Informatica® Cloud (Version Summer 2016)

CloudFileTransfer Connector Guide
# Table of Contents

**Preface** .......................................................... 5  
Informatica Resources. ........................................... 5  
  Informatica Network. .......................................... 5  
  Informatica Knowledge Base. ................................ 5  
  Informatica Documentation. .................................... 6  
  Informatica Product Availability Matrixes. ............... 6  
  Informatica Velocity. ......................................... 6  
  Informatica Marketplace. ...................................... 6  
  Informatica Global Customer Support. ....................... 6  

**Chapter 1: Introduction to CloudFileTransfer Connector** ............. 7  
  CloudFileTransfer Connector Overview. ...................... 7  
  CloudFileTransfer Connector Task and Object Types. .......... 7  

**Chapter 2: CloudFileTransfer Connections** .......................... 8  
  CloudFileTransfer Connection Overview. ..................... 8  
  Creating a CloudFileTransfer Connection. .................... 8  

**Chapter 3: Data Synchronization Tasks with CloudFileTransfer** ....... 10  
  CloudFileTransfer Sources in Data Synchronization Tasks. .... 10  
  CloudFileTransfer Targets in Data Synchronization Tasks. ........ 11  
  CloudFileTransfer Data Filters. ................................ 11  
  Data Synchronization Example. ................................ 12  

**Chapter 4: Mappings and Mapping Configuration Tasks with CloudFileTransfer** .......................................................... 15  
  Mappings CloudFileTransfer Overview. ....................... 15  
  Creating Mapping and Mapping Configuration task. ............... 15  

**Chapter 5: CloudFileTransfer Data Type Reference** ............... 19  
  Data Type Reference Overview. ................................. 19  

**Index** ........................................................... 20
Preface

The Informatica Cloud CloudFileTransfer Connector Guide contains information about how to set up and use CloudFileTransfer Connector. The guide explains how organization administrators and business users can use CloudFileTransfer Connector to read data from Google Cloud Storage and write data to Amazon Simple Storage Service (Amazon S3).

Informatica Resources

Informatica Network


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.

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.
- 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.
Introduction to CloudFileTransfer Connector

This chapter includes the following topics:

- CloudFileTransfer Connector Overview, 7
- CloudFileTransfer Connector Task and Object Types, 7

CloudFileTransfer Connector Overview

You can use CloudFileTransfer Connector to connect to Google Cloud and Amazon Simple Storage Service (Amazon S3) from Informatica Cloud.

Use CloudFileTransfer Connector to read data from Google Cloud Storage bucket and write data to Amazon S3 bucket. You can use CloudFileTransfer objects as sources and targets in Data Synchronization tasks, mappings, and Mapping Configuration tasks.

CloudFileTransfer Connector Task and Object Types

The following table lists the CloudFileTransfer object types that you can include in Informatica Cloud tasks:

<table>
<thead>
<tr>
<th>Task Type</th>
<th>Source</th>
<th>Target</th>
<th>Lookup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Synchronization</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Mapping</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Mapping Configuration</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
This chapter includes the following topics:

- CloudFileTransfer Connection Overview, 8
- Creating a CloudFileTransfer Connection, 8

CloudFileTransfer Connection Overview

Create a CloudFileTransfer connection to access Google Cloud Storage data and write to Amazon S3 from Informatica Cloud. You can create a connection on the Connections page or when you create a task. After you create a connection, it becomes available to all users who have access to the organization.

Use the connection when you create a Data Synchronization task or a Mapping Configuration task. You can specify the CloudFileTransfer source and target in mappings, Mapping Configuration tasks, or Data Synchronization tasks.

Creating a CloudFileTransfer Connection

To use CloudFileTransfer Connector in data synchronization task, you must create a connection in Informatica Cloud.

Follow the following steps to create CloudFileTransfer connection in Informatica Cloud.

1. In Informatica Cloud home page, click **Configure**.
2. The drop-down menu appears, select **Connections**. The Connections page appears.
3. Click **New** to create a connection. The New Connection page appears.

![New Connection Page]

4. Specify the following details.

<table>
<thead>
<tr>
<th>Connection Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connection Name</td>
<td>Name of the CloudFileTransfer connection.</td>
</tr>
<tr>
<td>Description</td>
<td>Description of the connection. The description cannot exceed 765 characters.</td>
</tr>
<tr>
<td>Type</td>
<td>Type of connection. Select the CloudFileTransfer connection.</td>
</tr>
<tr>
<td>Runtime Environment</td>
<td>The name of the runtime environment where you want to run the tasks.</td>
</tr>
<tr>
<td>Google Service Key Json Content</td>
<td>The authentication key to validate the user identity. To generate a private key in JSON, see <a href="https://cloud.google.com/storage/docs/authentication#service_accounts">https://cloud.google.com/storage/docs/authentication#service_accounts</a>. After you create the JSON file that contains the Authentication key, copy the contents of the JSON file and paste it to Google Private Key Json Content field.</td>
</tr>
<tr>
<td>AmazonS3 Access Key</td>
<td>The access key ID used to access the Amazon account resources. <strong>Note:</strong> Ensure that you have valid Amazon Web Services credentials before you create a connection.</td>
</tr>
<tr>
<td>AmazonS3 Secret Key</td>
<td>The secret access key used to access the Amazon account resources. This value is associated with the access key and uniquely identifies the account. You must specify this value if you specify the access key ID.</td>
</tr>
</tbody>
</table>

**Note:** The Secure agent, Google bucket and Amazon S3 bucket should locate in the same region to avoid performance issues. The file transfer slows down when the Secure Agent, Amazon S3 bucket and Google bucket are located in the different regions.

5. Click **Test** to evaluate the connection.

6. Click **Ok** to save the connection.
This chapter includes the following topics:

- CloudFileTransfer Sources in Data Synchronization Tasks, 10
- CloudFileTransfer Targets in Data Synchronization Tasks, 11
- CloudFileTransfer Data Filters, 11
- Data Synchronization Example, 12

CloudFileTransfer Sources in Data Synchronization Tasks

When you configure a Data Synchronization task to use a CloudFileTransfer source, you can configure the source properties.

The source properties appear on the Source page of the Data Synchronization Task wizard when you specify a CloudFileTransfer connection.

The general properties display the name and description of the CloudFileTransfer source. Configure the source properties for the source object.

The following table describes the CloudFileTransfer source properties:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connection</td>
<td>Name of the connection.</td>
</tr>
<tr>
<td>Source Type</td>
<td>Select Single.</td>
</tr>
<tr>
<td>Source Object</td>
<td>Source object for the Data Synchronization task. Select the GoogleCloudtoAmazonS3 object.</td>
</tr>
</tbody>
</table>
CloudFileTransfer Targets in Data Synchronization Tasks

When you configure a Data Synchronization task to perform operations on a CloudFileTransfer target, you can configure the target properties.

The target properties appear on the **Target** page of the Data Synchronization Task wizard when you specify a CloudFileTransfer connection. The general properties display the name and description of the CloudFileTransfer target.

The following table describes the CloudFileTransfer target properties:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connection</td>
<td>Name of the target connection.</td>
</tr>
<tr>
<td>Target Object</td>
<td>Select Transfer Status object from the list to know the status of the transfer of the files when you run a Data Synchronization task or Mapping Configuration task to upload files to Amazon S3.</td>
</tr>
<tr>
<td>Child Object</td>
<td>NA</td>
</tr>
</tbody>
</table>

CloudFileTransfer Data Filters

You can use data filters to fetch specific data of a particular object. The Data Synchronization task processes the data based on the filter field you assign to the object.

Perform the following steps to use data filters:

1. Select **Data Filters** tab in the Data Synchronization task.
   
   The Data Filters tab appears.

2. Click **New**.

   The **Data Filter** dialog box appears as shown in the following image:
3. Specify the following details:

<table>
<thead>
<tr>
<th>Field Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objects</td>
<td>Select object for which you want to assign filter fields.</td>
</tr>
<tr>
<td>Filter By</td>
<td>Select filter field. You can select the following filter fields from the list:</td>
</tr>
<tr>
<td></td>
<td>- GoogleBucketName: Name of the Google bucket from where you want to read the data.</td>
</tr>
<tr>
<td></td>
<td>- FileName: The name of the file you want to read from Google Cloud and write to Amazon S3.</td>
</tr>
<tr>
<td></td>
<td>- AmazonS3BucketName: Name of the Amazon bucket where you want to write the data.</td>
</tr>
<tr>
<td></td>
<td>- Prefix: The prefix of the file or folder name that you want to read from Google Cloud and write to Amazon S3.</td>
</tr>
<tr>
<td></td>
<td>- AmazonS3Subfolder: The name of the subfolder present in the Amazon S3 bucket.</td>
</tr>
<tr>
<td></td>
<td>You must mention the name of the Google bucket and Amazon S3 bucket that exist in your Google and Amazon S3 account. You must specify values to the GoogleBucketName and AmazonS3BucketName filter fields. The Data Synchronization or Mapping Configuration task fails if you do not specify values to the GoogleBucketName and AmazonS3BucketName filter fields or specify the Amazon S3 and Google bucket names that do not exist.</td>
</tr>
<tr>
<td>Operator</td>
<td>When you enter the filter fields, you must select an operator. Select the Equals operator as CloudFileTransfer Connector does not support other operators.</td>
</tr>
<tr>
<td>Filter Value</td>
<td>Enter filter value.</td>
</tr>
</tbody>
</table>

4. Click OK.

**Data Synchronization Example**

You are a data administrator in a product organization. You want to collate legacy sales data from Google Cloud Storage and archive it on Amazon S3. You can read data from Google Cloud Storage and use CloudFileTransfer Connector to upload data to Amazon S3. Configure a Data Synchronization task with the insert operation.

Perform the following steps to setup a Data Synchronization task in Informatica Cloud:

1. Click **Apps** on Informatica Cloud home page.
2. Select **Data Synchronization** from the menu.
   - The Data Synchronization page appears.
3. Click **New** to create Data Synchronization task.
4. Specify the task name.
5. Provide a description for the task.
6. Select the task operation.
7. Click Next. The Source tab appears.
   
The following image shows the details of the Source tab:

8. Select the Source Connection, Source Type, and Source Object to be used for the task. Use the CloudFileTransfer connection to select GoogleCloudToAmazonS3 object.
9. Click Next. The Target tab appears. The following image shows the details of Target tab.
10. Select the **Connection** and **Target Object** required for the task. Use the CloudFileTransfer connection to select TransferStatus object to know the status of the file upload from Google Cloud Storage to Amazon S3.

11. Click **Next**. The **Data Filters** page appears.

   In **Data Filters**, you must specify values of the GoogleBucketName and AmazonS3BucketName filter fields. Also, you can specify other filter fields. The following image shows the filter fields with sample values:

   ![Data Filters](image)

<table>
<thead>
<tr>
<th>Actions</th>
<th>Object</th>
<th>Filter by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GoogleCloudToAmazonS3</td>
<td>GoogleBucketName Equals 'Scoulcace'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AmazonS3BucketName Equals 'htestinfaace'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AmazonS3Subfolder Equals 'Prefixed'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Prefix Equals 'Subscription'</td>
</tr>
</tbody>
</table>

12. Click **Next**.

13. In **Field Mapping** tab, map source fields to target fields accordingly or click **Automatch** to map the fields of the source object to the target object.

   ![Field Map](image)

   The Field Map tab is shown in the following image:

14. Click **Next**.

   The **Schedule** tab appears where you can schedule the task accordingly to the requirement and save. When you run the Data Synchronization task, the Secure Agent writes the retrieved data from Google Cloud bucket to Amazon S3 bucket.
This chapter includes the following topics:

- Mappings and CloudFileTransfer Overview, 15
- Creating Mapping and Mapping Configuration task, 15

Mappings CloudFileTransfer Overview

Use the Informatica Cloud Mapping Designer to create a mapping. When you create a mapping, you configure a source or target to represent a single CloudFileTransfer object.

Describe the flow of data from source and target along with the required transformations before the Secure Agent writes data to the target. When you create a Mapping Configuration task, select the mapping that you want to use. Use the Mapping Configuration Task wizard to create a Mapping Configuration task. The Mapping Configuration task processes data based on the data flow logic you define in the mapping.

Creating Mapping and Mapping Configuration task

Perform the following steps to setup a Mapping Configuration task in Informatica Cloud:

1. In Design menu, click Mappings on Informatica Cloud home page.
2. Click **New Mapping**. The new mapping page appears.

3. Specify the name and description. Click **Ok**

4. Click **Source** to configure the source properties.

   The following image shows the details of **Source** tab:

5. Specify the source connection, select **Single Object** from the **Source Type** and select the GoogleCloudtoAmazonS3 object. Click **Ok**.

6. In the **Source** tab, select **Query Options**. In **Query Options** page, select **Not Parameterize** from the **Filter Condition** list. Configure the filter fields. You can configure the filter fields as per the following image:

   **Note:** You must mention the name of the Google bucket and Amazon S3 bucket that exist in your Google and Amazon S3 account. When you specify the name of the Amazon S3 or Google Cloud bucket that
does not exist or when you do not specify the Google and Amazon S3 buckets, the Data Synchronization task or Mapping Configuration task fails.

7. Click the target button to configure the target. Specify the target connection, target type, object and select insert operation. Click **OK**


9. In the **Definition** tab specify the task name, description, runtime environment, select task based on mapping, and import the mapping.
The following image shows the details of the Mapping Configuration Task wizard:

10. Configure source properties if any.
11. Save and run the task.
This chapter includes the following topic:
- Data Type Reference Overview, 19

Data Type Reference Overview

Informatica Cloud uses the following data types in mappings, Data Synchronization tasks, and Mapping Configuration tasks with CloudFileTransfer:

**CloudFileTransfer native data types**

CloudFileTransfer data types appear in the source and target transformations when you choose to edit metadata for the fields.

**Transformation data types**

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

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

The following table lists the Amazon S3 and Google Cloud Storage data types that Informatica Cloud supports and the corresponding transformation data types:

<table>
<thead>
<tr>
<th>CloudFileTransfer Native Data Type</th>
<th>Transformation Data Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>String</td>
<td>String</td>
<td>1 to 104,857,600 characters</td>
</tr>
</tbody>
</table>