Informatica® Cloud Data Integration
Summer 2018

Amazon S3 V2 Connector Guide

Informatica Cloud Data Integration Amazon S3 V2 Connector Guide
Summer 2018
October 2018

Publication Date: 2019-02-04

Table of Contents

CHAPTER 1

Introduction to Amazon S3 V2 Connector

This chapter includes the following topics:

Amazon S3 V2 Connector Overview

You can use Amazon S3 V2 Connector to connect Data Integration and Amazon S3. Use Amazon S3 V2 Connector to read or write Avro and Parquet file formats in Amazon S3.

You can create an Amazon S3 V2 connection and use the connection in mass ingestion tasks, mappings, or mapping tasks. Create a mass ingestion task to transfer files from an Amazon S3 source to any target that mass ingestion task supports and transfer files from any source that mass ingestion task supports to an Amazon S3 targets. Create a mapping task to process data based on the data flow logic defined in a mapping or integration template.

Amazon S3 V2 Connector supports Hosted Agent.

Note: Informatica recommends that you use Amazon S3 Connector if you want to run a synchronization task or mapping to read delimited files from and write delimited files to Amazon S3. For more information about using Amazon S3 Connector, see the *Informatica Cloud Data Integration Amazon S3 Connector User Guide*.

Amazon S3 V2 Supported Task Types and Object Types

The following table lists the Amazon S3 V2 object types that you can include in Data Integration tasks:

Task Type	Source	Target
Mass ingestion	Yes	Yes
Mapping	Yes	Yes

Introduction to Amazon S3

Amazon Storage Service (Amazon S3) is storage service in which you can copy data from source and simultaneously move data to any target. You can use Amazon S3 to transfer the files from a list of configured source connections to an Amazon S3 target. You can accomplish the tasks using the AWS Management Console web interface.

Amazon S3 stores data as objects within buckets. An object consists of a file and optionally any metadata that describes that file. Buckets are the containers for objects. You can have one or more buckets. When using the AWS Management Console, you can create folders to group objects, and you can nest folders.

Administration of Amazon S3 V2 Connector

As a user, you can use Amazon S3 V2 Connector after the organization administrator performs the following tasks:

- Manage Authentication. Use either of the following two methods:
  - Create an Access Key ID and Secret Access Key.
    Provide the values for access key ID and secret access key when you configure the Amazon S3 V2 connection. For more information about creating an access key ID and secret access key, see the AWS documentation.
  - Configure AWS Identity and Access Management (IAM) Authentication to enhance security.
    If you use IAM authentication, do not provide access key ID and secret access key explicitly in the Amazon S3 V2 connection.
- Create a master symmetric key if you want to enable client-side encryption.
- Create an AWS Key Management Service (AWS KMS)-managed customer master key if you want to enable server-side encryption.
- Create a minimal Amazon S3 bucket policy for Amazon S3 V2 Connector.

To use an Avro or Parquet file, you must ensure that only one Cloudera, Hortonworks or Amazon EMR license is available in the Secure Agent.

Create Minimal Amazon S3 Bucket Policy

The minimal Amazon S3 bucket policy restricts user operations and user access to particular Amazon S3 buckets by assigning an AWS Identity and Access Management (IAM) policy to users.

You can configure the IAM policy through the AWS console. Use AWS Identity and Access Management (IAM) authentication to securely control access to Amazon S3 resources. If you have valid AWS credentials and you want to use IAM authentication, you do not have to specify the access key and secret key when you create an Amazon S3 connection.

You can use the following minimum required actions for users to successfully read data from and write data to Amazon S3 bucket:

- PutObject
- GetObject
- DeleteObject
- ListBucket

Sample Policy:

```
{
"Version": "2012-10-17", "Statement": [ { "Effect": "Allow", "Principal": "*", "Action":
[ "s3:PutObject", "s3:GetObject", "s3:DeleteObject", "s3:ListBucket" ],"Resource":
[ "arn:aws:s3:::<bucket_name>/*", "arn:aws:s3:::<bucket_name>" ] } ]
}
```

IAM Authentication

Optional. You can configure IAM authentication when the Secure Agent runs on an Amazon Elastic Compute Cloud (EC2) system. Use IAM authentication for secure and controlled access to Amazon S3 resources when you run a session.

Perform the following steps to configure IAM authentication:

1. Create Minimal Amazon S3 Bucket Policy. For more information, see "Create Minimal Amazon S3 Bucket Policy" on page 7
2. Create the Amazon EC2 role. The Amazon EC2 role is used when you create an EC2 system in the S3 bucket. For more information about creating the Amazon EC2 role, see the AWS documentation.
3. Create an EC2 instance. Assign the Amazon EC2 role that you created in step #2 to the EC2 instance.
4. Install the Secure Agent on the EC2 system.

CHAPTER 2

Amazon S3 V2 Connections

This chapter includes the following topics:

Amazon S3 V2 Connections Overview

Amazon S3 V2 connection enables you to transfer files from an Amazon S3 source to an Amazon S3 target. You can use Amazon S3 V2 connections to read data from or write data to Amazon S3.

Use Amazon S3 V2 connections to specify sources or targets in a mass ingestion task, mapping and mapping task.

You can use AWS Identity and Access Management (IAM) authentication to securely control access to Amazon S3 resources. If you have valid AWS credentials and you want to use IAM authentication, you do not have to specify the access key and secret key when you create an Amazon S3 V2 connection.

Create an Amazon S3 V2 connection on the **Connections** page and associate it with a mass ingestion task, mapping, or mapping task. Define the source and target properties to read or write data to Amazon S3.

Note: If you enable both HTTP and SOCKS proxies, SOCKS proxy is used by default. If you want to use HTTP proxy instead of SOCKS proxy, set the value of the **DisableSocksProxy** property to true in the System property.

Amazon S3 V2 connection properties

When you set up an Amazon S3 V2 connection, you must configure the connection properties.

The following table describes the Amazon S3 V2 connection properties:

Property	Description
Connection Name	The name of the connection. The name is not case sensitive and must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters:~ ` ! $ % ^ & * ( ) - + = { [}] | \ : ; " ' < , > . ? /
Description	Optional. The description of the connection. The description cannot exceed 4,000 characters.
Type	The Amazon S3 V2 connection type.
Runtime Environment	The name of the runtime environment where you want to run the tasks.
Access Key	The access key ID for access to Amazon account resources.
Secret Key	The secret access key for access to Amazon account resources. The secret key is associated with the access key and uniquely identifies the account.
Folder Path	The complete path to Amazon S3 objects. The path must include the bucket name and any folder name. Do not use a slash at the end of the folder path. For example, `<bucket name>/<my folder name>`.
Master Symmetric Key	Optional. Provide a 256-bit AES encryption key in the Base64 format when you enable client-side encryption. You can generate a key using a third-party tool. **Note:** If you use a master symmetric key, you must replace the existing `JCE` files with the latest `JCE` files that are available in the Secure Agent installation location and restart the Secure Agent.

Property	Description
Customer Master Key ID	Optional. Specify the customer master key ID or alias name generated by AWS Key Management Service (AWS KMS). You must generate the customer master key for the same region where Amazon S3 bucket resides. You can specify the following master keys: **Customer generated customer master key** Enables client-side or server-side encryption. **Default customer master key** Enables client-side or server-side encryption. Only the administrator user of the account can use the default customer master key ID to enable client-side encryption.
Region Name	Select the AWS region in which the bucket you want to access resides. Select one of the following regions: - Asia Pacific(Mumbai) - Asia Pacific(Seoul) - Asia Pacific(Singapore) - Asia Pacific(Sydney) - Asia Pacific(Tokyo) - AWS GovCloud (US) - Canada(Central) - EU(Ireland) - EU(Frankfurt) - EU(London) - EU(Paris) - South America(Sao Paulo) - US East(Ohio) - US East(N. Virginia) - US West(N. California) - US West(Oregon) Default is US East (N. Virginia).

CHAPTER 3

Amazon S3 V2 Sources and Targets

This chapter includes the following topics:

Amazon S3 V2 Sources

You can use an Amazon S3 V2 object as a source in a mass ingestion task, mapping and mapping task.

When you configure the advanced source properties, you configure properties specific to Amazon S3 V2. You can download Amazon S3 V2 files in multiple parts, specify the location of the staging directory, and decompress the data when you read data from Amazon S3.

Client-side Encryption for Amazon S3 V2 Sources

Client-side encryption is a technique to encrypt data before transmitting the data to the Amazon S3 server.

You can read a client-side encrypted file in an Amazon S3 bucket. To read client-side encrypted files, you must provide a master symmetric key or customer master key in the connection properties. The Secure Agent encrypts the data by using the master symmetric key or customer master key.

To read a client-side encrypted file, perform the following tasks:

1. Provide a master symmetric key or customer master key in the connection properties.
2. Update the `JCE` file in the following directory: `<Secure Agent installation directory>\jre\lib\security`

   Note: You can download the `JCE` file that the JAVA environment on the Secure Agent machine supports from the Oracle website.
3. Select the client-side encrypted file in the mapping.

Source Types in Amazon S3 V2 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 Amazon S3 V2 advanced source properties:

File

You must enter the bucket name that contains the Amazon S3 file. If applicable, include the folder name that contains the target file in the `<bucket_name>/<folder_name>` format.

Amazon S3 V2 Connector provides the option to override the value of the **Folder Path** and **File Name** properties during run time.

If you do not provide the bucket name and specify the folder path starting with a slash (/) in the `/<folder_name>` format, the folder path appends with the folder path that you specified in the connection properties.

For example, if you specify the `/<dir2>` folder path in this property and `<my_bucket1>/<dir1>` folder path in the connection property, the folder path appends with the folder path that you specified in the connection properties in `<my_bucket1>/<dir1>/<dir2>` format.

If you specify the `<my_bucket1>/<dir1>` folder path in the connection property and `<my_bucket2>/<dir2>` folder path in this property, the Secure Agent writes the file in the `<my_bucket2>/<dir2>` folder path that you specify in this property.

Directory

You must select the source file when you create a mapping to select the source type as **Directory** at the run time. When you select the **Source Type** option as **Directory**, the value of the **File Name** is not honored.

For read operation, if you provide the **Folder Path** value during run time, the Secure Agent considers the value of the **Folder Path** from the advanced source properties. If you do not provide the **Folder Path** value during run time, the Secure Agent considers the value of the **Folder Path** that you specify during the connection creation.

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.

Working with Multiple Files

You can read multiple files, which are of flat format type, from Amazon S3 and write data to a target.

Note: Applicable when you create a mapping to read a file of flat format type.

To read multiple files, all files must be available in the same Amazon S3 bucket. When you want to read from multiple sources in the Amazon S3 bucket, you must create a `.manifest` file that contains all the source files with the respective absolute path or directory path. You must specify the `.manifest` file name in the following format: `<file_name>.manifest`

For example, the `.manifest` file contains source files in the following format:

```
{
 "fileLocations": [{
 "URIs": [
```

```
 "dir1/dir2/file_1.csv",
 "dir1/dir2/dir4/file_2.csv",
 "dirA/dirB/file_3.csv",
 "dirA/dirB/file_4.csv"
 ]
 }, {
 "URIPrefixes": [
 "dir1/dir2/",
 "dir1/dir2/"]
 }
 ],
 "settings":
}
```

The **Data Preview** tab displays the data of the first file available in the URI specified in the `.manifest` file. If the URI section is empty, the first file in the folder specified in URIPrefixes is displayed.

You can specify an asterisk (*) wildcard in the file name to fetch files from the Amazon S3 bucket. You can specify the asterisk (*) wildcard to fetch all the files or only the files that match the name pattern. Specify the wildcard character in the following format:

```
abc*.txt
abc.*
```

For example, if you specify `result*.txt`, all the file names starting with the term `result` and ending with the `.txt` file extension are read. If you specify `result.*`, all the file names starting with the term `result` are read regardless of the extension.

Use the wildcard character to specify files from a single folder. For example,

```
{
 "fileLocations": [{
 "URIs": [
 "automation/manual/AmazonS3_Input.csv"
 ]
 }, {
 "URIPrefixes": [
"automation/lookup/"
 ]
 },
 {
 "WildcardURIs": [

 "automation/new/**n*.csv"
    ]
 }
]
}
```

You cannot use the wildcard characters to specify folder names. For example,

```
{ "WildcardURIs": [ "multiread_wildcard/dir1*/", "multiread_wildcard/*/" ] }
```

Note: Amazon S3 V2 Connector supports only asterisk (*) wildcard character.

Source Partitioning

You can configure partitioning to optimize the mapping performance at run time when you read data from Amazon S3 sources.

Note: Applicable when you create a mapping to read a file of flat format type.

The partition type controls how the agent distributes data among partitions at partition points. You can define the partition type as passthrough partitioning. With partitioning, the Secure Agent distributes rows of source data based on the number of threads that you define as partition.

You can specify the value of the **Number of Partition** field in the **Partition** tab under the mapping task to configure partitioning for Amazon S3 V2 sources. The Secure Agent configures the partition for Amazon S3 V2 sources based on the value you enter in the **Number of Partition** field. By default, the value of the **Number of Partition** field is one.

The Secure Agent enables the partition according to the size of the Amazon S3 V2 source file. The file name is appended with a number starting from 0 in the following format: `<file name>_<number>`

If you enable partitioning and the precision for the source column is less than the maximum data length in that column, you might receive unexpected results. To avoid unexpected results, the precision for the source column must be equal to or greater than the maximum data length in that column for partitioning to work as expected.

Amazon S3 V2 Targets

You can use an Amazon S3 V2 object as a target in a mass ingestion task, mapping, or mapping task.

Specify the name and description of the Amazon S3 V2 target. Configure the Amazon S3 V2 target and advanced properties for the target object. If a mapping includes a flat file or an Amazon S3 target, you can choose to use an existing target or create a new target at run time.

Data Encryption in Amazon S3 V2 Targets

To protect data, you can encrypt the Amazon S3 files when you write the files to a target.

Note: You should not use the master symmetric key and customer master key together.

You can select the type of the encryption in the **Encryption Type** field under the Amazon S3 V2 advanced target properties on the **Schedule** page.

You can select the following types of encryption:

None

The data is not encrypted.

Server-side Encryption

Select **Server-side Encryption** as the encryption type if you want Amazon S3 to encrypt the data using Amazon S3-managed encryption keys when you write the files to the target.

Note: If you do not specify the customer master key ID in the connection properties, you must select **Server-side Encryption** as the encryption type.

Server-side Encryption with KMS

Select **Server-side Encryption with KMS** as the encryption type if you want Amazon S3 to encrypt the data using AWS KMS-managed customer master key encryption keys when you write the files to the target.

The AWS KMS-managed customer master key specified in the connection property must belong to the same region where Amazon S3 is hosted.

For example, if Amazon S3 is hosted in the **US West (Oregon)** region, you must use the AWS KMS-managed customer master key enabled in the same region when you select the **Server Side Encryption with KMS** encryption type.

Client-side Encryption

Select **Client-side Encryption** as the encryption type if you want the Secure Agent to encrypt the data when you write the files to the target. Client-side encryption uses a master symmetric key, which is a 256-bit AES encryption key in Base64 format or a customer master key.

To enable client-side encryption, perform the following tasks:

1. Ensure that an organization administrator creates a master symmetric key or customer master key ID when you create an Amazon S3 V2 connection.

   Note: The administrator user of the account can use the default customer master key ID to enable the client-side encryption.
2. Select **Client-side Encryption** as the encryption type in the advanced target properties.
3. Ensure that an organization administrator updates the security `JCE` files, required by the Amazon S3 client encryption policy, on the machine that hosts the Secure Agent.

The following table lists the encryption type for the support for various file types:

Encryption Type	Flat File	Avro File	Parquet File
Client-side Encryption	Yes	No	No
Server-side Encryption	Yes	No	No
Server-side Encryption with KMS	Yes	No	No

For information about the Amazon S3 client encryption policy, see the *Amazon S3 documentation*.

Overwriting Existing Files

You can choose to overwrite the existing target files.

Select the **Overwrite File(s) If Exists** option in the Amazon S3 V2 target advanced properties to overwrite the existing files. By default, the value of the **Overwrite File(s) If Exists** check box is true.

If you select the **Overwrite File(s) If Exists** option, the Secure Agent deletes the existing files with same file name and creates a new files with the same file name in the target directory.

If you do not select the **Overwrite File(s) If Exists** option, the Secure Agent does not delete the existing files in the target directory. The Secure Agent adds time stamp at the end of each target file name in the following format: `YYYYMMDD_HHMMSS_millisecond`. For example, the Secure Agent renames the target file name in the following format: `output.txt-20171220_091900_69844051`

Target Partitioning

You can configure partitioning to optimize the mapping performance at run time when you write data to Amazon S3 V2 targets.

Note: Applicable when you create a mapping to write a file of flat format type.

The partition type controls how the agent distributes data among partitions at partition points. You can define the partition type as passthrough partitioning. With partitioning, the Secure Agent distributes rows of target data based on the number of threads that you define as partition.

You can configure the **Merge Partition Files** options in the advanced target properties. You can specify whether the Secure Agent must merge the number of partition files as a single file or maintain separate files based on the number of partitions specified to write data to the Amazon S3 V2 targets.

If you do not select the **Merge Partition Files** option, separate files are created based on the number of partitions specified. The file name is appended with a number starting from 0 in the following format: `<file name>_<number>`

For example, the number of threads for the `Region.csv` file is three. If you do not select the **Merge Partition Files** option, the Secure Agent writes three separate files in the Amazon S3 V2 target in the following format:

```
<Region_0>
<Region_1>
<Region_2>
```

If you configure the **Merge Partition Files** option, the Secure Agent merges all the partitioned files as a single file and writes the file to Amazon S3 V2 target.

Distribution Column

You can write multiple files to Amazon S3 target from a single source. Configure the **Distribution Column** option in the advanced target properties.

Note: Applicable when you create a mapping to write a file of flat format type.

You can specify one column name in the **Distribution Column** field to create multiple target files during run time. When you specify the column name, the Secure Agent creates multiple target files in the column based on the column values that you specify in **Distribution Column**.

Each target file name is appended with the **Distribution Column** value in the following format:

```
<Target_fileName>+_+<Distribution column value>+<file extension>
```

Each target file contains all the columns of the table including the column that you specify in the **Distribution Column** field.

For example, the name of the target file is `Region.csv` that contains the values North America and South America. The following target files are created based on the values in the **Distribution Column** field:

```
Region_North America.csv
Region_South America.csv
```

You cannot specify two column names in the **Distribution Column** field. If you specify a column name that is not present in target field column, the task fails.

When you specify a column that contains value with special characters in the **Distribution Column** field, the Secure Agent fails to create target file if the corresponding Operating System do not support the special characters.

For example, the Secure Agent fails to create target file if the column contains date value in the following format: `YYYY/MM/DD`

Object Tag

You can add a tag to the object stored on the Amazon S3 bucket. Each tag contains a key value pair.

Note: Applicable when you create a mapping to write a file of flat format type.

Tagging an object helps to categorize the storage. You can add the object tags in the **Object Tags** field under the advanced target properties. Enter the object tag in the `Key=Value` format. You can also enter multiple object tags in the following format:

```
key1=Value1;key2=Value2
```

You can either enter the key value pairs or the specify the file path that contains the key value pairs. For example, you can specify the file path in the `C:\object\tags.txt` format. You can specify any file path on which the Secure Agent is installed.

When you upload new objects in the Amazon S3 bucket, you can add tags to the new objects or add tags to the existing objects. If the Secure Agent overrides a file that contains a tag in the Amazon S3 bucket, the tag is not retained. You must add a new tag for the overridden file. If you upload multiple files to the Amazon S3 bucket, each file that you upload must have the same set of tags associated with the multiple objects.

To add tags in the Amazon S3 V2 target object, you must add the `s3:PutObjectTagging` permission in the Amazon S3 policy. Following is the sample policy:

```
{
"Version": "2012-10-17",
"Id": "Policy1500966932533",
"Statement": [
{
"Sid": "Stmt1500966903029",
"Effect": "Allow",
"Principal": "*",
"Action": [
"s3:DeleteObject",
"s3:GetObject",
"s3:ListBucket",
"s3:PutObject",
"s3:PutObjectTagging"
],
"Resource": [
"arn:aws:s3:::<bucket_name>/*",
"arn:aws:s3:::<bucket_name>"
]
}
]
}
```

The following table lists the special characters that Amazon S3 V2 Connector supports during entering the key value pair:

Special Characters	Support
+	Yes
-	Yes
=	No
.	Yes
_	Yes
:	Yes
/	Yes

Rules and Guidelines for Tagging an Object

Use the following rules and guidelines for tagging an object:

- You can add maximum 10 tags for each object.
- When you enter a tag for an object, the tag must contain a unique tag key.

- The tag key can contain maximum 128 Unicode characters in length and tag values can contain maximum 256 Unicode characters in length.
- The key and values are case sensitive.

Data Compression in Amazon S3 V2 Sources and Targets

You can decompress data when you read data from Amazon S3 and compress the data when you write data to Amazon S3.

Configure the compression format in the **Compression Format** option under the advanced source and target properties. You must use the `.GZ` file name extension when you use the `Gzip` compression format. The source or target file in Amazon S3 V2 contains the same extension that you select in the **Compression Format** option.

When you run a task to read a compressed flat file, you must perform the following steps:

1. Select the required compressed file.
2. Navigate to **Formatting Options** property field.
3. Select **Import from schema file** option and upload the schema.
   The following figure shows a sample schema file for a flat file:

```
{"Columns":[{"Name":"f_varchar","Type":"string","Precision":"256","Scale":"0"},
{"Name":"f_char","Type":"string","Precision":"256","Scale":"0"},
{"Name":"f_smallint","Type":"string","Precision":"256","Scale":"0"},
{"Name":"f_integer","Type":"string","Precision":"256","Scale":"0"},
{"Name":"f_bigint","Type":"string","Precision":"256","Scale":"0"},
{"Name":"f_decimal_default","Type":"string","Precision":"256","Scale":"0"},
{"Name":"f_real","Type":"string","Precision":"256","Scale":"0"},
{"Name":"f_double_precision","Type":"string","Precision":"256","Scale":"0"},
{"Name":"f_boolean","Type":"string","Precision":"256","Scale":"0"},
{"Name":"f_date","Type":"string","Precision":"256","Scale":"0"},
{"Name":"f_timestamp","Type":"string","Precision":"256","Scale":"0"}]}
```

4. Select **Compression Format** as **GZIP** from the advanced source properties.

The following table lists the compression format for the support for various operations and file formats:

Compression format	Read	Write	Avro File	Parquet File
None	Yes	Yes	Yes	Yes
Gzip	Yes	Yes	No	Yes

CHAPTER 4

Mass ingestion Task with Amazon S3 V2 Connector

This chapter includes the following topics:

Mass ingestion Task Overview

Use mass ingestion tasks to transfer a large number of files of any file type between on-premises and cloud repositories, and to track and monitor file transfers.

Create an Amazon S3 V2 connection and use the connection to perform a mass ingestion task. When you create a mass ingestion task, select the source and target connection and then specify which files you want to move from the source to the target.

Example

You are a medical data analyst in a medical and pharmaceutical organization who maintains patient files. A patient file can contain data, such as patient details, doctor details, treatment history, and insurance in an Amazon S3 source file. You need to move the files that contains the patient details data from an Amazon S3 source file to a cloud-based environment for data analysis.

You can create a mass ingestion task to move all the files that contains the patient details data from an Amazon S3 source to an Amazon S3 target at once, instead of moving single row of data separately.

Before you begin

Before you create mass ingestion tasks, verify that the following conditions exist:

- The organization has the following licenses:
  - Mass Ingestion
  - Mass Ingestion Runtime
- The Mass Ingestion application is running on the Secure Agent.
- Source and target connections exist, based on the sources from where you want to transfer files and the targets to where you want to transfer files.

Amazon S3 V2 Sources in Mass ingestion Task

In a mass ingestion task, you can configure the Amazon S3 V2 source properties to transfer files from an Amazon S3 source to an Amazon S3 target or any target that mass ingestion task supports.

The following table describes the Amazon S3 V2 source properties that you can configure in a mass ingestion task:

Target Property	Description
Connection Type	Type of the source connection. Select **Amazon S3 V2** as the connection type.
Connection	Select the connection from a list of configured connections.

The following table describes the Amazon S3 V2 advanced source properties that you can configure in a mass ingestion task:

Option	Description
Folder Path	Amazon S3 folder path from where files are transferred, including bucket name. The Secure Agent must be able to access the folder. The default value is the folder path specified in the connection.
Include files from sub-folders	Transfer files from all sub-folders under the defined source directory.
Skip Duplicate Files	Do not transfer duplicate files. If files with the same name and creation date were transferred by the same mass ingestion task, the task does not transfer them again, and the files are marked as duplicate in the job log. If this option is not selected the task transfers all files.
File Pattern	File name pattern used to select the files to transfer. You can use the following wildcard character filters: - An asterisk (*) matches any number of characters. - A question mark (?) matches a single character.

Option	Description
File Encryption Type	Type of Amazon S3 file encryption to use during file transfer. The following options are available: - None. Files are not encrypted during transfer. - S3 server-side encryption. Amazon S3 encrypts the file using AWS-managed encryption keys. - S3 client-side encryption. Ensure that unrestricted policies are implemented for the AgentJVM, and that the master symmetric key for the connection is set.
S3 Accelerated Transfer	Select whether to use Amazon S3 Transfer Acceleration on the S3 bucket. To use Transfer Acceleration, accelerated transfer must be enabled for the bucket. The following options are available: - Disabled. Do not use Amazon S3 Transfer Acceleration. - Accelerated. Use Amazon S3 Transfer Acceleration. - Dualstack Accelerated. Use Amazon S3 Transfer Acceleration on a dual-stack endpoint.
Minimum Download Part Size	Minimum download part size when downloading a large file as a set of multiple independent parts, in megabytes.
Multipart Download Threshold	Multipart download minimum threshold to determine when to upload objects in multipleparts in parallel.

Amazon S3 V2 Targets in Mass ingestion Task

In a mass ingestion task, you can configure the Amazon S3 V2 target properties to transfer files from an Amazon S3 source or any source that mass ingestion task supports to an Amazon S3 target.

The following table describes the Amazon S3 V2 target properties that you can configure in a mass ingestion task:

Target Property	Description
Connection Type	Type of the target connection. Select **Amazon S3 V2** as the connection type.
Connection	Select the connection from a list of configured connections.

The following table describes the Amazon S3 V2 advanced target properties that you can configure in a mass ingestion task:

Option	Description
Folder Path	Amazon S3 folder path to where files are transferred, including bucket name. The Secure Agent must be able to access the folder. The default value is the folder path specified in the connection.
File Compression	Determines whether or not files are compressed before they are transferred to the target directory. The following options are available: - None. Files are not compressed. - GZIP. Files are compressed using GZIP compression.

Option	Description
File Encryption Type	Type of Amazon S3 file encryption to use during file transfer. The following options are available: - None. Files are not encrypted during transfer. - S3 server-side encryption. Amazon S3 encrypts the file using AWS-managed encryption keys. - S3 client-side encryption. Ensure that unrestricted policies are implemented for the AgentJVM, and that the master symmetric key for the connection is set.
S3 Accelerated Transfer	Select whether to use Amazon S3 Transfer Acceleration on the S3 bucket. To use Transfer Acceleration, accelerated transfer must be enabled for the bucket. The following options are available: - Disabled. Do not use Amazon S3 Transfer Acceleration. - Accelerated. Use Amazon S3 Transfer Acceleration. - Dualstack Accelerated. Use Amazon S3 Transfer Acceleration on a dual-stack endpoint.
Minimum Upload Part Size	Minimum upload part size when uploading a large file as a set of multiple independent parts, in megabytes. Use this option to tune the file load to Amazon S3.
Multipart Upload Threshold	Multipart download minimum threshold to determine when to upload objects in multipleparts in parallel.

Creating a Mass ingestion Task

You can create a mass ingestion task to transfer files from an Amazon S3 source to an Amazon S3 target.

1. In Data Integration, click **New** > **Tasks**.
2. Select **Mass Ingestion** and then click **Create**.

   The **Definition** tab appears.
3. In the **Definition** tab, configure the following properties:

Property	Description
Task Name	Name of the Mass Ingestion task. The names of Mass Ingestion tasks must be unique within the organization. Task names can contain alphanumeric characters, spaces, and the following special characters: _ . + - Task names are not case sensitive.
Location	Project folder in which the task resides. The default location is the last folder you viewed on the Explore page.
Runtime Environment	Runtime environment that contains the Secure Agent to run the task.
Description	Description of the task. Maximum length is 255 characters.

4. Click **Next**.

   The **Source** tab appears.

5. On the **Source Details** section, select the **Connection Type** as **Amazon S3 V2** and configure the Amazon S3 V2 source properties.
6. Click **View** to view the connection details.
7. Click **Test** to test the connection in the **View Connection** dialog.
8. Click **Next**.

   The **Target** tab appears.
9. On the **Target Details** section, select the **Connection Type** as **Amazon S3 V2** and configure the Amazon S3 V2 target properties.
10. Click **View** to view the connection details.
11. Click **Test** to test the connection in the **View Connection** dialog.
12. Click **Next**.

    The **Schedule** tab appears where you can select whether to run the task manually, or schedule the task to run at a specific time or when a file is ready.
13. Click **Run this task on schedule** to run the task on a schedule and select the schedule you want to use.

    If you do not want to run the task on a schedule, click **Do not run this task on a schedule** to run the task manually.
14. Click **Finish** to save and close the task wizard.

You can edit, run, or view the mass ingestion task on the **Explore** page after you create the mass ingestion task.

Viewing mass ingestion task details

You can view details about a mass ingestion task, including the source and target connections and the associated schedule.

1. On the **Explore** page, navigate to the task.
2. In the row that contains the task, click **Actions** and select **View**.

   The **Task Details** page appears with task, source, target, and schedule details.
3. You can edit or run the task that you selected to view. On the **Task Details** page, click **Edit** to modify the task or click **Run** to run the task.

Running a mass ingestion task

You can run a mass ingestion task in the following ways:

1. To run a mass ingestion task manually, on the **Explore** page, navigate to the task. In the row that contains the task, click **Actions** and select **Run**.

   Alternatively, you can run the task manually from the **Task Details** page. To access the **Task Details** page, click **Actions** and select **View**. In the **Task Details** page, select **Run**.
2. To run a mass ingestion task on a schedule, edit the task in the mass ingestion task wizard to associate the task with a schedule.

CHAPTER 5

Mappings and Mapping Tasks with Amazon S3 V2

This chapter includes the following topics:

Amazon S3 V2 Objects in Mappings

When you create a mapping, you can configure a Source or Target transformation to represent an Amazon S3 V2 object.

Amazon S3 V2 Source in Mappings

In a mapping, you can configure a Source transformation to represent an Amazon S3 V2 object as the source to read data from Amazon S3.

Specify the name and description of the Amazon S3 V2 source. Configure the Amazon S3 V2 source and advanced properties for the source object.

The following table describes the Amazon S3 V2 source properties that you can configure in a Source transformation:

Property	Description
Connection Name	Name of the Amazon S3 V2 source connection.
Source Type	Type of the source connection.

Property	Description
Object	Source object for the mapping.
Formatting Options	Amazon S3 format options. Opens the **Formatting Options** dialog box to define the format of the file. Configure the following format options: - **Format Type**: Select the type of the file format. You can select None, Flat, Avro, or Parquet file format. Default is None. Select the **Format Type** as **None** to read a file of binary format from Amazon S3. **Note:** When you create a mapping and if you do not click the **Formatting Options** tab, the Secure Agent considers the **Format Type** as **None** by default. - **Schema Source**: Specify the source of the schema. You can select **Read from data file** or **Import from schema file** option. - **Schema File**: You can upload a schema file. - **Delimiter**: Character used to separate columns of data. You can configure parameters such as comma, tab, colon, semicolon, or others. **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. - **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. **Note:** You can use the output text qualifier when a delimiter value is present in the data. - **Code Page**: Select the code page that the Secure Agent must use to read or write data. Amazon S3 V2 Connector supports the following code pages: - 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. - **Header Line Number**: Specify the line number that you want to use as the header when you read data from Amazon S3. You can also read a data from a file that does not have a header. Default is 1. To read data from a file that does not have a header, set the **Header Line Number** field to 0. To read data from a file that has a header, set the **Header Line Number** field to a value that is greater or equal to one. **Note:** This property is applicable during runtime and data preview. - **First Data Row**: Specify the line number of the file from where you want the Secure Agent to read data. 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 2. **Note:** This property is applicable during runtime and data preview. - **Target Header**: This property is not applicable when you read data from an Amazon S3 source. - **Distribution Column**: This property is not applicable when you read data from an Amazon S3 source.

The following table describes the Amazon S3 V2 advanced source properties that you can configure in a Source transformation:

Property	Description
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**. For more information about the source type, see "Source Types in Amazon S3 V2 Sources" on page 12.
Folder Path	Bucket name that contains the Amazon S3 source file. If applicable, include the folder name that contains the source file in the `<bucket_name>/<folder_name>` format. If you do not provide the bucket name and specify the folder path starting with a slash (/) in the `/<folder_name>` format, the folder path appends with the folder path that you specified in the connection properties. For example, if you specify the `/<dir2>` folder path in this property and `<my_bucket1>/<dir1>` folder path in the connection property, the folder path appends with the folder path that you specified in the connection properties in `<my_bucket1>/<dir1>/<dir2>` format. If you specify the `<my_bucket1>/<dir1>` folder path in the connection property and `<my_bucket2>/<dir2>` folder path in this property, the Secure Agent reads the file in the `<my_bucket2>/<dir2>` folder path that you specify in this property.
File Name	Name of the Amazon S3 resource file from which you want to read data.
Staging Directory	Amazon S3 staging directory. Ensure that the user has write permissions on the directory. In addition, ensure that there is sufficient space to enable staging of the entire file. When you specify the directory path, the Secure Agent create folders depending on the number of partitions that you specify in the following format: `InfaS3Staging<00/11><timestamp>_<partition number>` where, 00 represents read operation and 11 represents write operation. For example, `InfaS3Staging000703115851268912800_0` **Note:** The temporary files are created within the new directory. Default staging directory is the `/temp` directory on the machine that hosts the Secure Agent.
Hadoop Performance Tuning Options	This property is not applicable for Amazon S3 V2 Connector.
Compression Format	Decompresses data when you read data from Amazon S3. You can decompress the data in the following formats: - None - Gzip Default is None. **Note:** When you read a Parquet file, you can decompress the file using the none or gzip compression formats. For more information about the compression format, see "Data Compression in Amazon S3 V2 Sources and Targets" on page 18.
Download Part Size	Specifies the part size when you download an Amazon S3 object in multiple parts. Default is 5 MB. **Note:** Use this property when you run a mapping to read a file of flat format type.

Property	Description
Multiple Download Threshold	Downloads the Amazon S3 objects in multiple parts. Default is 10 MB. To download the object in multiple parts in parallel, you must ensure that the file size of an Amazon S3 object is greater than the value you specify in this property.
Tracing Level	This property is not applicable for Amazon S3 V2 Connector.

Amazon S3 V2 Targets in Mappings

In a mapping, you can configure a Target transformation to represent an Amazon S3 V2 object as the target to write data to Amazon S3.

Specify the name and description of the Amazon S3 V2 target. Configure the Amazon S3 V2 target and advanced properties for the target object.

The following table describes the Amazon S3 V2 target properties that you can configure in a Target transformation:

Property	Description
Connection	Name of the Amazon S3 V2 target connection.
Target Type	Type of the Amazon S3 V2 target connection.
Object	Name of the target object. You can select an existing object or create an object at runtime.
Create Target	Creates a target. 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 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. If you specify the path, the Secure Agent creates target object in the path you specify in this property and within the bucket that you specify in the **Folder Path** connection property. The Secure Agent creates target object in the following format: `<bucket_name>/<path_name>/<target_object_name>`. The Secure Agent only considers the bucket and ignores the path you specify in the **Folder Path** connection property. For example, specify the path as `folder1/folder2` and target object name as `Records`. Specify `<bucket_name>/folder3` as the **Folder Path** in the connection property. The Secure Agent creates the target object in the following location: `<bucket_name>/folder1/folder2/Records`. If you do not specify the path, the Secure Agent creates target object name within the bucket that you specify in the **Folder Path** connection property in the following format: `<bucket_name>/<target_object_name>`. For example, if you do not specify the path and specify the target object name as `Records`, the Secure Agent creates the target object within the bucket that you specify in the **Folder Path** connection property in the following location: `<bucket_name>/Records`. **Note:** When you write an Avro or Parquet file using the **Create Target** option, you cannot provide a Null data type.

Property	Description
Formatting Options	Amazon S3 format options. Opens the **Formatting Options** dialog box to define the format of the file. Configure the following format options: - **Format Type**: Select the type of the file format. You can select None, Flat, Avro, or Parquet file format. Default is None. Select the **Format Type** as **None** to write a file of binary format on Amazon S3. **Note:** When you create a mapping and if you do not click the **Formatting Options** tab, the Secure Agent considers the **Format Type** as **None** by default. - **Schema Source**: Specify the source of the schema. You can select **Read from data file** or **Import from schema file** option. - **Schema File**: You can upload a schema file. You cannot upload a schema file when you select the **Create Target** option. - **Delimiter**: Character used to separate columns of data. You can configure parameters such as comma, tab, colon, semicolon, or others. **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. - **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. **Note:** Use the text qualifier if you want to read data from a source that have a delimiter value present in the data to write the data to a target. Otherwise, the Secure Agent does not append the text qualifier in the target. - **Code Page**: Select the code page that the Secure Agent must use to read or write data. Amazon S3 V2 Connector supports the following code pages: - 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. - **Header Line Number**: Specify the line number that you want to use as the header when you write data to Amazon S3. You can also write data from a file that does not have a header. Default is 1. To write data to a file that does not have a header, set the **Header Line Number** field to 0. To write data to a file that has a header, set the **Header Line Number** field to a value that is greater or equal to one. **Note:** This property is applicable during data preview. - **First Data Row**: Specify the line number of the file from where you want the Secure Agent to write data. Default is 2. **Note:** This property is applicable during data preview. - **Target Header**: Select whether you want to write data to a target that contains a file with or without a header. You can select **With Header** or **Without Header** options. - **Distribution Column**: Specify the name of the column that is used to create multiple target files during run time. For more information about the distribution column, see "Distribution Column" on page 16.
Operation	Select the target operation. You can perform only insert operation on an Amazon S3 V2 target.

The following table describes the Amazon S3 V2 advanced target properties that you can configure in a Target transformation:

Property	Description
Overwrite File(s) If Exists	You can choose to overwrite the existing files. Select the check box if you want to overwrite the existing files. Default is true. For more information about overwriting the existing files, see "Overwriting Existing Files" on page 15.
Folder Path	Bucket name that contains the Amazon S3 target file. If applicable, include the folder name that contains the target file in the `<bucket_name>/<folder_name>` format. If you do not provide the bucket name and specify the folder path starting with a slash (/) in the `/<folder_name>` format, the folder path appends with the folder path that you specified in the connection properties. For example, if you specify the `/<dir2>` folder path in this property and `<my_bucket1>/<dir1>` folder path in the connection property, the folder path appends with the folder path that you specified in the connection properties in `<my_bucket1>/<dir1>/<dir2>` format. If you specify the `<my_bucket1>/<dir1>` folder path in the connection property and `<my_bucket2>/<dir2>` folder path in this property, the Secure Agent writes the file in the `<my_bucket2>/<dir2>` folder path that you specify in this property.
File Name	Name of the Amazon S3 resource file to which you want to write the source data.
Encryption Type	Method you want to use to encrypt data. Select one of the following values: - None - Client Side Encryption - Server Side Encryption - Server Side Encryption with KMS For more information, see "Data Encryption in Amazon S3 V2 Targets" on page 14.
Staging Directory	Amazon S3 V2 staging directory. Ensure that the user has write permissions on the directory. In addition, ensure that there is sufficient space to enable staging of the entire file. When you specify the directory path, the Secure Agent create folders depending on the number of partitions that you specify in the following format: `InfaS3Staging<00/11><timestamp>_<partition number>` where, 00 represents read operation and 11 represents write operation. For example, `InfaS3Staging000703115851268912800_0` **Note:** The temporary files are created within the new directory. Default staging directory is the `/temp` directory on the machine that hosts the Secure Agent.
File Merge	This property is not applicable for Amazon S3 V2 Connector.
Hadoop Performance Tuning Options	This property is not applicable for Amazon S3 V2 Connector.

Property	Description
Compression Format	Compresses data when you write data to Amazon S3. You can compress the data in the following formats: - None - Gzip Default is None. **Note:** When you write a Parquet file, you can compress the file using the none or gzip compression formats. For more information about the compression format, see "Data Compression in Amazon S3 V2 Sources and Targets" on page 18.
Object Tags	You can add single or multiple tags to the objects stored on the Amazon S3 bucket. You can either enter the key value pairs or specify the file path that contains the key value pairs. For more information about the object tags, see "Object Tag" on page 16. **Note:** Use this property when you run a mapping to write a file of flat format type.
TransferManager Thread Pool Size	Specifies the number of the threads to write data in parallel. Amazon S3 V2 Connector uses the `AWS TransferManager API` to upload a large object in multiple parts to Amazon S3. When the file size is more than 5 MB, you can configure multipart upload to upload object in multiple parts in parallel. If you set the value of the **TransferManager Thread Pool Size** to greater than 50, the value reverts to 50. Default is 10. **Note:** Use this property when you run a mapping to write a file of flat format type.
Merge Partition Files	Specifies whether the Secure Agent must merge all the partition files into a single file or maintain separate files based on the number of partitions specified to write data to the Amazon S3 V2 targets. Default is not selected.
Part Size	Specifies the part size of an object. Default is 5 MB. **Note:** Use this property when you run a mapping to write a file of flat format type.
Forward Rejected Rows	This property is not applicable for Amazon S3 V2 Connector.

Formatting Options for Avro or Parquet Files

You must set the appropriate **Formatting Options** when you select the Avro or Parquet format types.

Use the following guidelines when you select the Avro or Parquet format types and set the **Formatting Options**:

- If you select an Avro or Parquet format type and select **Read from data file** as the value of the **Schema Source** formatting option, you cannot configure the delimiter, escapeChar, and qualifier options.
- If you select an Avro 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 **Schema File** property field. You cannot configure the delimiter, escapeChar, and qualifier options.

- Set the appropriate **Formatting Options** for the Avro or Parquet format types that you select to avoid the following exception:

  ```
  invalid character encapsulated
  ```

Amazon S3 V2 Objects in Mapping Tasks

When you configure a mapping task, you can configure advanced properties for Amazon S3 V2 sources and targets.

Amazon S3 V2 Sources in Mapping Tasks

For Amazon S3 V2 source connections used in template-based mapping tasks, you can configure advanced properties in the **Sources** page of the Mapping Task wizard.

You can configure the following Amazon S3 V2 advanced properties:

Property	Description
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**. For more information about the source type, see "Source Types in Amazon S3 V2 Sources" on page 12.
Folder Path	Bucket name that contains the Amazon S3 source file. If applicable, include the folder name that contains the source file in the `<bucket_name>/<folder_name>` format. If you do not provide the bucket name and specify the folder path starting with a slash (/) in the `/<folder_name>` format, the folder path appends with the folder path that you specified in the connection properties. For example, if you specify the `/<dir2>` folder path in this property and `<my_bucket1>/<dir1>` folder path in the connection property, the folder path appends with the folder path that you specified in the connection properties in `<my_bucket1>/<dir1>/<dir2>` format. If you specify the `<my_bucket1>/<dir1>` folder path in the connection property and `<my_bucket2>/<dir2>` folder path in this property, the Secure Agent reads the file in the `<my_bucket2>/<dir2>` folder path that you specify in this property.
File Name	Name of the Amazon S3 resource file from which you want to read data.
Staging Directory	Amazon S3 staging directory. Ensure that the user has write permissions on the directory. In addition, ensure that there is sufficient space to enable staging of the entire file. When you specify the directory path, the Secure Agent create folders depending on the number of partitions that you specify in the following format: `InfaS3Staging<00/11><timestamp>_<partition number>` where, 00 represents read operation and 11 represents write operation. For example, `InfaS3Staging000703115851268912800_0` **Note:** The temporary files are created within the new directory. Default staging directory is the `/temp` directory on the machine that hosts the Secure Agent.

Property	Description
Hadoop Performance Tuning Options	This property is not applicable for Amazon S3 V2 Connector.
Compression Format	Decompresses data when you read data from Amazon S3. You can decompress the data in the following formats: - None - Gzip Default is None. **Note:** When you read a Parquet file, you can decompress the file using the none or gzip compression formats. For more information about the compression format, see "Data Compression in Amazon S3 V2 Sources and Targets" on page 18.
Download Part Size	Specifies the part size when you download an Amazon S3 object in multiple parts. Default is 5 MB. **Note:** Use this property when you run a mapping to read a file of flat format type.
Multiple Download Threshold	Downloads the Amazon S3 objects in multiple parts. Default is 10 MB. To download the object in multiple parts in parallel, you must ensure that the file size of an Amazon S3 object is greater than the value you specify in this property.
Tracing Level	This property is not applicable for Amazon S3 V2 Connector.

Amazon S3 V2 Targets in Mapping Tasks

For Amazon S3 V2 target connections used in mapping tasks, you can configure advanced properties in the **Targets** page of the Mapping Task wizard.

You can configure the following Amazon S3 V2 advanced properties:

Property	Description
Overwrite File(s) If Exists	You can choose to overwrite the existing files. Select the check box if you want to overwrite the existing files. Default is true. For more information about overwriting the existing files, see "Overwriting Existing Files" on page 15.
Folder Path	Bucket name that contains the Amazon S3 target file. If applicable, include the folder name that contains the target file in the `<bucket_name>/<folder_name>` format. If you do not provide the bucket name and specify the folder path starting with a slash (/) in the `/<folder_name>` format, the folder path appends with the folder path that you specified in the connection properties. For example, if you specify the `/<dir2>` folder path in this property and `<my_bucket1>/<dir1>` folder path in the connection property, the folder path appends with the folder path that you specified in the connection properties in `<my_bucket1>/<dir1>/<dir2>` format. If you specify the `<my_bucket1>/<dir1>` folder path in the connection property and `<my_bucket2>/<dir2>` folder path in this property, the Secure Agent writes the file in the `<my_bucket2>/<dir2>` folder path that you specify in this property.
File Name	Name of the Amazon S3 resource file to which you want to write the source data.

Property	Description
Encryption Type	Method you want to use to encrypt data. Select one of the following values: - None - Client Side Encryption - Server Side Encryption - Server Side Encryption with KMS For more information, see "Data Encryption in Amazon S3 V2 Targets" on page 14.
Staging Directory	Amazon S3 V2 staging directory. Ensure that the user has write permissions on the directory. In addition, ensure that there is sufficient space to enable staging of the entire file. When you specify the directory path, the Secure Agent create folders depending on the number of partitions that you specify in the following format: `InfaS3Staging<00/11><timestamp>_<partition number>` where, 00 represents read operation and 11 represents write operation. For example, `InfaS3Staging000703115851268912800_0` **Note:** The temporary files are created within the new directory. Default staging directory is the `/temp` directory on the machine that hosts the Secure Agent.
File Merge	This property is not applicable for Amazon S3 V2 Connector.
Hadoop Performance Tuning Options	This property is not applicable for Amazon S3 V2 Connector.
Compression Format	Compresses data when you write data to Amazon S3. You can compress the data in the following formats: - None - Gzip Default is None. **Note:** When you write a Parquet file, you can compress the file using the none or gzip compression formats. For more information the compression format, see "Data Compression in Amazon S3 V2 Sources and Targets" on page 18.
Object Tags	You can add single or multiple tags to the objects stored on the Amazon S3 bucket. You can either enter the key value pairs or specify the file path that contains the key value pairs. For more information about the object tags, see "Object Tag" on page 16. **Note:** Use this property when you run a mapping to write a file of flat format type.
TransferManager Thread Pool Size	Specifies the number of the threads to write data in parallel. Amazon S3 V2 Connector uses the `AWS TransferManager API` to upload a large object in multiple parts to Amazon S3. When the file size is more than 5 MB, you can configure multipart upload to upload object in multiple parts in parallel. If you set the value of the **TransferManager Thread Pool Size** to greater than 50, the value reverts to 50. Default is 10. **Note:** Use this property when you run a mapping to write a file of flat format type.
Merge Partition Files	Not applicable.

Property	Description
Part Size	Specifies the part size of an object. Default is 5 MB. **Note:** Use this property when you run a mapping to write a file of flat format type.
Forward Rejected Rows	This property is not applicable for Amazon S3 V2 Connector.

Specifying a Target

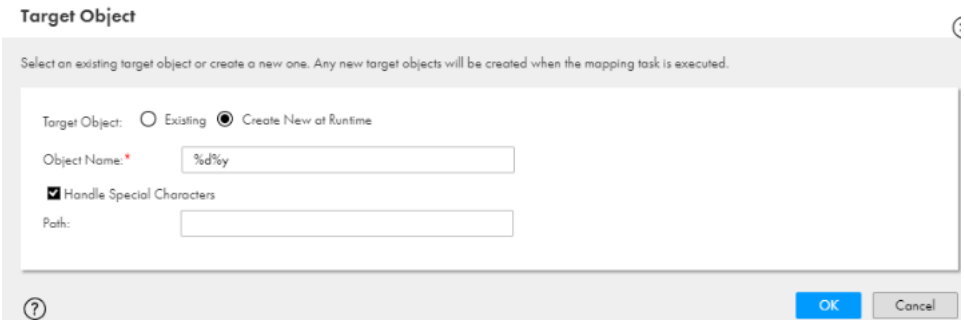
You can use an existing target or create a target to hold the results of a mapping. If you choose to create the target, the Secure Agent creates the target when you run the task.

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. You must specify a `.csv` target file name.
   - To create a target file at run time, enter the name for the target file including the extension. For example, `Accounts.csv`.
   - 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 in the following format to include all the time stamp information: `Accounts_%d%m%y%T.csv`.

     Note: If you enable **Handle Special Characters**, the Secure Agent ignores the input and output parameters in **Create Target**.
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.

   The following image shows the **Target Object** box:



9. Specify the advanced properties for the target, if needed.

Amazon S3 V2 Target File Parameterization

You can parameterize the file name and target folder location for Amazon S3 V2 target objects to pass the file name and folder location at run time. If the folder does not exist, the Secure Agent creates the folder structure dynamically.

Parameterization Using Timestamp

You can append time stamp information to the file name to show when the file is created.

When you specify a file name for the target file, include special characters based on Apache STRFTIME function formats that the mapping task uses to include time stamp information in the file name. You must enable **Handle Special Characters** options to add special characters to the file name. You can use the STRFTIME function formats in a mapping.

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.
%Y	Year including the century, for example 2015.
%T	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.

Parameterization Using a Parameter File

You can parameterize an Amazon S3 V2 target file using a parameter file.

Perform the following steps to parameterize an Amazon S3 V2 target file using a parameter file:

1. Create an Amazon S3 V2 target object.
2. Specify the values of the **Target File Name** as `$p1` and **Target Object Path** as `$p2` in the **Create Target** option.
3. Define the parameters that you added for the target object name and target object path in the parameter file.

For example,

```
$p1=filename
$p2=path
```

4. Place the parameter file in the following location:
   `<Informatica Cloud Secure Agent\apps\Data_Integration_Server\data\userparameters>`
5. Specify the parameter file name in **Schedule** tab of the mapping task.
6. Save and run the mapping task.

CHAPTER 6

Data Type Reference

This chapter includes the following topics:

Data Type Reference Overview

Data Integration uses the following data types in mass ingestion tasks, mappings, and mapping tasks with Amazon S3:

Amazon S3 native data types

Amazon S3 data types appear in the Fields tab for 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 a mapping.

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

Amazon S3 and Transformation Data Types

The following table lists the Amazon S3 data types that the Secure Agent supports and the corresponding transformation data types:

Amazon S3 Data Type	Transformation Data Type	Description
String	String	1 to 104,857,600 characters

Avro Amazon S3 File Data Types and Transformation Data Types

Avro Amazon S3 file data types map to transformation data types that the Secure Agent uses to move data across platforms.

The following table lists the Avro Amazon S3 file data types that the Secure Agent supports and the corresponding transformation data types:

Amazon S3 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

Parquet Amazon S3 File Data Types and Transformation Data Types

Parquet Amazon S3 file data types map to transformation data types that the Secure Agent uses to move data across platforms.

The following table lists the Parquet Amazon S3 file data types that the Secure Agent supports and the corresponding transformation data types:

Amazon S3 File Data Type	Transformation Data Type	Range and Description
Boolean	Integer	TRUE (1) or FALSE (0)
Double	Double	Precision 15
Float	Double	Precision 15

Amazon S3 File Data Type	Transformation Data Type	Range and Description
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
String	String	-1 to 104,857,600 characters

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 7

Troubleshooting

This chapter includes the following topics:

Troubleshooting Overview

Use the following sections to troubleshoot errors in Amazon S3 V2 Connector.

Java Heap Size Configuration

"ERROR java.lang.OutOfMemoryError: GC overhead limit exceeded." occurs when you run a Mapping task to write large number of records.

To resolve this issue, perform the following tasks to configure the JVM options in the Secure Agent to increase the memory for the Java heap size:

1. Select **Administer** > **Runtime Environments**.
2. On the **Runtime Environments** page, select the Secure Agent for which you want to increase memory 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 Service.
5. Edit the **JVMOption1** as **-Xms2048m**.

   Note: Specify the maximum and minimum heap size based on the data you want to process.
6. Restart the Secure Agent manually.

Troubleshooting for Amazon S3 V2 Connector

Informatica Cloud Data Integration Amazon S3 V2 Connector Frequently Asked Questions

For information about Amazon S3 V2 Connector frequently asked questions, see https://kb.informatica.com/h2l/HowTo%20Library/1/1207-InformaticaCDIAmazonS3V2ConnectorFAQs-H2L.pdf

How to configure AWS IAM authentication for Amazon S3 V2 Connector?

For information about configuring AWS IAM authentication, see https://kb.informatica.com/h2l/HowTo%20Library/1/1199-ConfiguringAWSIAMforAmazonS3andAmazonS3V2Connectors-H2L.pdf

INDEX