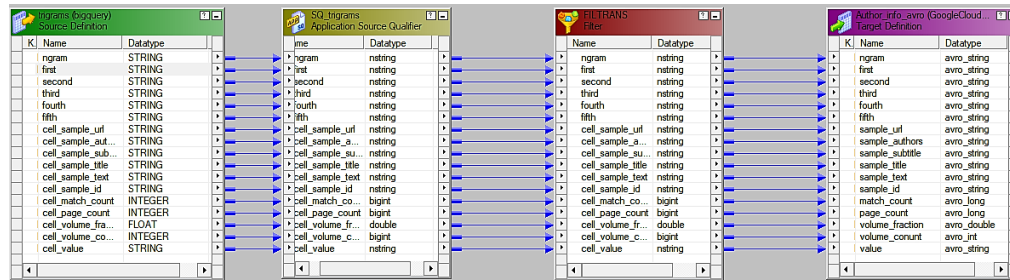
You run the IT department of a library and use Google BigQuery to store information about books in the library. You want to use the Google Cloud Storage to store information for particular authors.

You can extract author and book details from a Google BigQuery source and use PowerExchange for Google Cloud Storage to load the data to a target Avro file in Google Cloud Storage.

The following procedure shows how to move data from Google BigQuery to Google Cloud Storage:

1. Import the Google BigQuery source table.
2. Import the Google Cloud Storage target file.
3. Create a mapping with the Google BigQuery source and the Google Cloud Storage target.
4. Use a **Filter Transformation** to filter details for a particular author.

The following image shows the example mapping:



5. Create a session and configure it to load the data to the Google Cloud Storage target.

The mapping contains the following objects:

Google BigQuery Source Definition

The mapping source definition is a Google BigQuery source table. In the **Source Analyzer**, import the Google BigQuery `trigrams` source table. The PowerCenter Integration Service reads the author and book details from the Google BigQuery source.

Transformations

The `FILTRANS` filter transformation filters the details for a particular author.

Google Cloud Storage Target Definition

The mapping contains a Google Cloud Storage target definition.

In the **Target Designer**, import a Google Cloud Storage target definition.

CHAPTER 5

Google Cloud Storage Sessions

This chapter includes the following topics:

- [Google Cloud Storage Sessions Overview, 23](#)
- [Google Cloud Storage Connections, 23](#)
- [Directory Source in Google Cloud Storage Sources, 24](#)
- [Configure Google Cloud Storage Source Session Properties, 25](#)
- [Configure Google Cloud Storage Target Session Properties, 25](#)
- [Java Heap Memory Configuration, 26](#)

Google Cloud Storage Sessions Overview

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

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

Google Cloud Storage Connections

Create a Google Cloud Storage connection to read data from a Google Cloud Storage source and write data to a Google Cloud Storage target. You can use Google Cloud Storage connections in mappings. When you create a Google Cloud Storage connection, you can configure a connection mode based on how you want to read and write the data.

PowerExchange for Google Cloud Storage Connections

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

The following table describes the Google Cloud Storage 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.
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.

Configuring a Google Cloud Storage Connection

Configure a Google Cloud Storage connection in the **Workflow Manager** to define the connection attributes that the PowerCenter Integration Service uses to connect to the Google Cloud Storage 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 **Google Cloud Storage** and click **OK**.
The **Application Connection Editor** dialog box appears.
4. Enter a name for the Google Cloud Storage connection.
5. Enter the Google Cloud Storage connection attributes.
6. Click **OK** to create a Google Cloud Storage connection.

Directory Source in Google Cloud Storage Sources

You can select the **Is Directory** option under the Google Cloud Storage source session properties to read all the files in a Google Cloud Storage folder.

PowerExchange for Google Cloud Storage provides the option to override the value of the **Google Cloud Storage Path** and **Source File Name** properties during run time. When you select the **Is Directory** option, the value of the **Source File Name** is not honored.

When you run a session to read data from Google Cloud Storage and you provide the **Google Cloud Storage Path** value in the session properties, the PowerCenter Integration Service considers the value of the **Google Cloud Storage Path** from the session properties. If you do not provide the **Google Cloud Storage Path** value during run time, the PowerCenter Integration Service considers the value of the **Google Cloud Storage Path** that you specify in the source definition.

Note: When you select the **Is Directory** option, PowerExchange for Google Cloud Storage does not read the files available in the sub-folder.

Use the following rules and guidelines to select the **Is Directory** option:

- You cannot specify wildcard characters in the **Google Cloud Storage Path** source session property.
- All the source files in the directory must contain the same metadata.
- All the files must have data in the same format. For example, delimiters, header fields, and escape characters must be same.

Configure Google Cloud Storage Source Session Properties

You can configure the session properties for a Google Cloud Storage 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 Cloud Storage source session:

Property	Description
Google Cloud Storage Path	Optional. Overrides the bucket name or folder path of the Google Cloud Storage file that you selected in the Google Cloud Storage source definition. Use the following format: <code>gs://<bucket name></code> or <code>gs://<bucket name>/<folder name></code>
Source File Name	Optional. Overrides the Google Cloud Storage source file name specified in the Google Cloud Storage source definition.
Is Directory	Select this property to read all the files available in the folder specified in the Google Cloud Storage Path property.
Tracing Level	Amount of detail displayed in the session log for the transformation. You can choose Normal, Verbose Initialization, or Verbose Data. Default is Normal.
Output is Deterministic	Indicates whether the transformation generates consistent output data between session runs.

Configure Google Cloud Storage Target Session Properties

You can configure the session properties for a Google Cloud Storage 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 Cloud Storage target session:

Property	Description
Google Cloud Storage Path	Optional. Overrides the bucket name or folder path of the Google Cloud Storage file that you selected in the Google Cloud Storage target definition. This property is required when the source is not a flat file. Use the following format: <code>gs://<bucket name></code> or <code>gs://<bucket name>/<folder name></code>
Target File Name	Optional. Overrides the Google Cloud Storage target file name.

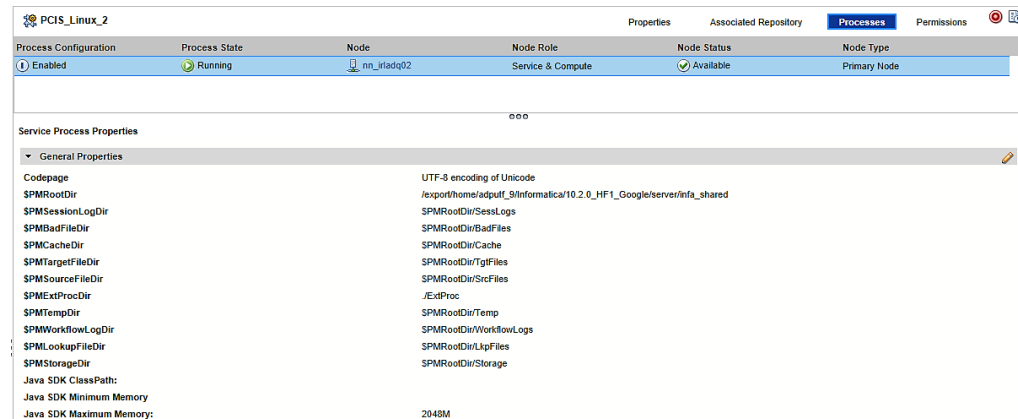
Java Heap Memory Configuration

When the Google Cloud Storage source or target contains large amount of data, configure the memory for the Java heap size in the node that runs the PowerCenter Integration Service. You must ensure that the minimum physical memory available on the server machine for each session is in the range of 300 to 500 MB.

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, edit the **General Properties** section.
3. Specify the maximum heap size in **Java SDK Maximum Memory** limit based on the amount data you want to process.

Specify the maximum and minimum heap size based on the data you want to process. The minimum physical memory available for each session must be in the range of 300 to 500 MB.

The following image shows the **Processes** tab where you can configure the **Java SDK Maximum Memory** for the PowerCenter Integration Service:



4. Click **Ok**.
5. Recycle the PowerCenter Integration Service.

APPENDIX A

Google Cloud Storage Data Type Reference

This appendix includes the following topics:

- [Data Type Reference Overview, 27](#)
- [CSV Google Cloud Storage File Data Types and Transformation Data Types, 27](#)
- [Avro Google Cloud Storage File Data Types and Transformation Data Types, 28](#)
- [JSON Google Cloud Storage File Data Types and Transformation Data Types, 29](#)
- [Parquet Google Cloud Storage File Data Types and Transformation Data Types, 30](#)

Data Type Reference Overview

PowerCenter uses the following data types in Google Cloud Storage mappings:

- Google Cloud Storage native data types. Google Cloud Storage data types appear in Google Cloud Storage 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.

CSV Google Cloud Storage File Data Types and Transformation Data Types

CSV Google Cloud Storage file data types map to transformation data types that the PowerCenter Integration Service uses to move data across platforms.

The following table lists the Google Cloud Storage data types that the PowerCenter Integration Service supports and the corresponding transformation data types:

Google Cloud Storage Data Type	Transformation Data Type	Description
BIGINT	Bigint	Precision of 19 digits, scale of 0
DATETIME	Date/Time	Jan 1, 0001 A.D. to Dec 31, 9999 A.D. Precision of 29, scale of 9 (precision to the nanosecond) Combined date/time value.
NSTRING	String	1 to 104,857,600 characters
NUMBER	Decimal	For transformations that support precision up to 28 digits, the precision is 1 to 28 digits, and the scale is 0 to 28. If you specify the precision greater than the maximum number of digits, the PowerCenter Integration Service converts decimal values to double in high precision mode.
STRING	String	1 to 104,857,600 characters
Timestamp with Time Zone	timestampWithTZ	Precision of 36, scale of 9

Avro Google Cloud Storage File Data Types and Transformation Data Types

Avro Google Cloud Storage file data types map to transformation data types that the PowerCenter Integration Service uses to move data across platforms.

The following table lists the Avro Google Cloud Storage file data types that the PowerCenter Integration Service supports and the corresponding transformation data types:

Avro Google Cloud Storage File Data Type	Transformation Data Type	Range and Description
BOOLEAN	Integer	TRUE (1) or FALSE (0)
BYTES	Binary	Precision 4000
DOUBLE	Double	Precision 15
FIXED	Binary	1 to 104,857,600 bytes
FLOAT	Double	Precision 15
INT	Integer	-2,147,483,648 to 2,147,483,647 Precision 10, scale 0

Avro Google Cloud Storage File Data Type	Transformation Data Type	Range and Description
LONG	Bigint	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision 19, scale 0
NULL	Integer	-2,147,483,648 to 2,147,483,647 Precision 10, scale 0
STRING	String	1 to 104,857,600 characters

JSON Google Cloud Storage File Data Types and Transformation Data Types

JSON Google Cloud Storage file data types map to transformation data types that the PowerCenter Integration Service uses to move data across platforms.

The following table lists the JSON Google Cloud Storage file data types that the PowerCenter Integration Service supports and the corresponding transformation data types:

JSON Google Cloud Storage File Data Type	Transformation Data Type	Range and Description
BIGINT	Bigint	Precision of 19 digits, scale of 0
BOOLEAN	Integer	TRUE (1) or FALSE (0)
DOUBLE	Double	Precision 15
INTEGER	Integer	-2,147,483,648 to 2,147,483,647 Precision of 10, scale of 0
STRING	String	1 to 104,857,600 characters

Note: PowerExchange for Google Cloud Storage does not support the following JSON data types:

- Object
- Array
- Date/Timestamp
- Enum
- Union

Parquet Google Cloud Storage File Data Types and Transformation Data Types

Google Cloud Storage file data types map to transformation data types that the PowerCenter Integration Service uses to move data across platforms.

The following table lists the Google Cloud Storage file data types that the PowerCenter Integration Service supports and the corresponding transformation data types:

Parquet Google Cloud Storage File Data Type	Transformation	Description
BINARY (UTF8)	String	1 to 104,857,600 characters
BOOLEAN	Integer	TRUE (1) or FALSE (0)
DOUBLE	Double	Precision of 15 digits
Int32	Integer	-2,147,483,648 to 2,147,483,647 Precision of 10, scale of 0
Int64	Bigint	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision of 19, scale of 0

The Parquet schema that you specify to read or write a Parquet file must be in smaller case. Parquet does not support case-sensitive schema.

INDEX

A

Avro Google Cloud Storage file data types
transformation data types [28](#)

C

configuration
overview [11](#)

D

directory source
Google Cloud Storage sources [24](#)

F

filter [21](#)

G

Google Cloud Storage
access control lists [8](#)
components [8](#)
connections [23](#)
features [8](#)
introduction [8](#)
mappings [21](#)
source [14](#)
target [14](#)
Google Cloud Storage components
buckets [8](#)
objects [8](#)
projects [8](#)
Google Cloud Storage connections
properties [24](#)
connection configuration [24](#)
Google Cloud Storage files
data format [7](#)

I

Import
connection properties [14](#)
source definition [14](#)
target definition [14](#)

J

java heap size [26](#)
JSON Google Cloud Storage file data types
transformation data types [29](#)

M

mapping
example [21](#)
filter [21](#)
source [21](#)
target [21](#)

P

Parquet Google Cloud Storage file data types
transformation data types [30](#)
plug-in registration
administrator tool [12](#)
PowerExchange for Google Cloud Storage
configuration [11](#)
EBF [7](#)
example [7](#)
overview [7](#)
plug-in registration [11](#)

R

rules and guidelines
FileName port [20](#)

S

session
overview [23](#)
source properties [25](#)
target properties [25](#)

W

working with FileName port [19](#)