



Informatica® Data Integration - Free & PayGo

Google Cloud Storage V2 Connector

© Copyright Informatica LLC 2018, 2023

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

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

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

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

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

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

NOTICES

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

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

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

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

Publication Date: 2023-04-04

Table of Contents

Preface	5
Informatica Resources.	5
Informatica Documentation.	5
Informatica Intelligent Cloud Services web site.	5
Informatica Intelligent Cloud Services Communities.	5
Informatica Intelligent Cloud Services Marketplace.	5
Data Integration connector documentation.	6
Informatica Knowledge Base.	6
Informatica Intelligent Cloud Services Trust Center.	6
Informatica Global Customer Support.	6
 Chapter 1: Introduction to Google Cloud Storage V2 Connector.....	7
Google Cloud Storage V2 Connector assets.	7
Administration of Google Cloud Storage V2 Connector.	7
Introduction to Google Cloud Storage.	8
 Chapter 2: Google Cloud Storage V2 connections.....	11
Google Cloud Storage V2 connection properties.	11
Configure proxy settings.	12
Configure proxy settings for NTLM authentication.	12
 Chapter 3: Mappings for Google Cloud Storage.....	13
Directory source in Google Cloud Storage sources.	13
Pushdown optimization.	13
Informatica encryption for Google Cloud Storage V2 sources.	14
Importing encrypted source files.	14
Fixed-width file formats.	15
Handling dynamic schemas.	15
Google Cloud Storage V2 sources in mappings.	15
Google Cloud Storage file formatting options.	18
Data compression in Google Cloud Storage V2 sources.	19
Rules and guidelines for mappings and mapping tasks.	20
 Appendix A: Data type reference.....	22
Flat Google Cloud Storage file data types and transformation data types.	22
Avro Google Cloud Storage file data types and transformation data types.	23
JSON Google Cloud Storage file data types and transformation data types.	24
ORC Google Cloud Storage file data types and transformation data types.	24
Parquet Google Cloud Storage file data types and transformation data types.	25

Index..... 27

Preface

Use *Google Cloud Storage V2 Connector* to learn how to read from Google Cloud Storage by using Data Integration. Learn to create a connection, develop and run mappings, mapping tasks, and data transfer tasks in Data Integration.

Informatica Resources

Informatica provides you with a range of product resources through the Informatica Network and other online portals. Use the resources to get the most from your Informatica products and solutions and to learn from other Informatica users and subject matter experts.

Informatica Documentation

Use the Informatica Documentation Portal to explore an extensive library of documentation for current and recent product releases. To explore the Documentation Portal, visit <https://docs.informatica.com>.

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

Informatica Intelligent Cloud Services web site

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

Informatica Intelligent Cloud Services Communities

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

Access the Informatica Intelligent Cloud Services Community at:

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

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

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

Informatica Intelligent Cloud Services Marketplace

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

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

Data Integration connector documentation

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

Informatica Knowledge Base

Use the Informatica Knowledge Base to find product resources such as how-to articles, best practices, video tutorials, and answers to frequently asked questions.

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

Informatica Intelligent Cloud Services Trust Center

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

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

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

To subscribe, go to <https://status.informatica.com/> and click **SUBSCRIBE TO UPDATES**. You can then choose to receive notifications sent as emails, SMS text messages, webhooks, RSS feeds, or any combination of the four.

Informatica Global Customer Support

You can contact a Customer Support Center by telephone or online.

For online support, click **Submit Support Request** in Informatica Intelligent Cloud Services. You can also use Online Support to log a case. Online Support requires a login. You can request a login at <https://network.informatica.com/welcome>.

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

CHAPTER 1

Introduction to Google Cloud Storage V2 Connector

You can use Google Cloud Storage V2 Connector to securely read data from Google Cloud Storage.

You can use Google Cloud Storage V2 objects as sources in mappings and mapping tasks.

Use Google Cloud Storage V2 Connector to read from Avro, flat, and Parquet file formats for mappings. You can read from primitive data types for Avro, Parquet, JSON, and ORC files.

When you run a task or mapping, the Secure Agent uses the Google Cloud Storage API to perform the specified operation and reads data from the Google Cloud Storage files.

Google Cloud Storage V2 Connector assets

Create assets in Data Integration to integrate data using Google Cloud Storage V2 Connector.

When you use Google Cloud Storage V2 Connector, you can include the following Data Integration assets:

- Data transfer task
- Mapping
- Mapping task

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

Administration of Google Cloud Storage V2 Connector

Before you use Google Cloud Storage V2 Connector, you must complete the following prerequisite tasks:

1. Ensure that you have a Google service account to access Google Cloud Storage.
2. Ensure that 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 Cloud Storage connection in Data Integration.
3. Ensure that you have enabled the Google Cloud Storage JSON API for your service account.

Google Cloud Storage V2 Connector uses the Google API to integrate with Google Cloud Storage.

4. Verify that you have read access to the Google Cloud Storage bucket that contains the source file.
5. When you read data from a Google Cloud Storage file in a mapping, you must have the required permissions to run the mapping successfully.

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.

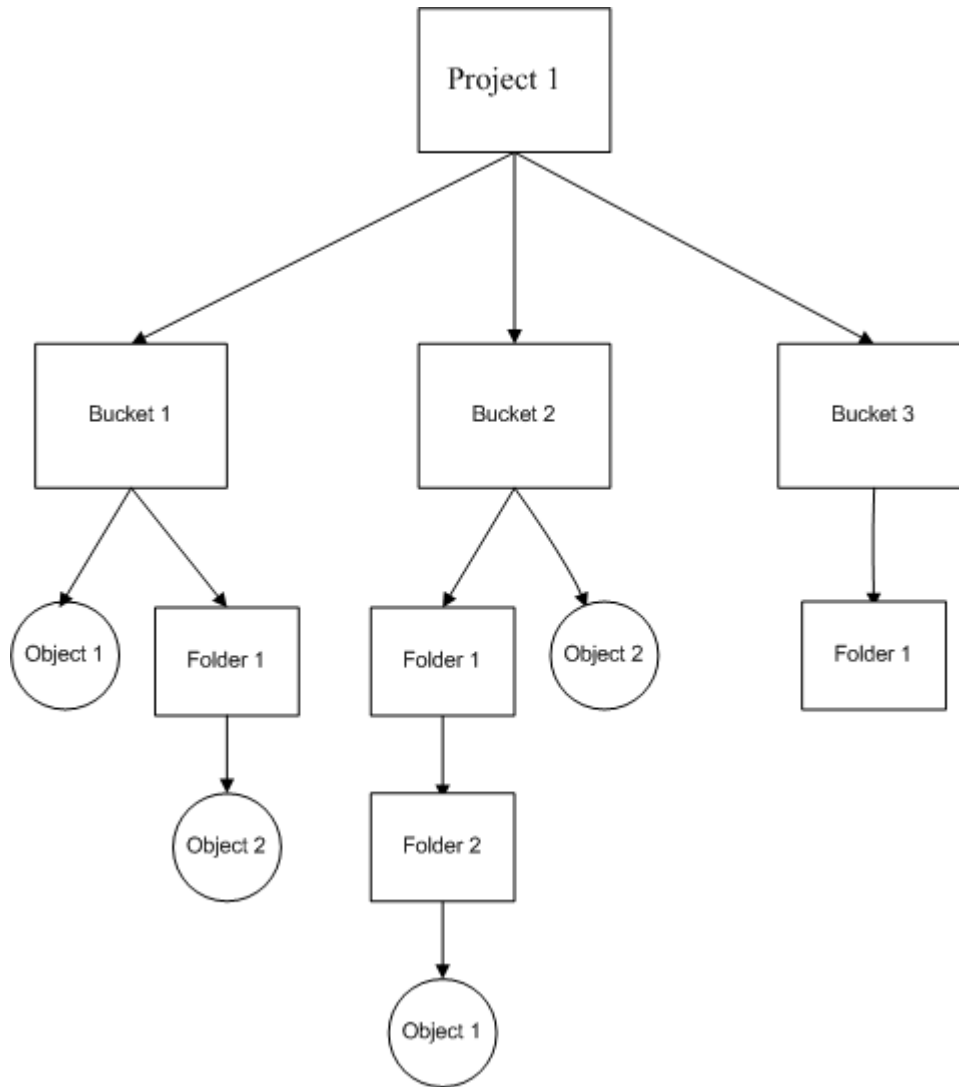
You can use Google Cloud Storage to stream multimedia, store custom data analytics pipelines, or distribute large data objects to users through direct download.

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 read data from 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 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

Google Cloud Storage V2 connections

Create a Google Cloud Storage V2 connection to securely read data from Google Cloud Storage files.

You can use a Google Cloud Storage V2 connection to specify sources in mappings and mapping tasks.

Google Cloud Storage V2 connection properties

When you create a Google Cloud Storage V2 connection, configure the connection properties.

The following table describes the Google Cloud Storage connection properties:

Property	Description
Connection Name	Name of the connection. Each connection name must be unique within the organization. Connection names can contain alphanumeric characters, spaces, and the following special characters: _ . + -, Maximum length is 255 characters.
Description	Description of the connection. Maximum length is 4000 characters.
Type	The Google Cloud Storage V2 connection type.
Runtime Environment	Name of the runtime environment where you want to run the tasks. Specify a Secure Agent or a Hosted Agent.
Service Account ID	The client_email value in the JSON file that you download after you create a service account.
Service Account Key	The private_key value in the JSON file that you download after you create a service account.

Property	Description
Project ID	The project_id value in the JSON file that you download after you create a service account. If you created multiple projects with the same service account, enter the ID of the project that contains the bucket that you want to connect to.
Bucket Name	The Google Cloud Storage bucket name that you want to connect to. When you select a source object , the Package Explorer lists files and folder available in the specified Google Cloud Storage bucket. If you do not specify a bucket name, you can select a bucket from the Package Explorer to select a source.

Configure proxy settings

If your organization uses an outgoing proxy server to connect to the internet, the agent connects to Informatica Intelligent Cloud Services through the proxy server.

You can configure the Secure Agent to use the proxy server on Windows and Linux. You can use the unauthenticated or authenticated proxy server.

Use one of the following methods to configure the proxy settings:

- Configure the Secure Agent through the Secure Agent Manager on Windows or shell command on Linux. For instructions, see the topic "Configure the proxy settings on Windows" or "Configure the proxy settings on Linux," in *Getting Started* in the Data Integration documentation.
- Configure the JVM options for the DTM in the Secure Agent properties. For instructions, see the Knowledge Base article [000185646](#).

Contact your network administrator for the correct proxy settings.

Configure proxy settings for NTLM authentication

You can use a proxy server that uses NTLM authentication to connect to Google Cloud Storage.

To configure the proxy settings for NTLM authentication, perform the following steps:

1. In Administrator, select **Runtime Environments**.
2. Select the Secure Agent for which you want to configure from the list of available Secure Agents.
3. In the upper-right corner, click **Edit**.
4. In the **System Configuration Details** section, select the **Type** as **DTM** for the Data Integration Server.
5. Edit the **JVMOption1** and add the following value:
`-Dhttp.auth.ntlm.domain=<domain name>`
6. Select the **Type** as **Platform** for the Data Integration Server.
7. Edit the **INFA_DEBUG** property and add the following value:
`-Dhttp.auth.ntlm.domain=<domain name>`
8. Click **Save**.
9. Restart the Secure Agent.

CHAPTER 3

Mappings for Google Cloud Storage

When you configure a mapping, you describe the flow of data from the source to the target.

A mapping defines reusable data flow logic that you can use in mapping tasks.

When you create a mapping, you define the Source and Target transformations to represent a Google Cloud Storage V2 object. Use the Mapping Designer in Data Integration to add the Source transformations in the mapping canvas and configure the Google Cloud Storage V2 source properties.

You can use Monitor to monitor the jobs.

Directory source in Google Cloud Storage sources

You can select the **Is Directory** option under the advanced properties for a Google Cloud Storage source object to read all the files in a Google Cloud Storage folder and sub-folder.

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

- You cannot read files available in a sub-folder within a sub-folder.
- All the source files in the directory must contain the same schema.
- When you read files in a Google Cloud Storage folder in Delimited format, all the files must have data in the same format. For example, delimiters, header fields, and escape characters must be same.
- When you run a mapping to read data from the files in a Google Cloud Storage folder and sub-folder using the Is Directory source advanced property, ensure that the folder and the sub-folder does not contain a file with the same schema. Otherwise, the mapping fails.
- All the source files in the directory must contain the same metadata and must have the data in the same format. The file format, delimiters, header fields, escape characters, and compression format must be same for all the files. If the guideline is not followed, the mapping might not process any of the files or process the files until the mapping files.

Pushdown optimization

You can enable full pushdown optimization when you want to load data from Google Cloud Storage sources to your data warehouse in Google BigQuery. While loading the data to Google BigQuery, you can transform

the data as per your data warehouse model and requirements. When you enable full pushdown on a mapping task, the mapping logic is pushed to the GCP environment to leverage GCP commands. For more information, see the help for Google BigQuery V2 Connector.

If your use case involves loading data to any other supported cloud data warehouse, see the connector help for the applicable cloud data warehouse.

Informatica encryption for Google Cloud Storage V2 sources

You can download a flat source file that is encrypted using the Informatica crypto libraries in the local machine or staging location and decrypt the source files.

Informatica encryption is applicable only when you run mappings on the Secure Agent machine installed on the Google Cloud Platform virtual machine.

When you configure a mapping, you can enable Informatica encryption for delimited flat files, fixed-width files, and binary files.

To read a source file that is encrypted using the Informatica crypto libraries, perform the following tasks:

1. Ensure that the organization administrator has permission to the Informatica crypto libraries license when you create a Google Cloud Storage V2 connection.
2. Select **Informatica Encryption** as the encryption type in the advanced source properties.
3. Ensure that you encrypt or decrypt the file within the same organization ID.

Importing encrypted source files

When you read an Informatica encrypted source file and select the **Informatica Encryption** as the encryption type, Secure Agent fails to import the encrypted file.

To import the data successfully, perform the following tasks:

1. Select the **Is Encrypted File** option in the Google Cloud Storage V2 connection.
2. Import an encrypted file from Google Cloud Storage and select **Delimited** as the **Format**.
3. Select **Import from Schema File** as the **Schema Source** in the formatting options.
4. Upload a schema file in the JSON format in the **Schema File** property to override the schema of the encrypted file.
5. Select **Informatica Encryption** as the encryption type in the advanced source properties.

Alternatively, you can select a dummy source file that contains the same metadata as in the Informatica encrypted source file from where you want to read data. To override the file name of the dummy source file, enter the file name of the Informatica encrypted source file in the **Source File Name** advanced source property. Then, select **Informatica Encryption** as the encryption type in the advanced source properties.

Fixed-width file formats

You can use a fixed-width flat file as a source in mappings and mapping tasks.

When you configure a Source transformation and select the fixed-width flat file type, you must select the most appropriate fixed-width file format to use based on the data in the fixed-width flat file.

See the following exceptions before you use a fixed-width flat file:

- When you use a fixed-width flat file as a source, you cannot edit the metadata for the fields.
- When you use a Secure Agent installed on a Linux machine and create a fixed-width file format, ensure that the sample file uses the `\n` character as the new line symbol, and that source files use the same symbol.
- When you use a Secure Agent installed on a Windows machine and create a fixed-width file format, ensure that the sample file uses the `\r\n` character as the new line symbol, and that source files use the same symbol.
- When you create a fixed-width file format, ensure that the sample flat file only uses UTF-8 character set encoding.

Handling dynamic schemas

When you add a mapping to a mapping task, you can choose how Data Integration handles changes in the data object schemas. To refresh the schema every time the task runs, you can enable dynamic schema handling in the task.

Configure schema change handling on the **Schedule** page when you configure the task. You can configure asynchronous or dynamic schema change handling.

Google Cloud Storage V2 sources in mappings

To read data from a Google Cloud Storage file, configure a Google Cloud Storage object as the Source transformation in a mapping.

Specify the name and description of the Google Cloud Storage source. Configure the source and advanced properties for the source object.

The following table describes the properties that you can configure for a Google Cloud Storage source:

Property	Description
Connection	Name of the Google Cloud Storage V2 connection. Select a source connection, or click New Parameter to define a new parameter for the source connection. If you want to overwrite the parameter at runtime, select the Allow parameter to be overridden at run time option when you create a parameter.
Source Type	Type of the Google Cloud Storage source object available. You can read data to a single Google Cloud Storage source object.

Property	Description
Object	Name of the source object for the mapping. Note: Ensure that the column names in the source object do not begin with special characters to successfully read data from the source.
Parameter	A parameter file where you define values that you want to update without having to edit the task. Select a parameter for the source object, or click New Parameter to define a new parameter for the source object. The Parameter property appears only if you select Parameter as the source type. When you define the value for the source object parameter, ensure that you provide a valid file name. If you want to overwrite the parameter at runtime, select the Allow parameter to be overridden at run time option when you create a parameter. When the task runs, the agent uses the parameters from the file that you specify in the task advanced session properties.
Format	Specifies the file format that the Google Cloud Storage V2 Connector uses to read data from Google Cloud Storage. You can select the following file format types: <ul style="list-style-type: none"> - Flat - Avro - Parquet - JSON - Orc - None Note: If you select None as the format type, Google Cloud Storage V2 Connector reads data from Google Cloud Storage files in binary format. Open the Formatting Options dialog box to configure the formatting options for the file. For more information about format options, see “Google Cloud Storage file formatting options” on page 18 .
Filter	This attribute is not applicable for Google Cloud Storage V2 Connector.
Sort	This attribute is not applicable for Google Cloud Storage V2 Connector.

The following table describes the advanced properties that you can configure for a Google Cloud Storage source:

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 object. Use the following format: <code>gs://<bucket name></code> or <code>gs://<bucket name>/<folder name></code> Note: You cannot specify wildcard characters in the Google Cloud Storage path.
Source File Name	Optional. Overrides the Google Cloud Storage source file name that you specified in the Source transformation. Note: Does not apply when you configure Is Directory option to read multiple files from a directory.
Is Directory	Select this property to read all the files available in the folder specified in the Google Cloud Storage Path property. Note: If you do not provide the Google Cloud Storage Path value during run time, the Secure Agent considers the value of the Google Cloud Storage Path that you specify when you select a Google Cloud Storage source file in the Source transformation.

Property	Description
Allow Wildcard Characters	Indicates whether you want to use wildcard characters for the directory sources. If you select this option, you can use the question mark (?) and asterisk (*) wildcard characters in the folder path or file name.
Encryption Type	Method to decrypt data. You can select one of the following encryption types: <ul style="list-style-type: none"> - Informatica Encryption - None Default is None .
Compression Format	Method to read compressed data from Google Cloud Storage. You can read the compressed data in the following formats: <ul style="list-style-type: none"> - None - Gzip Select None to read from the compressed Avro or Parquet file. Select Gzip to read from the compressed Flat file. Default is None .

You can set the tracing level in the advanced properties session to determine the amount of details that logs contain.

The following table describes the tracing levels that you can configure:

Tracing Level	Description
Terse	The Secure Agent logs initialization information, error messages, and notification of rejected data.
Normal	The Secure Agent logs initialization and status information, errors encountered, and skipped rows due to transformation row errors. Summarizes session results, but not at the level of individual rows.
Verbose Initialization	In addition to normal tracing, the Secure Agent logs additional initialization details, names of index and data files used, and detailed transformation statistics.
Verbose Data	In addition to verbose initialization tracing, the Secure Agent logs each row that passes into the mapping. Also notes where the Secure Agent truncates string data to fit the precision of a column and provides detailed transformation statistics.

Google Cloud Storage file formatting options

When you select the format of a Google Cloud Storage file, you can configure the formatting options.

The following table describes the Google Cloud Storage V2 file formatting options that you can configure:

Option	Description
Schema Source	You must specify the schema of the source file. You can select one of the following options to specify a schema: <ul style="list-style-type: none">- Read from data file. Google Cloud Storage V2 Connector imports the schema from the file in Google Cloud Storage.- Import from Schema File. Imports schema from a schema definition file in your local machine.
Read from data file	Not applicable.
Schema File	You can upload a schema file.

The following table describes the file formatting options that you must configure if you select Delimited as the Flat File Type:

Option	Description
Delimiter	Character used to separate columns of data. You can configure parameters such as comma, tab, colon, semicolon, or others. To set a tab as a delimiter, you must type the tab character in any text editor. Then, copy and paste the tab character in the Delimiter field. If you specify a multibyte character as a delimiter in the source object, the mapping fails. Note: To set a tab as a delimiter, you must type the tab character in any text editor. Then, copy and paste the tab character in the Delimiter field.
Escape Char	Character immediately preceding a column delimiter character embedded in an unquoted string, or immediately preceding the quote character in a quoted string.
Qualifier	Quote character that defines the boundaries of data. You can set qualifier as single quote or double quote.
Qualifier Mode	Specify the qualifier behavior for the object. You can select one of the following options: <ul style="list-style-type: none">- Minimal. Default mode. Applies qualifier to data enclosed within a delimiter value or a special character.- All. Applies qualifier to all data.- Non_Numeric. Not applicable.- All_Non_Null. Not applicable.

Option	Description
Code Page	<p>Select the code page that the Secure Agent must use to read data.</p> <p>You can select from the following code pages:</p> <ul style="list-style-type: none"> - MS Windows Latin 1. Select for ISO 8859-1 Western European data. - UTF-8. Select for Unicode and non-Unicode data. - Shift-JIS. Select for double-byte character data. - ISO 8859-15 Latin 9 (Western European). - ISO 8859-2 Eastern European. - ISO 8859-3 Southeast European. - ISO 8859-5 Cyrillic. - ISO 8859-9 Latin 5 (Turkish). - IBM EBCDIC International Latin-1. <p>This property applies only to mappings.</p>
Header Line Number	<p>Specify the line number that you want to use as the header when you read data from Google Cloud Storage. You can also read a file that doesn't have a header. Default is 1.</p> <p>To read data from a file with no header, specify the value of the Header Line Number field as 0. To read data from a file with a header, set the value of the Header Line Number field to a value that is greater than or equal to one.</p> <p>Ensure that the value of the Header Line Number field is lesser than or equal to the value of the First Data Row field.</p> <p>This property is applicable during runtime and data preview to read a file.</p>
First Data Row	<p>Specify the line number from where you want the Secure Agent to read data. You must enter a value that is greater or equal to one.</p> <p>To read data from the header, the value of the Header Line Number and the First Data Row fields should be the same. Default is 1.</p> <p>This property is applicable during runtime and data preview to read a file.</p> <p>This property applies only to mappings.</p>
Is Escape Character Data Retained	Not applicable.
Max Rows To Preview	Not applicable.
Row Delimiter	Not applicable.

To use a fixed-width flat file type, select the **Fixed Width File Format** to use. If you do not have a fixed-width file format, click **New > Components > Fixed Width File Format** to create one.

Data compression in Google Cloud Storage V2 sources

You can decompress the data when you read data from Google Cloud Storage V2 .

Configure the compression format in the **Compression Format** option under the advanced source properties.

The following table lists the supported compression formats for different file formats:

Compression format	Avro File	Binary File	Flat File	JSON File	Parquet File
Deflate	Yes	No	No	No	No
Gzip	No	No	Yes	No	Yes
None	Yes	Yes	Yes	Yes	Yes
Snappy	Yes	No	No	No	Yes

To read a compressed file from Google Cloud Storage V2, the compressed file must have specific extensions. If the extensions used to read the compressed file are not valid, the Secure Agent does not process the file. The following table describes the extensions that are appended based on the compression format that you use:

Compression format	File Name Extension
Deflate	.deflate
Gzip	.GZ
Snappy	.snappy

Use the following guidelines when you configure data compression:

- Data compression is supported at the file level. You cannot use data compression for a directory.
- When **Is Directory** property is selected at source, the files within the directory are read sequentially.
- When you download a Gzip compressed file for the Google Cloud Platform console, you must remove the content encoding metadata of the object manually. Select **Edit object metadata** of the object and remove **Gzip** from the **Content-Encoding** field.

Rules and guidelines for mappings and mapping tasks

Use the following guidelines when you create mappings:

- You must set the appropriate formatting options when you select the Flat, Avro, JSON, ORC, or Parquet format types.
- When you override the source file using the **Source File Name** advanced property, ensure that the file contains valid header columns.
- If you select the delimited format type and select Import from schema file as the value of the Schema Source formatting option, you can only upload a schema file in the JSON format.
The following sample shows a schema file for a delimited file:

```
{ "Columns": [{"Name": "f_varchar", "Type": "string", "Precision": "256", "Scale": "0"},  
{"Name": "f_char", "Type": "string", "Precision": "256", "Scale": "0"},  
{"Name": "f_integer", "Type": "string", "Precision": "256", "Scale": "0"}]}
```

- If you select the Avro, JSON, or Parquet format type and select **Read from data file** as the value of the **Schema Source** formatting option, you cannot view the schema of the file.
- When you select a file of JSON format that contains a large amount of data, you cannot preview data.
- When you read data from a flat file that contains a newline character, you must enclose the data within double quotation marks. Otherwise, the Secure Agent displays incorrect number of success rows in the session log.
- If you select None format type and select **Is Directory** advanced property, the connector reads only one file.

APPENDIX A

Data type reference

Data Integration uses the following data types in mappings and mapping tasks with Google Cloud Storage:

Google Cloud Storage native data types

Google Cloud Storage data types appear in the **Fields** tab for Source transformations when you choose to edit metadata for the fields.

Transformation data types

Set of data types that appear in the remaining transformations. They are internal data types based on ANSI SQL-92 generic data types, which Data Integration uses to move data across platforms. Transformation data types appear in all remaining transformations in a mapping or mapping task.

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

Flat Google Cloud Storage file data types and transformation data types

Flat Google Cloud Storage file data types map to transformation data types that the Secure Agent uses to move data across platforms.

The following table lists the Google Cloud Storage data types that the Secure Agent supports and the corresponding transformation data types:

Flat Google Cloud Storage Data Type	Transformation Data Type	Description
BigInt	BigInt	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision of 19, scale of 0
Number	Decimal	Precision 1 to 28 digits, scale 0 to 28
Nstring	Text	1 to 104,857,600 characters Fixed-length or varying-length string.
STRING	String	1 to 104,857,600 characters Default precision is 256. You can increase the value up to 104857600 characters.

Note: BigInt, Number, and Nstring data types are applicable only when you select **Import from Schema File** as the **Schema Source** in the file formatting options.

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 Secure Agent uses to move data across platforms.

The following table lists the Avro Google Cloud Storage file data types that the Secure Agent 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
FLOAT	Double	Precision 15
INT	Integer	-2,147,483,648 to 2,147,483,647 Precision 10, scale 0
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 Default precision is 256. You can increase the value up to 104857600 characters.

Note: Google Cloud Storage V2 Connector does not support the following Avro complex data types:

- Enum
- Fixed

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 Secure Agent uses to move data across platforms.

The following table lists the JSON Google Cloud Storage file data types that the Secure Agent 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 Default precision is 256. You can increase the value up to 104857600 characters.

Note: The following JSON complex data types are not applicable for Google Cloud Storage V2 Connector:

- Date/Timestamp
- Enum
- Union

ORC Google Cloud Storage file data types and transformation data types

ORC Google Cloud Storage file data types map to transformation data types that the Secure Agent uses to move data across platforms.

The following table lists the ORC Google Cloud Storage file data types that the Secure Agent supports and the corresponding transformation data types:

ORC Google Cloud Storage File Data Type	Transformation Data Type	Range and Description
BigInt	BigInt	-9223372036854775808 to 9,223,372,036,854,775,807
Boolean	Integer	TRUE (1) or FALSE (0)

ORC Google Cloud Storage File Data Type	Transformation Data Type	Range and Description
Char	String	1 to 104,857,600 characters
Date	Date/Time	Jan 1, 1753 A.D. to Dec 31, 4712 A.D. (precision to microsecond)
Double	Double	Precision of 15 digits
Float	Double	Precision of 15 digits
Integer	Integer	-2,147,483,648 to 2,147,483,647
SmallInt	Integer	-32,768 to 32,767
String	String	1 to 104,857,600 characters
Timestamp	Date/Time	1 to 19 characters Precision 19 to 26, scale 0 to 6
TinyInt	Integer	-128 to 127
Varchar	String	1 to 104,857,600 characters

Parquet Google Cloud Storage file data types and transformation data types

Parquet Google Cloud Storage file data types map to transformation data types that the Secure Agent uses to move data across platforms.

The following table lists the Google Cloud Storage file data types that the Secure Agent supports and the corresponding transformation data types:

Parquet Google Cloud Storage File Data Type	Transformation	Description
BIGINT	Bigint	Precision of 19 digits, scale of 0.
BOOLEAN	Integer	TRUE (1) or FALSE (0)
DATE	Date/Time	January 1, 0001 to December 31, 9999.
DECIMAL	Decimal	Precision 1 to 28 digits, scale 0 to 28. Note: You cannot use decimal values with precision greater than 28.

Parquet Google Cloud Storage File Data Type	Transformation	Description
DOUBLE	Double	Precision of 15 digits.
FLOAT	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
Int96	Date/Time	Jan1,0001 to Dec 31, 9999. Precision of 29, scale of 9.
String	String	-1 to 104,857,600 characters.
Time	Date/Time	Time of the day. Precision to microsecond.
Timestamp	Date/Time	January 1, 0001 00:00:00 to December 31, 9999 23:59:59.997. Precision to microsecond. Note: You cannot set precision to nanoseconds.

Note: The following Parquet complex data types are not applicable for Google Cloud Storage V2 Connector:

- Byte_Array
- Enum
- Fixed Length Byte Array
- Union

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

INDEX

A

Avro Google Cloud Storage file data types
transformation data types [23](#)

C

Cloud Application Integration community
URL [5](#)
Cloud Developer community
URL [5](#)
connections
Google Cloud Storage V2 [11](#)

D

data encryption
Informatica encryption [14](#)
Data Integration community
URL [5](#)
data type reference
overview [22](#)
directory source
Google Cloud Storage sources [13](#)
dynamic schema handling [15](#)

G

Google Cloud Storage
access control lists [8](#)
components [8](#)
features [8](#)
introduction [8](#)
native data types [22](#)
Google Cloud Storage components
buckets [8](#)
objects [8](#)
projects [8](#)
Google Cloud Storage connections
overview [11](#)
Google Cloud Storage Connector
administration [7](#)
Google Cloud Storage V2
connection properties [11](#)
Google Cloud Storage V2 Connector
overview [7](#)
supported task and object types [7](#)

I

Informatica Global Customer Support
contact information [6](#)

Informatica Intelligent Cloud Services
web site [5](#)

J

JSON Google Cloud Storage file data types
transformation data types [24](#)

M

maintenance outages [6](#)
mapping
Google Cloud Storage V2 sources [15](#)
mapping advanced properties
Google Cloud Storage V2 source [15](#)
mapping properties
Google Cloud Storage V2 source [15](#)

O

ORC file data types
transformation data types [24](#)

P

Parquet Google Cloud Storage file data types
transformation data types [25](#)

R

rules and guidelines
Avro format types [20](#)
parquet format types [20](#)

S

status
Informatica Intelligent Cloud Services [6](#)
system status [6](#)

T

transformation
data types [22](#)
trust site
description [6](#)

U

upgrade notifications [6](#)

W

web site [5](#)