

Frequently Asked Questions for Amazon S3 V2 Connector

Abstract

You can use Amazon S3 V2 Connector to read from or write Avro and Parquet files to Amazon S3. This article lists the frequently asked questions about using Amazon S3 V2 Connector to read data from or write data to Amazon S3.

Supported Versions

- Informatica Cloud® Data Integration Amazon S3 V2 Summer 2018

Table of Contents

Overview.	2
General Questions.	2
Amazon S3 V2 Source Types.	5
Using an Existing Amazon S3 V2 Target Object.	6
Creating Amazon S3 V2 Target Objects.	6
Parameterizing an Amazon S3 V2 Target File.	8
Using Files with Header and without Header.	9

Overview

You can use Amazon S3 V2 Connector to create an Amazon S3 V2 connection to read from or write Avro and Parquet files to Amazon S3. You can use the Amazon S3 V2 connection in mass ingestion tasks, mappings, or mapping tasks.

For more information about Amazon S3 V2 Connector, see the *Informatica Cloud Data Integration Amazon S3 V2 Connector User Guide*.

General Questions

When should I use Amazon S3 V2 Connector?

If you want to run a mapping to read or write Avro and Parquet files or run a mass ingestion task, Informatica recommends that you use Amazon S3 V2 Connector.

If you want to run a mapping or a synchronization task to read or write delimited files to Amazon S3, Informatica recommends that you use Amazon S3 Connector.

Can I use Amazon S3 V2 Connector to read or write Avro and Parquet files?

Yes. You can use Amazon S3 V2 Connector to read from or write Avro and Parquet files to Amazon S3.

Note: When you use an Avro or Parquet file, ensure that one Cloudera, Hortonworks or Amazon EMR license is available in the Secure Agent to successfully fetch the metadata.

What are the tasks that Amazon S3 V2 Connector supports?

You can use Amazon S3 V2 Connector to run the following tasks:

Mass Ingestion Task

You can 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.

Mapping Task

You can create a mapping task to process data based on the data flow logic defined in a mapping or integration template.

Can I configure the AWS Identity and Access Management (IAM) authentication when I use Amazon S3 V2 Connector?

Yes. You can configure the IAM authentication when the Secure Agent runs on an Amazon Elastic Compute Cloud (EC2) system.

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

Can I read a client-side encrypted file in an Amazon S3 bucket?

Yes. You can read a client-side encrypted file in an Amazon S3 bucket. You must provide a master symmetric key or customer master key in the connection properties.

Does Amazon S3 V2 Connector support client-side and server-side encryptions for Amazon S3 V2 targets?

Yes. You can enable the client-side and server-side encryptions for Amazon S3 V2 targets if you want to encrypt the data while uploading the files to the buckets.

The client-side encryption uses a master symmetric key or AWS KMS-managed customer master key to encrypt data. The server-side encryption uses an Amazon S3-managed encryption key (SSE-S3) or AWS KMS-managed customer master key (SSE-KMS) to encrypt data.

The following table lists the encryption types that Amazon S3 V2 Connector supports:

Encryption Type	Flat File	Avro File	Parquet File
Client-side Encryption	Yes	No	No
Server-side Encryption (SSE-S3)	Yes	No	No
Server-side Encryption (SSE-KMS)	Yes	No	No

Can I use Amazon S3 V2 Connector to compress or decompress data?

Yes. You can use Amazon S3 V2 Connector to decompress data when you read data from Amazon S3 and compress the data when you write data to Amazon S3.

The following table lists the supported compression formats:

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

Is there any difference in the behavior when I run a mapping directly compared to when I run a mapping task that uses the mapping?

No. There is no difference in the behavior when you run a mapping directly compared to when you run a mapping task that uses the mapping.

What are the data types that Amazon S3 V2 Connector supports for flat, Avro, and Parquet files?

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

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

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

Avro	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

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

Parquet	Transformation Data Type	Range and Description
Boolean	Integer	TRUE (1) or FALSE (0)
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

Parquet	Transformation Data Type	Range and Description
Int96	Binary	12-byte signed integer
String	String	-1 to 104,857,600 characters

Amazon S3 V2 Source Types

What are the various source types that Amazon S3 V2 Connector supports?

Amazon S3 V2 Connector supports the following type of sources from where you can read data:

- File
- Directory

What happens if I provide a bucket name and folder path in the connection property and specify directory as the source type in the advanced source property?

If you provide a bucket name and folder path in the `<bucket_name>/<path_name>` format in the connection property and specify directory as the source type in the advanced source property, the Secure Agent reads the file from the bucket name and folder path that you specified in the connection property.

What happens if I specify only a bucket name in the connection property and directory as the source type in the advanced source property?

If you specify only a bucket name in the `<bucket_name>` format in the connection property and directory as the source type in the advanced source property, the Secure Agent reads all the files that are present inside the bucket that you specified in the connection property.

What happens if I specify different folder paths in the connection and advanced source properties when I specify directory as the source type?

If you specify different folder paths in the connection and advanced source properties, the Secure Agent reads the file from the folder path that you specified in the advanced source property. The Secure Agent ignores the folder path that you specified in the connection property.

For example, if you specify the `<my_bucket1>/<dir1>` folder path in the connection property and `<my_bucket2>/<dir2>` folder path in the advanced source property, the Secure Agent reads the file from the `<my_bucket2>/<dir2>` folder path that you specified in the advanced source property.

What happens if I specify a folder path name starting with a slash (/) in the advanced source property when I specify directory as the source type?

If you specify a folder path starting with a slash (/) in the `/<folder_name>` format in the advanced source property, the Secure Agent appends the folder path to the folder path that you specified in the connection properties.

For example, if you specify the `<my_bucket1>/<dir1>` folder path in the connection property and `/<dir2>` folder path in the advanced source property, the Secure Agent appends the `/<dir2>` folder path to the `<my_bucket1>/<dir1>` folder path that you specified in the connection properties in the following format:
`<my_bucket1>/<dir1>/<dir2>`

Using an Existing Amazon S3 V2 Target Object

What happens if I select an existing target object and specify a folder path starting with a slash (/) in the advanced target property?

If you select an existing target object and specify a folder path starting with a slash (/) in the `/<folder_name>` format in the advanced target property, the Secure Agent appends the folder path in the advanced target property to the folder path in the connection property.

For example, if you specify the `<my_bucket1>/<dir1>` folder path in the connection property and `/<dir2>` folder path in the advanced target property, the Secure Agent appends the `/<dir2>` folder path to the `<my_bucket1>/<dir1>` folder path and writes the data to the following location: `<my_bucket1>/<dir1>/<dir2>/<existing_target_object>`

What happens if I select an existing target object and specify different bucket names in the connection and advanced target properties?

If you specify different bucket names in the connection and advanced target properties when you select an existing target object, the Secure Agent considers the bucket name that you specified in the advanced target properties. The Secure Agent ignores the bucket name that you specified in the connection property.

For example, if you specify the `<my_bucket1>` bucket name in the connection property and `<my_bucket2>` bucket name in the advanced target property, the Secure Agent writes the file to the following location: `<my_bucket2>/<existing_target_object>`

What happens if I specify a folder path starting with a slash (/) in advanced target property that is not present in the bucket that I specified in the connection property when I select an existing target object?

The Secure Agent creates a new folder and appends the new folder to the folder path that you specified in the connection property. The Secure Agent creates the new folder in the following location: `<bucket_name>/<folder_name1>/<new_folder_name2>`

For example, if you specify the `<my_bucket1>/<dir1>` folder path in the connection property and `/<dir2>` folder path in the advanced target property, the Secure Agent creates and appends the `/<dir2>` new folder with the `<my_bucket1>/<dir1>` folder path that you specified in the connection property. The Secure Agent writes the file to the following location: `<my_bucket1>/<dir1>/<dir2>/<existing_target_object>`

Creating Amazon S3 V2 Target Objects

Can I create a new Amazon S3 V2 target object?

Yes. You can create a new Amazon S3 V2 target object using the **Create Target** option at runtime. You can also provide a path for the target object.

Can I specify a Null data type when I create a new Amazon S3 V2 target object to write an Avro or Parquet file?

No. You cannot specify a Null data type when you create a new Amazon S3 V2 target object to write an Avro or Parquet file.

When I use the Create Target option and run a mapping, the Secure Agent creates a new target file. What happens to the new target file if I rerun the mapping?

When you rerun the mapping, the Secure Agent replaces the target file each time you run the mapping or the mapping task.

What happens if I do not specify a bucket name and specify only a path in the Create Target option?

The Secure Agent creates the new target object in the path that you specified in the **Create Target** option and within the bucket that you specified in the connection property. The Secure Agent creates the new target object in the following location: `<bucket_name>/<path_name>/<new_target_object_name>`.

For example, if you specify the <bucket_name> bucket name in the connection property, new target object name as `Records`, and path as <dir1>/<dir2> in the **Create Target** option, the Secure Agent creates the new target object in the following location: <bucket_name>/<dir1>/<dir2>/`Records`

What happens if I specify only a path in the Create Target option that is not present in the bucket that I specify in the connection property?

The Secure Agent creates a new folder within the bucket that you specified in the connection property and creates the target object in the following location: <con_bucket_name>/<new_folder_name>/<new_target_object_name>

The Secure Agent considers the bucket name and ignores the path that you specified in the connection property.

For example, if you specify the <bucket_name>/<dir1> folder path in the connection property, new target object name as `Records`, and the path as <dir2> in the **Create Target** option, the Secure Agent creates the <dir2> new folder within the bucket that you specified in the connection property. The Secure Agent creates the new target object in the following location: <bucket_name>/<dir2>/`Records`

What happens if I specify only a bucket name in the connection property and do not specify a path in the Create Target option?

The Secure Agent creates the new target object inside the bucket that you specified in the connection property in the following location: <con_bucket_name>/<new_target_object_name>

For example, if you specify the <bucket_name> bucket name in the connection property and the new target object name as `Records` in the **Create Target** option, the Secure Agent creates the target object in the following location: <bucket_name>/`Records`

What happens if I specify only a bucket name in the connection property and specify a path in the advanced target property?

The Secure Agent creates the new target object in the path that you specified in the advanced target property within the bucket that you specified in the connection property in the following location: <con_bucket_name>/<advanced_property_path_name>/<new_target_object_name>

For example, if you specify the <bucket_name> bucket name in the connection property, path as <dir1>/<dir2> in the advanced target property, and new target object name as `Records` in the **Create Target** option, the Secure Agent creates the new target object in the following location: <bucket_name>/<dir1>/<dir2>/`Records`

What happens if I specify a bucket name and folder path in the connection property and specify a path in the Create Target option?

The Secure Agent creates the new target object in the path that you specified in the **Create Target** option within the bucket that you specified in the connection property. The Secure Agent ignores the path that you specified in the connection property and creates the new target object in the following location: <con_bucket_name>/<Create_Target_path_name>/<new_target_object_name>

For example, if you specify the <bucket_name>/<dir1> bucket name in the connection property, target object name as `Records`, and path as <dir2>/<dir3> in the **Create Target** option, the Secure Agent creates the new target object in the following location: <con_bucket_name>/<dir2>/<dir3>/`Records`

What happens if I specify a bucket name and folder path in the connection property and do not specify a path in the Create Target option?

The Secure Agent creates the new target object within the bucket that you specified in the connection property and creates the new target object in the following location: <con_bucket_name>/<new_target_object_name>

For example, if you specify the `<bucket_name>/<dir1>` bucket name in the connection property, new target object name as `Records` in the **Create Target** option, the Secure Agent creates the new target object in the following location: `<bucket_name>/Records`

What happens if I specify a bucket name and folder path in the connection and advanced target properties when I create a new target object?

The Secure Agent creates the new target object in the path within the bucket that you specified in the advanced target property and creates the new target object in the following location: `<bucket_name>/<path_name>/<new_target_object_name>`

The Secure Agent ignores the bucket name and folder path specified in the connection property.

For example, if you specify the `<bucket_name1>/<dir1>` bucket name in the connection property, the `<bucket_name2>/<dir2>` bucket name in the advanced target property, and new target object name as `Records` in the **Create Target** option, the Secure Agent creates the new target object in the following location: `<bucket_name2>/<dir2>/Records`

Parameterizing an Amazon S3 V2 Target File

Can I parameterize the file name and path for Amazon S3 target files?

Yes. You can parameterize the file name and path for the Amazon S3 target files at run time.

You can parameterize the file name and path for a newly created target file using time stamp. You can parameterize the file name of an existing target file or a new target file using a parameter file.

How can I parameterize an Amazon S3 target file name using time stamp?

When you configure the Amazon S3 V2 target properties in a mapping and select an object, set the following properties in the **Target Object** dialog box:

1. In the **Target Object** field, select the **Create Target** option to create a new Amazon S3 target file.
2. In the **Object Name** field, enter the object name for the new target file and add special characters based on Apache STRFTIME function formats to include time stamp information in the file name.
3. Select the **Handle Special Characters** check box.
4. Click **OK**.

The Secure Agents creates a new target file with the specified file name and appends the time stamp information to the file name to show the file creation time when you run the mapping task.

For example, if you enter the object name as `File_%d_%y.txt` in the **Object Name** field, the Secure Agent creates a new target file with the `File_21_18.txt` file name.

How can I parameterize an Amazon S3 target file using a parameter file?

To parameterize an Amazon S3 target file using a parameter file, perform the following steps:

1. When you configure the Amazon S3 V2 target properties in a mapping and select an object, set the following properties in the **Target Object** dialog box:
 - a. In the **Target Object** field, select an existing Amazon S3 file or create a new Amazon S3 target file.
 - b. In the **Object Name** field, enter `$p1` as the object name for the target file.
 - c. In the **Path** field, enter `$p2` as the path if you create a new target file.

Note: This property is applicable only when you select the **Create Target** option in the **Target Object** field to create a new Amazon S3 target file.

- d. Click **OK**.

2. Define the parameters that you added for the file name and path in the parameter file.
For example, \$p1=filename and \$p2=path.
3. Place the parameter file in the following location:
<Informatica Cloud Secure Agent\apps\Data_Integration_Server\data\userparameters>
4. In **Schedule** tab of the mapping task, specify the parameter file name in the **Parameter File Name** field.
5. Save and run the mapping task.

For an existing file, the Secure Agent overwrites the existing target file name with the parameterized file name when you run the mapping task. For a newly created target file, the Secure Agent creates a new target file with the parameterized file name in the specified path when you run the mapping task.

For example, you want to parameterize an existing target file name and you define the parameter as \$p1=Amazon_S3 for the object name in the parameter file. The Secure Agent overwrites the existing target file name to Amazon_S3.

You want to parameterize a newly created target file and you define the parameter as \$p1=Amazon_S3 and \$p2=Target_path for the object name and path in the parameter file. The Secure Agents creates a new target file with the Amazon_S3 file name in the Target_path path.

Using Files with Header and without Header

Can I read or write an Amazon S3 file that does not contain a header?

Yes. You can read or write an Amazon S3 file that does not contain a header.

How can I read data from an Amazon S3 file that does not contain a header and write the data to an Amazon S3 file with a header?

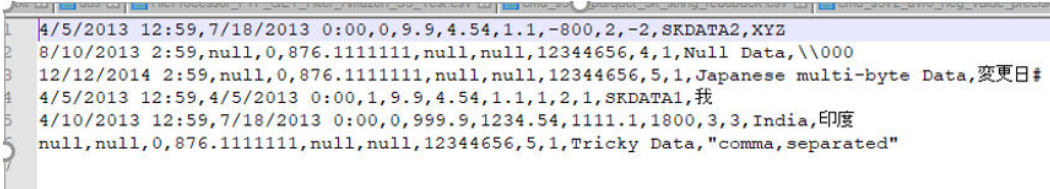
Set the following properties in the **Formatting Options** tab when you configure the Amazon S3 V2 source and target properties in a mapping:

1. In the **Header Line Number** source property, enter **0** as the line number.
2. In the **First Data Row** source property, enter **1** as the line number.
3. In the **Target Header** target property, select the **With Header** option.

When you run the mapping task, the Secure Agent reads data from the file that does not contain a header and writes the data to the file adding a header.

Note: If you select the **Create Target** option, the Secure Agent reads data from the file that does not contain a header and generates a file with a header in the target.

For example, you want to read data from the following Amazon S3 file that does not contain a header:



```

1 4/5/2013 12:59,7/18/2013 0:00,0,9.9,4.54,1.1,-800,2,-2,SKDATA2,XYZ
2 8/10/2013 2:59,null,0,876.1111111,null,null,12344656,4,1,Null Data,\\000
3 12/12/2014 2:59,null,0,876.1111111,null,null,12344656,5,1,Japanese multi-byte Data,変更日#
4 4/5/2013 12:59,4/5/2013 0:00,1,9.9,4.54,1.1,1,2,1,SKDATA1,我
5 4/10/2013 12:59,7/18/2013 0:00,0,999.9,1234.54,1111.1,1800,3,3,India,印度
6 null,null,0,876.1111111,null,null,12344656,5,1,Tricky Data,"comma,separated"

```

After you set the required properties in the **Formatting Options** tab, the Secure Agent displays the data in the **Data Preview** tab in the following format:

Data Preview (X)

Connection: s3_v2_win_new		Object: info.qa.bucket/headerless files//headerless.csv		
FIELD0	FIELD1	FIELD2	FIELD3	FIELD4
4/5/2013 12:59	7/18/2013 0:00	0	9.9	4.54
8/10/2013 2:59		0	876.1111111	
12/12/2014 2:59		0	876.1111111	
4/5/2013 12:59	4/5/2013 0:00	1	9.9	4.54

When you run the mapping task, the Secure Agent reads data from the file that does not contain a header and writes the data to the file adding a header in the following format:

```
FIELD0, FIELD1, FIELD2, FIELD3, FIELD4, FIELD5
4/5/2013 12:59, 7/18/23 0:00, 0, 9.9, 4.54
8/10/2013 2:59, null, 0, 876.1111111, null
12/12/2014 2:59, null, 0, 876.1111111, null
```

How can I read data from an Amazon S3 file that does not contain a header and write data to an Amazon S3 file without a header?

Set the following properties in the **Formatting Options** tab when you configure the Amazon S3 V2 source and target properties in a mapping:

1. In the **Header Line Number** source property, enter **0** as the line number.
2. In the **First Data Row** source property, enter **1** as the line number.
3. In the **Target Header** target property, select the **Without Header** option.

When you run the mapping task, the Secure Agent reads data from the file that does not contain a header and writes data to the file without a header.

Note: If you select the **Create Target** option, the Secure Agent reads data from the file that does not contain a header and generates a file without a header in the target.

For example, the Secure Agent reads data from the file that does not contain a header and writes the data to the file without a header the following format when you run the mapping task:

```
4/5/2013 12:59, 7/18/23 0:00, 0, 9.9, 4.54, 1.1
8/10/2013 2:59, null, 0, 876.1111111, null
12/12/2014 2:59, null, 0, 876.1111111, null
```

How can I read data from an Amazon S3 file that contains a header and write data to an Amazon S3 file without a header?

Set the following properties in the **Formatting Options** tab when you configure the Amazon S3 V2 source properties in a mapping:

1. In the **Header Line Number** property, enter the line number that you want to use as the header when you read data from Amazon S3.
2. In the **First Data Row** property, enter the line number of the Amazon S3 file from where you want the Secure Agent to read data.
If you want to read data from the header, the value of the **Header Line Number** and the **First Data Row** properties should be the same.

3. In the **Target Header** target property, select the **Without Header** option.

When you run the mapping task, the Secure Agent reads data from the file that contains a header and writes data to the file without a header.

Note: If you select the **Create Target** option, the Secure Agent reads data from the file that contains a header and generates a file without a header in the target.

For example, you want to read data from the following Amazon S3 file that have a header:

```
#comments
comments
"f_varchar","f_char","f_smallint","f_integer","f_bigint"
4/5/2013 12:59,7/18/23 0:00,0,9.9,4.54,1.1
8/10/2013 2:59,null,0,876.1111111,null
12/12/2014 2:59,null,0,876.1111111,null
```

Set the **Header Line Number** property to **3** and **First Data Row** property to **4** in the **Formatting Options** tab. The Secure Agent displays the data of an Amazon S3 file that have a header in the **Data Preview** tab in the following format:

Data Preview			
f_varchar	f_char	f_smallint	f_integer
我	SKDATA1	1	2
The two letters are unrelated...	SKDATA2	-2	2
印度	India	3	3

The Secure Agent reads data from the file that contains a header and writes the data to the file without a header in the following format when you run the mapping task:

```
4/5/2013 12:59,7/18/23 0:00,0,9.9,4.54,1.1
8/10/2013 2:59,null,0,876.1111111,null
12/12/2014 2:59,null,0,876.1111111,null
```

How can I read data from an Amazon S3 file that contains a header and write data to an Amazon S3 file with a header?

Set the following properties in the **Formatting Options** tab when you configure the Amazon S3 V2 source properties in a mapping:

1. In the **Header Line Number** property, enter the line number that you want to use as the header when you read data from Amazon S3.
2. In the **First Data Row** property, enter the line number of the Amazon S3 file from where you want the Secure Agent to read data.
3. In the **Target Header** target property, select the **With Header** option.

When you run the mapping task, the Secure Agent reads data from the file that contains a header and writes data to the file with a header.

Note: If you select the **Create Target** option, the Secure Agent reads data from the file that contains a header and generates a file with a header in the target.

For example, you want to read data from the following Amazon S3 file that have a header:

```
#comments
comments
"f_varchar","f_char","f_smallint","f_integer","f_bigint"
4/5/2013 12:59,7/18/23 0:00,0,9.9,4.54,1.1
8/10/2013 2:59,null,0,876.1111111,null
12/12/2014 2:59,null,0,876.1111111,null
```

Set the **Header Line Number** property to **3** and **First Data Row** property to **4** in the **Formatting Options** tab.

The Secure Agent reads data from the file that contains a header and writes the data to the file with a header in the following format when you run the mapping task:

```
f_varchar,f_char,f_smallint,f_integer,f_bigint
4/5/2013 12:59,7/18/23 0:00,0,9.9,4.54,1.1
8/10/2013 2:59,null,0,876.1111111,null
12/12/2014 2:59,null,0,876.1111111,null
```

Author

Salam Subhashree

Acknowledgements

The author would like to acknowledge Aditi Tellar Nayak, for her technical assistance with this article.