Informatica

Informatica® Cloud Data Integration

Microsoft Azure Blob Storage V3 Connector

Informatica Cloud Data Integration Microsoft Azure Blob Storage V3 Connector
April 2024

Publication Date: 2024-04-17

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 Microsoft Azure Blob Storage V3 Connector 7

Microsoft Azure Blob Storage V3 Connector assets 7

Chapter 2: Connections for Microsoft Azure Blob Storage V3 8

Prepare for authentication 8

- Shared Key authentication 8
- Shared access signature authentication 9

Connect to Microsoft Azure Blob Storage V3 10

- Before you begin 10
- Connection details 11
- Authentication types 11

Proxy server settings 12

Chapter 3: Mappings for Microsoft Azure Blob Storage V3 14

Data compression in Microsoft Azure Blob Storage V3 sources and targets 14

Directory source in Microsoft Azure Blob Storage sources 15

Microsoft Azure Blob Storage V3 sources in mappings 16

Microsoft Azure Blob Storage V3 targets in mappings 17

- Specifying a target 19
- Target time stamps 19

File formatting options 20

Microsoft Azure Blob Storage V3 target file parameterization 22

- Microsoft Azure Blob Storage V3 target file parameterization through a parameter file 22

Rules and guidelines for mappings and mapping tasks 22

Chapter 4: Data type reference 23

Microsoft Azure Blob Storage V3 and transformation data types 23

Avro data types and transformation data types 24

JSON data types and transformation data types 24

Parquet data types and transformation data types. 25

Chapter 5: Troubleshooting. 26

Troubleshooting a connection. 26

Troubleshooting a mapping or mapping task. 26

Index. 28

Preface

Use *Microsoft Azure Blob Storage V3 Connector* to learn how to read from or write to Microsoft Azure Blob Storage. Learn to create a connection, develop and run mappings, mapping tasks, and data transfer tasks in Cloud 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, on the Informatica Intelligent Cloud Services Status page, click **SUBSCRIBE TO UPDATES**. You can 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 Global Support Center through the Informatica Network or by telephone.

To find online support resources on the Informatica Network, click **Contact Support** in the Informatica Intelligent Cloud Services Help menu to go to the **Cloud Support** page. The **Cloud Support** page includes system status information and community discussions. Log in to Informatica Network and click **Need Help** to find additional resources and to contact Informatica Global Customer Support through email.

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 Microsoft Azure Blob Storage V3 Connector

You can use Microsoft Azure Blob Storage V3 Connector to securely read from or write to Microsoft Azure Blob Storage.

Use Microsoft Azure Blob Storage V3 Connector to read and write delimited files and complex files such as Avro, JSON, and Parquet.

You can use a Microsoft Azure Blob Storage V3 connection to access delimited, Avro, and Parquet files that are block blobs or append blobs.

You cannot read and write nested and multi-line indented JSON files.

You can use Microsoft Azure Blob Storage V3 objects as sources and targets in mappings and mapping tasks.

You can switch mappings to advanced mode to include transformations and functions that enable advanced functionality.

Microsoft Azure Blob Storage V3 Connector assets

Create assets in Data Integration to integrate data using Microsoft Azure Blob Storage V3 Connector.

When you use Microsoft Azure Blob Storage V3 Connector, you can include the following Data Integration assets:

- Data transfer task
- Mapping
- Mapping task

You cannot configure a Lookup transformation in a mapping and mapping task.

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

CHAPTER 2

Connections for Microsoft Azure Blob Storage V3

Create a Microsoft Azure Blob Storage V3 connection to securely read data from or write data to Microsoft Azure Blob Storage.

You can use a Microsoft Azure Blob Storage V3 connection to specify sources and targets in mappings and mapping tasks.

Prepare for authentication

You can use Microsoft Azure Blob Storage V3 Connector to connect to Microsoft Azure Blob Storage using shared key authentication or shared access signature authentication.

Before you configure authentication, create a storage account to use with Microsoft Azure Blob Storage and create a blob container in the storage account. For more information on how to create a storage account and a blob container, see the Informatica How-To Library article,
Prerequisites to create a Microsoft Azure Blob Storage V3 connection.

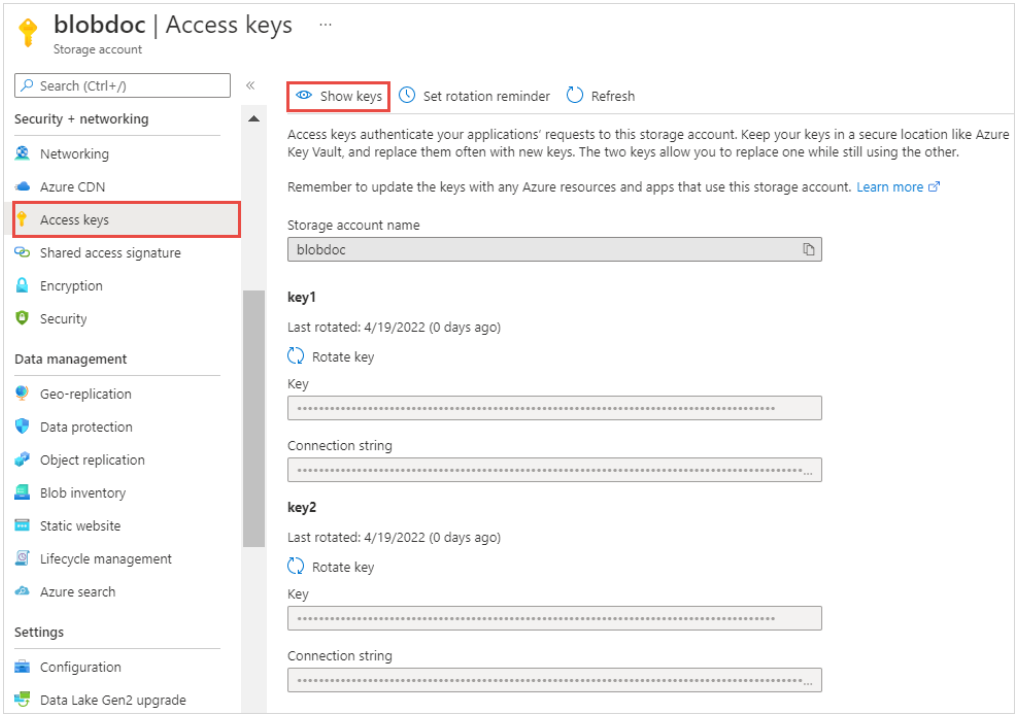
Before you configure the connection properties, you also need to keep the authentication details handy based on the authentication type that you want to use.

Shared Key authentication

To connect to Microsoft Azure Blob Storage using shared key authentication, you need the storage account name and account key.

1. Open the storage account.
2. Under **Security + Networking**, click **Access keys**.

3. Click **Show keys**.



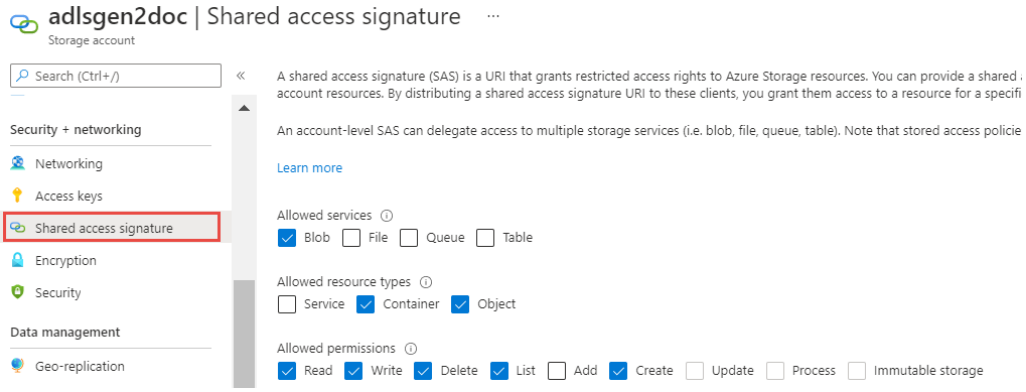
4. Make a note of the storage account name and account key. You can use key1 or key2.

Shared access signature authentication

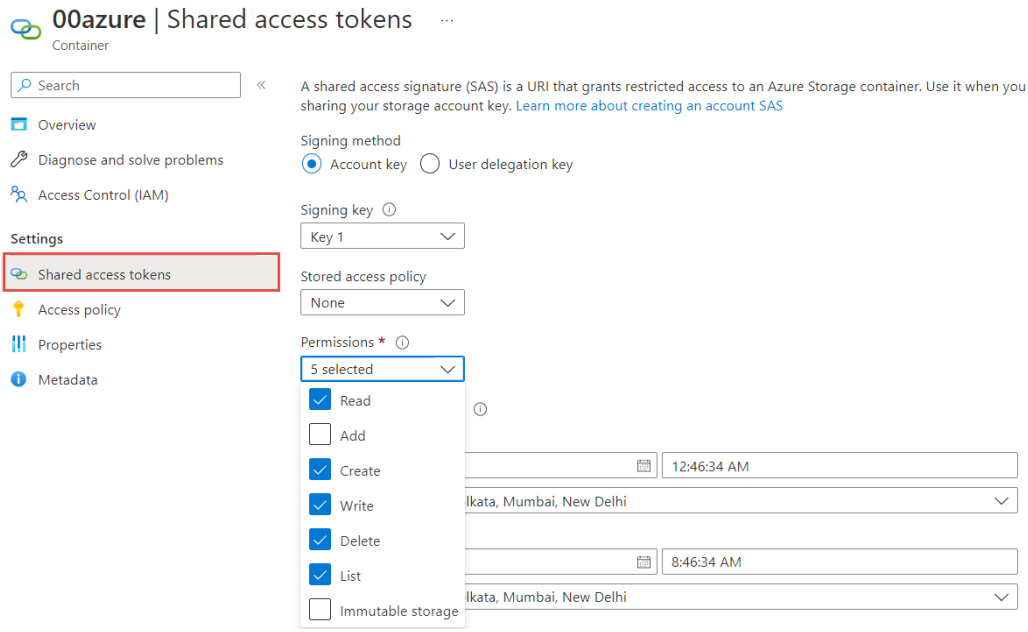
To connect to Microsoft Azure Blob Storage using shared access signature, you need to configure the minimum permissions for shared access signature authentication and generate the SAS token in the Azure portal.

You can generate the SAS token for the storage account or for the container.

- To generate the SAS token for the storage account, on the Azure portal, go to **Security + Networking**, and click **Shared access signature**.
  The following image shows the minimum permissions required for shared access signature authentication:



- To generate the SAS token for the Blob container, go to **Settings** of the container, and click **Shared access tokens**.
  The following image shows the minimum permissions required for shared access signature authentication:



Note: If you use the User delegation key signing method, ensure that you have the **Storage Blob Data Owner** role for the container or the storage account.

Connect to Microsoft Azure Blob Storage V3

Let's configure the Microsoft Azure Blob Storage V3 connection properties to connect to Microsoft Azure Blob Storage.

Before you begin

Before you get started, you'll need to get information from your Microsoft Azure Blob Storage account based on the authentication type that you want to configure.

Check out "Prepare for authentication" on page 8 to learn more about the authentication prerequisites.

Connection details

The following table describes the basic 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	Microsoft Azure Blob Storage V3
Use Secret Vault	Stores sensitive credentials for this connection in the secrets manager that is configured for your organization. This property appears only if secrets manager is set up for your organization. When you enable the secret vault in the connection, you can select which credentials that the Secure Agent retrieves from the secrets manager. If you don't enable this option, the credentials are stored in the repository or on a local Secure Agent, depending on how your organization is configured. For information about how to configure and use a secrets manager, see "Secrets manager configuration" in the Administrator help.
Runtime Environment	The name of the runtime environment where you want to run tasks. Select a Secure Agent, Hosted Agent, or serverless runtime environment.
Account Name	Microsoft Azure Blob Storage account name.

Authentication types

You can configure shared key authentication and shared access signature authentication types to access Microsoft Azure Blob Storage.

Select the required authentication method and then configure the authentication-specific parameters.

Shared key authentication

Shared key authentication uses the account key to connect to Microsoft Azure Blob Storage.

The following table describes the connection properties for shared key authentication:

Property	Description
Account Key	The account key for the Microsoft Azure Blob Storage account.
Container Name	The name of the blob container in the Microsoft Azure Blob Storage account.
Endpoint Suffix	Types of Microsoft Azure endpoints. Select one of the following options: - core.windows.net. Connects to Azure endpoints. - core.usgovcloudapi.net. Connects to Azure Government endpoints. - core.chinacloudapi.cn. Not applicable. Default is core.windows.net.

Shared access signature authentication

Shared access signature authentication uses the SAS token to connect to Microsoft Azure Blob Storage. Use the SAS token to grant access to the resources in the storage account or container for a specific time range without sharing the account key.

Note:

The following table describes the connection properties for shared access signature authentication:

Property	Description
SAS Token	The shared access signature token generated in the Azure portal to authenticate successfully and gain access to the Microsoft Azure Blob Storage resources.
Container Name	The name of the blob container in the Microsoft Azure Blob Storage account.
Endpoint Suffix	Types of Microsoft Azure endpoints. Select one of the following options: - core.windows.net. Connects to Azure endpoints. - core.usgovcloudapi.net. Connects to Azure Government endpoints. - core.chinacloudapi.cn. Not applicable. Default is core.windows.net.

Proxy server settings

If your organization uses an outgoing proxy server to connect to the Internet, the Secure 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 proxy server that requires only the host and port address for configuration.

To configure proxy settings for the Secure Agent, use one of the following methods:

- Configure the Secure Agent through the Secure Agent Manager on Windows or shell command on Linux.

  For instructions, see "Configure the proxy settings on Windows" or "Configure the proxy settings on Linux" in *Getting Started* in the Data Integration help .

- Configure proxy server through the JVM options. To do this, perform the following steps:
  1. Log in to Informatica Intelligent Cloud Services.
  2. Open Administrator and select **Runtime Environments**.
  3. Select the Secure Agent for which you want to configure the proxy server.
  4. On the upper-right corner of the page, click **Edit**.
  5. In the **System Configuration Details** section, select the **Type** as **DTM** for the Data Integration Service.
  6. Add the following parameters in any **JVMOption** field and specify appropriate values for each parameter:

Parameter	Description
-DproxyEnabled=	Required. Set the value to true to enable proxy server.
-Dhttp.proxyHost=	Required. Host name of the outgoing HTTP proxy server.
-Dhttp.proxyPort=	Required. Port number of the outgoing HTTP proxy server.

     Example for HTTP:

     ```
     JVMOption1=-DproxyEnabled=true
     JVMOption2=-Dhttp.proxyHost=<proxy_server_hostname>
     JVMOption3=-Dhttp.proxyPort=8081
     ```

  7. Click **Save**.
     The Secure Agent restarts to apply the settings.

To configure proxy settings for the serverless runtime environment, see "Using a proxy server" in *Runtime Environments* in the Administrator help.

CHAPTER 3

Mappings for Microsoft Azure Blob Storage V3

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

When you create a mapping, you define the Source transformation and Target transformation to represent a Microsoft Azure Blob Storage V3 object. Use the Mapping Designer in Data Integration to add the Source or Target transformations in the mapping canvas and configure the Microsoft Azure Blob Storage V3 source and target properties.

In advanced mode, the Mapping Designer updates the mapping canvas to include transformations and functions that enable advanced functionality.

You can use Monitor to monitor the jobs.

Data compression in Microsoft Azure Blob Storage V3 sources and targets

You can decompress data when you read data from Microsoft Azure Blob Storage and compress the data when you write data to Microsoft Azure Blob Storage.

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

For the Flat resource type, select only the Gzip compression format. The following table lists the compression formats for Avro and Parquet resource types:

Compression format	Avro File	Flat File	JSON File	Parquet File
None	Yes	Yes	Yes	Yes
Deflate*	Yes	N/A	No	No
Gzip	No	Yes	No	Yes
Bzip2	N/A	N/A	No	N/A
Lzo	N/A	N/A	No	N/A

Compression format	Avro File	Flat File	JSON File	Parquet File
Snappy*	Yes	N/A	No	Yes
**Select None to decompress the Deflate and Snappy file formats.*				

To read a compressed file from Microsoft Azure Blob Storage, 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`
Bzip2	`.BZ2`
Lzo	`.LZO`
Snappy	`.snappy`

Directory source in Microsoft Azure Blob Storage sources

You can select the type of source from which you want to read data.

You can select the following type of sources from the **Source Type** option under the advanced source properties:

- File
- Directory

Use the following rules and guidelines to select **Directory** as the source type:

- 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.
- All the files under a specified directory are parsed. The files under subdirectories are not parsed.
- The connector does not perform any validation if there are multiple blob formats in the directory you select and might result into errors.

Microsoft Azure Blob Storage V3 sources in mappings

In a mapping, you can configure a Source transformation to represent a Microsoft Azure Blob Storage V3 object.

The following table describes the Microsoft Azure Blob Storage V3 source properties that you can configure in a Source transformation:

Property	Description
Connection	Name of the source connection. Select a source connection or click **New Parameter** to define a new parameter for the source connection.
Source Type	Source type. Select one of the following types: - Single Object - Parameter: Select Parameter to define the source type when you configure the mapping task.
Object	Name of the source object. You can drill-down and select an object from a sub-folder to fetch metadata from a particular object. When you run a task, the Secure Agent reads data from the container you specified either in connection properties or in the advance properties.
Parameter	Select an existing 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.
Format	Specifies the file format that the Microsoft Azure Blob Storage V3 Connector uses to read data from Microsoft Azure Blob Storage. You can select the following file format types: - Flat - Avro - Parquet - JSON Default is **None**. You must select the **Format Type** as **None** to read binary files. For more information, see "File formatting options" on page 20.

The following table describes the Microsoft Azure Blob Storage V3 advanced source properties that you can configure in a Source transformation:

Property	Description
Number of concurrent connections to Blob Store	The number of concurrent connections to Blob Store to upload files. Default is 4.
Source Type	Select the type of source from which you want to read data. You can select the following source types: - File - Directory Default is File.

Property	Description
Blob Name Override	Overrides the default file name.
Blob Container Override	Overrides the default container name. When you read data from a directory and override the Blob container, ensure that files in the Blob container that you override with are not empty. When you generate the SAS token at the container-level, the default container name and the container name that you specify for the container override must be the same.
Compression Format	Decompresses data when you read data from Microsoft Azure Blob Storage. You can decompress the data in the following formats: - None. Select None to decompress deflate and snappy file formats. - Gzip - Bzip2 - Lzo Default is None.
Tracing Level	Sets the amount of detail that appears in the log file. You can choose terse, normal, verbose initialization, or verbose data. Default is normal.

Microsoft Azure Blob Storage V3 targets in mappings

In a mapping, you can configure a Target transformation to represent a Microsoft Azure Blob Storage V3 object.

The following table describes the Microsoft Azure Blob Storage V3 target properties that you can configure in a Target transformation:

Property	Description
Connection	Name of the target connection. Select a target connection or click **New Parameter** to define a new parameter for the target connection.
Target Type	Target type. Select one of the following types: - Single Object. - Parameter. Select **Parameter** to define the target type when you configure the task.
Object	Name of the target object. You can select an existing object or create a new target at run time. When you select **Create New at Runtime**, enter a name and path for the target object and select the source fields that you want to use. By default, all source fields are used. The **Path** attribute is not applicable. The target name can contain alphanumeric characters. You can use only a period (.), an underscore (_), an at the rate sign (@), a dollar sign ($), and a percentage sign (%) special characters in the file name. You can use parameters defined in a parameter file in the target name.
Parameter	Select an existing parameter for the target object or click **New Parameter** to define a new parameter for the target object. The **Parameter** property appears only if you select Parameter as the target type.

Property	Description
Format	Specifies the file format that the Microsoft Azure Blob Storage V3 Connector uses to read data from Microsoft Azure Blob Storage. You can select the following file format types: - Flat - Avro - Parquet - JSON Default is **None**. You must select the **Format Type** as **None** to read binary files. For more information, see "File formatting options" on page 20.
Operation	Target operation. Select **Insert**. You can only insert data to a Microsoft Azure Blob Storage target.

The following table describes the Microsoft Azure Blob Storage V3 advanced target properties that you can configure in a Target transformation:

Property	Description
Number of concurrent connections to Blob Store	The number of concurrent connections to Blob Store to upload files. Default is 4.
Blob Name Override	Overrides the default file name. You must use this property when you want to write compressed blob files to Microsoft Azure Blob Storage.
Blob Container Override	Overrides the default container name. When you create a new target at the run time and select the blob container override property, the Secure Agent generates an empty header file in the container specified in the connection. When you specify the blob container override in the target, ensure that you specify the file that you want to write to the target in the blob name override property. When you generate the SAS token at the container-level, the default container name and the container name that you specify for the container override must be the same.
Compression Format	Compresses data when you write data to Microsoft Azure Blob Storage. You can compress the data in the following formats: - None - Deflate - Gzip - Bzip2 - Lzo - Snappy Default is None.
Write Strategy	Appends block to a blob, when you select append blob. Applicable to `.csv` files only.
Blob Type	Writes data to a block blob or an append blob.
Forward Rejected Rows	Not applicable.

Specifying a target

You can use an existing target or create a target to hold the results of a mapping. If you choose to create the target, the agent creates the target when you run the task.

To specify the target properties, follow these steps:

1. Select the Target transformation in the mapping.
2. On the **Incoming Fields** tab, configure field rules to specify the fields to include in the target.
3. To specify the target, click the **Target** tab.
4. Select the target connection.
5. For the target type, choose **Single Object** or **Parameter**.
6. Specify the target object or parameter.
   - To create a target file at run time, enter the name for the target file including the extension, for example, `Accounts.csv`.

     Note: When you read from a flat file, ensure that the file contains some data and not the header alone. If the file has only a header, the header is not written to the target.
   - If you want the file name to include a time stamp, click **Handle Special Characters** and add special characters to the file name. For example, add the special characters shown here to include all the time stamp information: `Accounts_%d%m%y%T.csv`.
   - If you want the folder name to include a time stamp, click **Handle Special Characters** and add the folder name separated with a back slash (\) followed by the file name. For example, `%Y%m%d \Target_filename_%m.csv`.

   Note: The Handle Special Characters option is not applicable to mappings in advanced mode.
7. Click **Formatting Options** if you want to configure the formatting options for the file, and click **OK**.
8. Click **Select** and choose a target object. You can select an existing target object or create a new target object at run time and specify the object name.
9. Specify Advanced properties for the target, if needed.

Target time stamps

When you create a target at run time in a mapping, you can append time stamp information to the file name to show when the file is created.

When you specify the file name for the target file, include special characters based on Linux STRFTIME function formats that the mapping task uses to include time stamp information in the file name. The time stamp is based on the organization's time zone.

You cannot append time stamp information to the file name for mappings in advanced mode.

The following table describes some common STRFTIME function formats that you might use in a mapping or mapping task:

Special Character	Description
%d	Day as a two-decimal number, with a range of 01-31.
%m	Month as a two-decimal number, with a range of 01-12.
%y	Year as a two-decimal number without the century, with range of 00-99.

Special Character	Description
%Y	Year including the century, for example 2015.
%T	Applicable only to flat files. Time in 24-hour notation, equivalent to %H:%:M:%S.
%H	Hour in 24-hour clock notation, with a range of 00-24.
%I	Hour in 12-hour clock notation, with a range of 01-12.
%M	Minute as a decimal, with a range of 00-59.
%S	Second as a decimal, with a range of 00-60.
%p	Either AM or PM.

Note: For complex files, instead of %T you can use the equivalent %H_%M_%S.

File formatting options

When you select the format of a Microsoft Azure Blob Storage file, you can configure the formatting options.

The following table describes the formatting options for Avro, Parquet, JSON, and delimited flat files:

Property	Description
Schema Source	The schema of the source or target file. Select one of the following options to specify a schema: - Read from data file. Imports the schema from the file in Microsoft Azure Blob Storage. - Import from schema file. Imports schema from a schema definition file in your local machine.
Schema File	The schema definition file in the agent machine from where you want to upload the schema. You cannot upload a schema file when you select the Create Target option.

The following table describes the formatting options for flat files:

Property	Description
Flat File Type	The type of flat file. Select one of the following options: - Delimited. Reads a flat file that contains column delimiters. - Fixed Width. Not applicable.
Delimiter	Character used to separate columns of data. You can set values as comma, tab, colon, semicolon, or others. You can't set a tab as a delimiter directly in the **Delimiter** field. To set a tab as a delimiter, you must type the tab character in any text editor. Then, copy the character to the **Delimiter** field.

Property	Description
EscapeChar	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 text strings. You can configure parameters such as single quote or double quote. You can use the output text qualifier when a delimiter value is present in the data.
Qualifier Mode	Specify the qualifier behavior for the target object. You can select one of the following options: - Minimal. Default mode. Applies qualifier to data enclosed within a delimiter value or a special character. - All. Applies qualifier to all data.
Code Page	Select the code page that the Secure Agent must use to read or write data. Microsoft Azure Blob Storage V3 Connector supports only UTF-8. Ignore rest of the code pages.
Header Line Number	Specify the line number that you want to use as the header when you read data from Microsoft Azure Blob Storage. 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.
First Data Row	Specify the line number from where you want to start reading the data. Enter a value greater than or equal to one. 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.
Target Header	Select whether you want to write data to a target that contains a header or without a header in the flat file. You can select With Header or Without Header options. This property is not applicable when you read data from a Microsoft Azure Blob Storage V3 source.
Distribution Column	Not applicable.
escapeCharacterDataRetained	Not applicable.
maxRowsToPreview	Not applicable.
rowDelimiter	Not applicable.

The following table describes the formatting options for JSON files:

Property	Description
Data elements to sample	Not applicable.
Memory available to process data	Not applicable.

Microsoft Azure Blob Storage V3 target file parameterization

When you parameterize the file name and target folder location for Microsoft Azure Blob Storage V3 target objects, you can pass the file name and folder location at run time. If the folder does not exist, the Secure Agent creates the folder structure dynamically.

Microsoft Azure Blob Storage V3 target file parameterization through a parameter file

You can parameterize a Microsoft Azure Blob Storage V3 target file using a parameter file.

To parameterize a Microsoft Azure Blob Storage V3 target file using a parameter file, create a Microsoft Azure Blob Storage V3 target object and add parameters in the target object name and target object path. Define the parameter that you added for the target object in the parameter file. Then, place the parameter file in the following location and run the mapping task:

```
<Informatica Cloud Secure Agent\apps\Data_Integration_Server\data\userparameters>
```

Rules and guidelines for mappings and mapping tasks

Consider the following rules and guidelines when you configure mappings and mapping tasks:

- When you edit the metadata, all native data types change to Bigint and you cannot change the scale and precision of data types except for the string data type.
- When you write a JSON file to Microsoft Azure Data Lake Blob Storage, ensure that the column names do not contain unicode characters.
- Ensure that the scale of a double data type is not set to 0 in the target file when you read data from or write data to Microsoft Azure Blob Storage.
- The data preview and mapping fail if you read an Avro file that contains binary fields.
- Ensure that the field names in the source or target object do not contain special characters or unicode characters.
- You cannot preview data when you read or write a compressed file.
- When you write an Avro or a Parquet file, ensure that the file does not contain null values, else incorrect data in written in the target for the fields with null values.
- You cannot select append blob as blob type when you read or write a JSON file.

CHAPTER 4

Data type reference

Data Integration uses the following data types in Microsoft Azure Blob Storage V3 mappings and mapping tasks:

- Microsoft Azure Blob Storage V3 native 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. These are internal data types based on ANSI SQL-92 generic data types, which the Secure Agent uses to move data across platforms. They appear in all transformations in a mapping.

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

Microsoft Azure Blob Storage V3 and transformation data types

The following table lists the Microsoft Azure Blob Storage V3 data types that the Secure Agent supports and the corresponding transformation data types:

Microsoft Azure Blob Storage V3 Native Data Type	Transformation Data Type	Range and Description
String	String	1 to 104,857,600 characters

Avro data types and transformation data types

Avro file data types map to transformation data types that the Secure Agent uses to move data across platforms.

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

Avro 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

JSON data types and transformation data types

JSON complex file data types map to transformation data types that the Secure Agent uses to move data across platforms.

The following table lists the JSON complex file data types that the Data Integration supports and the corresponding transformation data types:

JSON Data Type	Transformation Data Type	Range and Description
boolean	integer	The default transformation type for boolean is integer. You can specify string data type with values of True and False. True is equivalent to the integer 1 and False is equivalent to the integer 0.
Number (double)	double	-1.79769313486231570E+308 to +1.79769313486231570E+308. Precision 15.

JSON Data Type	Transformation Data Type	Range and Description
Number (float)	double	-1.79769313486231570E+308 to +1.79769313486231570E+308. Precision 15.
Number (int)	integer	-2,147,483,648 to 2,147,483,647 Precision 10, scale 0
Number (long)	bigint	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision 19, scale 0.
string	string	1 to 104,857,600 characters.

Parquet data types and transformation data types

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

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

Parquet Data Type	Transformation Data Type	Range and Description
Boolean	Integer	TRUE (1) or FALSE (0)
Byte_Array	Binary	Arbitrarily long byte array
Double	Double	Precision 15
Float	Double	Precision 15
Int32	Integer	-2,147,483,648 to +2,147,483,647
Int64	Bigint	-9,223,372,036,854,775,808 to +9,223,372,036,854,775,807 8-byte signed integer
Int96	Binary	12-byte signed integer

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.

CHAPTER 5

Troubleshooting

Use the following sections to troubleshoot errors in mappings.

Troubleshooting a connection

The session log does not log the proxy server details

When you configure a proxy server through the Informatica Cloud Secure Agent user interface, the session log does not log the proxy server details.

Configure the proxy server by setting the JVM Options for your Secure Agent in the Administrator service.

Troubleshooting a mapping or mapping task

[ERROR] Exception: java.io.IOException: Too many open files

When you run a mapping on a Linux machine to read a large file, the mapping might fail with the following error:

```
[ERROR] Exception: java.io.IOException: Too many open files
```

To resolve this issue, perform the following steps:

1. Increase the value of file-max that is the maximum File Descriptors enforced on a kernel level. To change the file descriptor setting, edit the kernel parameter file `/etc/sysctl.conf` and add `fs.file-max=[new value]` to it.

   For example:

   ```
   # vi /etc/sysctl.conf
   fs.file-max = 400000
   ```

2. Set the ulimit. The ulimit must be less than file-max.
   To change the ulimit setting, edit the file `/etc/security/limits.conf` and set the hard and soft limits in it.

   For example:

   ```
   # vi /etc/security/limits.conf
   * soft nofile 40000
   * hard nofile 40000
   ```

When I write a JSON file, the mapping task fails with a Java heap space error.

When you write a JSON file of size 1 GB or more, the task fails with a Java heap space error.

Set the JVM options for type DTM to increase the -Xms and -Xmx values in the system configuration details of the Secure Agent.

When I use the create a new target at runtime to write an Avro file, the schema is created with primitive data types without providing an option to include null values.

You must manually edit the schema to allow null values as required. For example:

```
{"type":"record","name":"Azure_Avro_CT_N","fields":[
{"name":"c_custkey" , "type":["int","null"]},
{"name":"c_name" , "type":"string"},
{"name":"c_address" , "type":"string"},
{"name":"c_nationkey" , "type":["long","null"]}
```

The same error message is displayed for every failed mapping.

You can verify the error message in the session log.

INDEX

C

Cloud Application Integration community
  URL 5
Cloud Developer community
  URL 5
complex file format
  JSON 24
connections
  Microsoft Azure Blob Storage V3 10
create target
  adding time stamps 19
  target file parameterization 19, 22

D

data compression
  sources and targets 14
Data Integration community
  URL 5
data type reference
  overview 23
data types
  avro 24
  parquet 25
directory source
  Microsoft Azure Blob Storage sources 15

I

Informatica Global Customer Support
  contact information 6
Informatica Intelligent Cloud Services
  web site 5

M

maintenance outages 6
mappings
  target properties 17
  source properties 16
Microsoft Azure Blob Storage connection
  overview 8
Microsoft Azure Blob Storage V3
  connection properties 10
Microsoft Azure Blob Storage V3 Connector
  overview 7

P

parameterization through a parameter file 22

S

Source transformation
  Microsoft Azure Blob Storage properties 16
sources in mappings 16
specifying targets 19
status
  Informatica Intelligent Cloud Services 6
system status 6

T

Target transformation
  Microsoft Azure Blob Storage properties 17
targets in mappings 17
tracing level 16
trust site
  description 6

U

upgrade notifications 6

W

web site 5