Informatica

Informatica® Cloud Data Integration

Amazon S3 V2 Connector

Informatica Cloud Data Integration Amazon S3 V2 Connector
February 2021

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

NOTICES

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

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

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

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

Publication Date: 2021-02-17

Table of Contents

Preface 5

Informatica Resources. 5

Informatica Documentation. 5

Informatica Intelligent Cloud Services web site. 5

Informatica Intelligent Cloud Services Communities. 5

Informatica Intelligent Cloud Services Marketplace. 5

Data Integration connector documentation. 6

Informatica Knowledge Base. 6

Informatica Intelligent Cloud Services Trust Center. 6

Informatica Global Customer Support. 6

Chapter 1: Introduction to Amazon S3 V2 Connector. 7

Amazon S3 V2 Connector overview. 7

Amazon S3 V2 supported task types and object types. 8

Introduction to Amazon S3 8

Authentication methods. 8

Administration of Amazon S3 V2 Connector. 8

Create minimal Amazon S3 bucket policy. 9

IAM authentication. 9

Temporary security credentials using AssumeRole 10

Credential profile file authentication. 13

Chapter 2: Amazon S3 V2 connections. 15

Amazon S3 V2 connections overview. 15

Amazon S3 V2 connection properties. 16

Rules and guidelines for AWS regions. 19

Configuring proxy settings. 19

Configuring proxy settings on Windows. 20

Configuring proxy settings on Linux. 21

Chapter 3: Amazon S3 V2 sources and targets. 22

Amazon S3 V2 sources. 22

Data encryption in Amazon S3 V2 sources 23

Source types in Amazon S3 V2 sources. 24

Working with multiple files. 25

Source partitioning. 26

Reading source objects path. 26

Pushdown optimization. 27

Amazon S3 V2 targets. 27

Data encryption in Amazon S3 V2 targets. 28

Overwriting existing files. 29
Target partitioning. 30
Distribution column. 30
Writing to multiple target objects. 31
Object tag. 31
Data compression in Amazon S3 V2 sources and targets. 33
Reading a compressed flat file. 34

Chapter 4: Mappings and mapping tasks with Amazon S3 V2. 35
Amazon S3 V2 objects in mappings. 35
Amazon S3 V2 source in mappings. 35
Amazon S3 V2 targets in mappings. 38
Amazon S3 V2 lookups. 42
Amazon S3 V2 file formatting options. 42
Rules and guidelines for setting formatting options. 44
Rules and guidelines for creating an elastic mapping. 44
Amazon S3 V2 objects in mapping tasks. 45
Amazon S3 V2 sources in mapping tasks. 45
Amazon S3 V2 targets in mapping tasks. 47
Specifying a target. 50
Amazon S3 V2 target file parameterization. 50
Parameterization using timestamp 50
Parameterization using a parameter file. 51
Directory-level partitioning. 52
Rules and guidelines for reading from and writing to a partition folder. 53
Elastic mapping example. 54

Chapter 5: Data type reference. 56
Data type reference overview. 56
Flat file data types and transformation data types. 57
Avro Amazon S3 file data types and transformation data types. 57
Enabling Date, Decimal, and Timestamp types. 59
JSON Amazon S3 file data types and transformation data types. 59
ORC Amazon S3 file data types and transformation data types. 60
Parquet Amazon S3 file data types and transformation data types. 61

Chapter 6: Troubleshooting. 63
Troubleshooting overview. 63
Troubleshooting for Amazon S3 V2 Connector. 63
Troubleshooting FAQ. 64

Index. 66

Preface

Use *Amazon S3 V2 Connector* to learn how to read from or write to Amazon S3 by using Cloud Data Integration. Learn to create an Amazon S3 V2 connection, develop and run mass ingestion tasks, mapping tasks, mappings, and elastic mappings in Cloud Data Integration.

Informatica Resources

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

Informatica Documentation

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

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

Informatica Intelligent Cloud Services web site

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

Informatica Intelligent Cloud Services Communities

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

Access the Informatica Intelligent Cloud Services Community at:

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

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

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

Informatica Intelligent Cloud Services Marketplace

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

https://marketplace.informatica.com/

Data Integration connector documentation

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

Informatica Knowledge Base

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

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

Informatica Intelligent Cloud Services Trust Center

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

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

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

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

Informatica Global Customer Support

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

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

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

CHAPTER 1

Introduction to Amazon S3 V2 Connector

This chapter includes the following topics:

- Amazon S3 V2 Connector overview, 7
- Amazon S3 V2 supported task types and object types, 8
- Introduction to Amazon S3 , 8
- Authentication methods, 8
- Administration of Amazon S3 V2 Connector, 8

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, flat, binary, ORC, and Parquet file formats for mappings and Avro, flat, ORC, Parquet and JSON for elastic mappings in Amazon S3.

You can create an Amazon S3 V2 connection and use the connection in mappings or mapping tasks. You can also use the Amazon S3 V2 connection in elastic mappings. You can read and write primitive data types for Avro, Parquet, JSON and ORC files. You can read and write hierarchical data types only for Avro, Parquet and JSON in elastic mappings. For more information about elastic mappings, see *Administrator* and *Mappings*.

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

You cannot use Amazon S3 V2 Connector to read or write Avro, JSON, ORC, and Parquet files on Windows.

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
Mapping	Yes	Yes
Elastic mapping	Yes	Yes

Introduction to Amazon S3

Amazon Simple 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 nest folders.

Authentication methods

Amazon S3 V2 Connector supports the following authentication methods:

- **Basic authentication**: You can configure the basic authentication by providing the access key and secret key values.
- **IAM authentication**: You can configure IAM authentication when the Secure Agent runs on an Amazon Elastic Compute Cloud (EC2) system.
- **Temporary security credentials via AssumeRole**: You can configure the temporary security credentials using **AssumeRole** to access the AWS resources from the same or different AWS accounts.
- **Credential profile file authentication**: You can access the Amazon S3 credentials from a credential file that contains the access key, secret key, and the session token.
- **Federated user single sign-on**: You can configure federated user single sign-on to securely control access to the Amazon S3 resources.

Administration of Amazon S3 V2 Connector

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

- Create a minimal Amazon S3 bucket policy for Amazon S3 V2 Connector.

- To run a mapping that reads from or writes to a complex file, ensure that either Cloudera 5.8, Cloudera 6.1, Hortonworks 2.5, or Hortonworks 2.6 license is enabled.
- Ensure that Cloudera 5.8 or Cloudera 6.1 license is enabled to preview the data successfully when you create an elastic mapping that reads from or writes to an Avro, JSON, ORC, or Parquet file.

Prerequisites for client-side and server-side encryption

- Create a master symmetric key if you want to enable client-side encryption.
- Create an AWS Key Management Service (AWS KMS) policy and an AWS KMS-managed customer master key if you want to enable the encryption with KMS.

Prerequisites for Informatica encryption

- To use the Informatica Encryption method to encrypt or decrypt data of a binary or flat file, ensure that the Informatica crypto library license is enabled.

Prerequisites for temporary security credentials via AssumeRole

- Ensure that you have the **sts:AssumeRole** permission and a trust relationship established within the AWS accounts to use the temporary security credentials to access the AWS resources.
- Create the temporary security credentials policy to use the temporary security credentials to access the AWS resources.

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.

In elastic mappings, you can use different AWS accounts within the same AWS region. Make sure that the Amazon S3 bucket policy confirms access to the AWS accounts used in elastic mappings.

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

Optionally, if you do not provide the access key and the secret key in the connection, Amazon S3 V2 Connector uses AWS credentials provider chain that looks for credentials in the following order:

1. The **AWS_ACCESS_KEY_ID** and **AWS_SECRET_ACCESS_KEY** or **AWS_ACCESS_KEY** and **AWS_SECRET_KEY** environment variables.

2. The **aws.accessKeyId** and **aws.secretKey** java system properties.
3. The credential profiles file at the default location, `~/.aws/credentials`.
4. The instance profile credentials delivered through the Amazon EC2 metadata service.

You can configure IAM authentication when the Secure Agent runs on an Amazon Elastic Compute Cloud (EC2) system. When you use a serverless runtime environment, you cannot configure IAM authentication.

Perform the following steps to configure IAM authentication on EC2:

1. Create minimal Amazon S3 bucket policy.
2. Create the Amazon EC2 role. The Amazon EC2 role is used when you create an EC2 system. For more information about creating the Amazon EC2 role, see the AWS documentation.
3. Link the minimal Amazon S3 bucket policy with the Amazon EC2 role.
4. Create an EC2 instance. Assign the Amazon EC2 role that you created in step #2 to the EC2 instance.
5. Install the Secure Agent on the EC2 system.

Use IAM authentication for secure and controlled access to Amazon S3 resources when you run a session.

Temporary security credentials using AssumeRole

You can use the temporary security credentials using AssumeRole to access the AWS resources from the same or different AWS accounts.

Ensure that you have the **sts:AssumeRole** permission and a trust relationship established within the AWS accounts to use the temporary security credentials. The trust relationship is defined in the trust policy of the IAM role when you create the role. The IAM role adds the IAM user as a trusted entity allowing the IAM users to use the temporary security credentials and access the AWS accounts. For more information about how to establish the trust relationship, see the AWS documentation.

When the trusted IAM user requests for the temporary security credentials, the AWS Security Token Service (AWS STS) dynamically generates the temporary security credentials that are valid for a specified period and provides the credentials to the trusted IAM users. The temporary security credentials consist of access key ID, secret access key, and secret token.

To use the dynamically generated temporary security credentials, provide the value of the **IAM Role ARN** connection property when you create an Amazon S3 V2 connection. The IAM Role ARN uniquely identifies the AWS resources. Then, specify the time duration in seconds during which you can use the temporarily security credentials in the **Temporary Credential Duration** advanced source and target properties.

External ID

You can specify the external ID for a more secure cross-account access to the Amazon S3 bucket when the Amazon S3 bucket is in a different AWS account.

You can optionally specify the external ID in the AssumeRole request to the AWS Security Token Service (STS).

The external ID must be a string.

The following sample shows an external ID condition in the assumed IAM role's trust policy:

```
"Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::AWS_Account_ID : user/user_name"
      },
      "Action": "sts:AssumeRole",
        "Condition": {
          "StringEquals": {
            "sts:ExternalId": "dummy_external_id"
          }
        }
      }
    ]
```

Temporary security credentials policy

To use the temporary security credentials to access the AWS resources, both the IAM user and IAM role require policies.

The following section lists the policies required for the IAM user and IAM role:

IAM user

An IAM user must have the `sts:AssumeRole` policy to use the temporary security credentials in the same or different AWS account.

The following sample policy allows an IAM user to use the temporary security credentials in an AWS account:

```
{
"Version":"2012-10-17", "Statement":{ "Effect":"Allow", "Action":"sts:AssumeRole",
"Resource":"arn:aws:iam::<ACCOUNT-HYPHENS>:role/<ROLE-NAME>" }
}
```

IAM role

An IAM role must have a `sts:AssumeRole` policy and a trust policy attached with the IAM role to allow the IAM user to access the AWS resource using the temporary security credentials. The policy specifies the AWS resource that the IAM user can access and the actions that the IAM user can perform. The trust policy specifies the IAM user from the AWS account that can access the AWS resource.

The following policy is a sample trust policy:

```
{
"Version":"2012-10-17", "Statement":[{ "Effect":"Allow", "Principal":
{ "AWS":"arn:aws:iam::AWS-account-ID:root" },
"Action":"sts:AssumeRole" }
 ]
  }
   }
```

Here, in the `Principal` attribute, you can also provide the ARN of IAM user who can use the dynamically generated temporary security credentials and to restrict further access. For example,

```
"Principal" : { "AWS" : "arn:aws:iam:: AWS-account-ID :user/ user-name " }
```

To use the temporary security credentials with AWS Key Management Service (AWS KMS)-managed customer master key and enable the encryption with KMS, you must create a KMS policy.

You can perform the following operations to use the temporary security credentials and enable the encryption with KMS:

- GenerateDataKey
- DescribeKey
- Encrypt
- Decrypt
- ReEncrypt

Sample policy:

```
{
"Version":"2012-10-17", "Statement":[{ "Effect":"Allow", "Action":
[ "kms:GenerateDataKey", "kms:DescribeKey", "kms:Encrypt", "kms:Decrypt",
"kms:ReEncrypt*" ], "Resource":"*"
    }
  ]
   }
```

Temporary security credentials using AssumeRole for EC2

You can use temporary security credentials using AssumeRole for an Amazon EC2 role to access the AWS resources from the same or different AWS accounts.

The Amazon EC2 role would be able to assume another IAM Role from the same or different AWS account without requiring a permanent access key and secret key. The Amazon EC2 role can also assume another IAM role from a different region.

Consider the following prerequisites when you use temporary security credentials using AssumeRole for EC2:

- To use temporary security credentials using AssumeRole for EC2, install the Secure Agent on an AWS service such as Amazon EC2.
- The EC2 role attached to the AWS EC2 service must not have access to Amazon S3 but needs to have permission to assume another IAM role.
- The IAM role that needs to be assumed by the EC2 role must have a permission policy and a trust policy attached to it.
  For more information about the policies for EC2 roles and IAM roles, see "Temporary security credentials policy" on page 11.

To configure an EC2 role to assume the IAM role provided in the **IAM Role ARN** connection property, select the **Use EC2 Role to Assume Role** check box in the connection properties.

Rules and guidelines for using the temporary security credentials

Consider the following guidelines when you use the temporary security credentials:

- The IAM user or IAM role that requests for the temporary security credentials must not have access to any AWS resources.
- Only authenticated IAM users or IAM roles can request for the temporary security credentials from the AWS Security Token Service (AWS STS).
- Before you run a task, ensure that you have enough time to use the temporary security credentials for running the task. You cannot extend the time duration of the temporary security credentials for an ongoing task. For example, when you read from and write to Amazon Redshift and if the temporary security credentials expire, you cannot extend the time duration of the temporary security credentials that causes the task to fail.
- After the temporary security credentials expire, AWS does not authorize the IAM users or IAM roles to access the resources using the credentials. You must request for new temporary security credentials before the previous temporary security credentials expire in a mapping.
- For elastic mappings, the temporary security credentials do not expire even after the configured time in the **Temporary Credential Duration** advanced source property elapses.
- Do not use the root user credentials of an AWS account to use the temporary security credentials. You must use the credentials of an IAM user to use the temporary security credentials.
- Using temporary security credentials to read data from or write data to a complex file such as Avro, ORC, or Parquet file depends on the Hadoop distribution in your environment. Currently, Amazon S3 V2

Connector supports temporary security credentials for complex files on Cloudera 6.1, Hortonworks 2.5, and Hortonworks 2.6 distributions. However, to read data from or write data to a flat file using the temporary security credentials, no Hadoop distribution is required by Amazon S3 V2 Connector.

- If both the source and target in a mapping point to the same Amazon S3 bucket, use the same Amazon S3 connection in the Source and Target transformations. If you use two different Amazon S3 connections, configure the same values in the connection properties for both the connections.
- If the source and target in a mapping point to different Amazon S3 buckets, you can use two different Amazon S3 connections.
  You can configure different values in the connection properties for both the connections. However, you must select the **Use EC2 Role to Assume Role** check box in the connection property. You must also specify the same value for the **Temporary Credential Duration** field in the source and target properties.
- In a mapping, if you configure two or more Amazon S3 data sources for the same Amazon S3 bucket with different IAM roles, either of the IAM roles must be able to access the other data source as well.
- In a mapping, if you configure one Amazon S3 data source with user credentials and the other Amazon S3 data source with an IAM role, consider the following rules:
  - The user credentials for the first data source must also be able to assume the IAM role of the second Amazon S3 data source.
  - The IAM role that you configured for the second data source must also have access to the first Amazon S3 data source.

Credential profile file authentication

You can provide the credentials required to establish the connection with Amazon S3 through the credential profile file that contains an access key and secret key. The credential profile file contains an access key, a secret key, and a session token when you use temporary security credentials.

You can use permanent IAM credentials or temporary security credentials with a session token when you use credential profile file authentication.

If you do not specify the credential profile file path, the default credential file path is used. If you do not specify the profile name, the credentials are used from the default profile in the credential file.

Consider the following rules for a credential profile file:

- The credential file must be on the same machine where you installed the Secure Agent.
- The credential profile file name must end with `.credentials`.
- If you do not specify the credential profile path, the Secure Agent uses the credential profile file present in the following default location in your home directory:
  `~/.aws/credentials`

  Note: On Windows, you can refer to your home directory by using the environment variable `%UserProfile%`. On Unix-like systems, you can use the environment variable `$HOME`.

A sample credential profile file:

```
[default]

aws_access_key_id = 1233333

aws_secret_access_key = abcabcabc


[test-profile]

aws_access_key_id = 1233333
```

```
aws_secret_access_key = abcabcabc

aws_session_token = jahaheieomdrtflmlioerp
```

The aws_access_key_id and aws_secret_access_key specify the AWS access key and secret key used as part of credentials to authenticate the user.

The aws_session_token specifies an AWS session token used as part of the credentials to authenticate the user. A session token is required only if you specify temporary security credentials.

CHAPTER 2

Amazon S3 V2 connections

This chapter includes the following topics:

- Amazon S3 V2 connections overview, 15
- Amazon S3 V2 connection properties, 16
- Configuring proxy settings, 19

Amazon S3 V2 connections overview

Amazon S3 V2 connection enables you to read data from or write data to Amazon S3.

Use Amazon S3 V2 connections to specify sources and targets in mappings, elastic mappings, and mapping tasks.

You can use AWS Identity and Access Management (IAM) authentication to securely control access to Amazon S3 resources. You can also use temporary security credentials and federated user single sign-on to securely control access to Amazon S3 resources.

Create an Amazon S3 V2 connection on the **Connections** page and associate it with a 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	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. Description of the connection. The description cannot exceed 4,000 characters.
Type	The Amazon S3 V2 connection type.
Runtime Environment	Name of the runtime environment where you want to run the tasks. Specify a Secure Agent, Hosted Agent, or serverless runtime environment.
Access Key	Access key to access the Amazon S3 bucket. Provide the access key value based on the following authentication methods: - Basic authentication. Provide the actual access key value. - IAM authentication. Do not provide the access key value. - Temporary security credentials via assume role. Provide access key of an IAM user with no permissions to access Amazon S3 bucket. - Assume role for EC2. Do not provide the access key value. - Credential profile file authentication. Do not provide the access key value. - Federated user single sign-on. Do not provide the access key value.
Secret Key	Secret access key to access the Amazon S3 bucket. The secret key is associated with the access key and uniquely identifies the account. Provide the access key value based on the following authentication methods: - Basic authentication. Provide the actual access secret value. - IAM authentication. Do not provide the access secret value. - Temporary security credentials via assume role. Provide access secret of an IAM user with no permissions to access Amazon S3 bucket. - Assume role for EC2. Do not provide the access secret value. - Credential profile file authentication. Do not provide the access secret value. - Federated user single sign-on. Do not provide the access secret value.
IAM Role ARN	The ARN of the IAM role assumed by the user to use the dynamically generated temporary security credentials. Enter the value of this property if you want to use the temporary security credentials to access the AWS resources. **Note:** Even if you remove the IAM role that enables the Secure Agent to access the Amazon S3 bucket, and create a connection, the test connection is successful. For more information about how to obtain the ARN of the IAM role, see the AWS documentation.
External Id	Optional. Specify the external ID for a more secure access to the Amazon S3 bucket when the Amazon S3 bucket is in a different AWS account.

Property	Description
Use EC2 Role to Assume Role	Optional. Select the check box to enable the EC2 role to assume another IAM role specified in the IAM Role ARN option. **Note:** The EC2 role must have a policy attached with a permission to assume an IAM role from the same or different account. By default, the Use EC2 Role to Assume Role check box is not selected.
Folder Path	Bucket name or complete folder path to the Amazon S3 objects. 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, replace the existing `JCE` files with the latest `JCE` files that are available in the Secure Agent installation location and restart the Secure Agent.
Customer Master Key ID	Optional. Specify the customer master key ID or alias name generated by AWS Key Management Service (AWS KMS) or the Amazon Resource Name (ARN) of your custom key for cross-account access. **Note:** Cross-account access is not applicable to elastic mappings. 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.

Property	Description
Region Name	The AWS region of the bucket that you want to access. Select one of the following regions: - Asia Pacific (Mumbai) - Asia Pacific (Seoul) - Asia Pacific (Singapore) - Asia Pacific (Sydney) - Asia Pacific (Tokyo) - Asia Pacific (Hong Kong) - AWS GovCloud (US) - AWS GovCloud (US-East) - Canada (Central) - China (Beijing) - China (Ningxia) - EU (Ireland) - EU (Frankfurt) - EU (London) - EU (Paris) - EU (Stockholm) - South America (Sao Paulo) - Middle East (Bahrain) - US East (Ohio) - US East (N. Virginia) - US West (N. California) - US West (Oregon) Default is US East (N. Virginia). **Note:** Middle East (Bahrain) region is not applicable for elastic mappings.
Federated SSO IdP	SAML 2.0-enabled identity provider for the federated user single sign-on to use with the AWS account. Amazon S3 V2 connector supports only `ADFS 3.0` identity provider. Select `None` if you do not want to use federated user single sign-on. **Note:** Federated user single sign-on is not applicable to elastic mappings.
Other Authentication Type	Select **NONE** or **Credential Profile File Authentication**. Select the Credential Profile File Authentication option to access the Amazon S3 credentials from a credential file that contains the access key and secret key. Specify the credential profile file path and the profile name to establish the connection with Amazon S3. You can use permanent IAM credentials or temporary session tokens when you configure the Credential Profile File Authentication. Default is NONE.
Credential Profile File Path	Optional. Specify the credential profile file path. If you do not specify the credential profile path, the Secure Agent uses the credential profile file present in the following default location in your home directory: `~/.aws/credentials` **Note:** Mass Ingestion Databases has not been certified with the **Credential Profile File Path** and **Profile Name** connection properties. Mass Ingestion Databases finds AWS credentials by using the default credential provider chain that is implemented by the DefaultAWSCredentialsProviderChain class, which includes the credential profile file.
Profile Name	Optional. Name of the profile in the credential profile file used to get the credentials. If you do not specify the profile name, the credentials from the default profile in the credential profile file are used.

Federated user single sign-on connection properties

Configure the following properties when you select `ADFS 3.0` in **Federated SSO IdP**:

Property	Description
Federated User Name	User name of the federated user to access the AWS account through the identity provider.
Federated User Password	Password for the federated user to access the AWS account through the identity provider.
IdP SSO URL	Single sign-on URL of the identity provider for AWS. Not applicable for a streaming ingestion task.
SAML Identity Provider ARN	ARN of the SAML identity provider that the AWS administrator created to register the identity provider as a trusted provider.
Role ARN	ARN of the IAM role assumed by the federated user.

Note: Amazon S3 V2 connector supports reading from or writing to complex files such as Avro, Parquet, and ORC only on Hortonworks 2.5 distribution for the federated single sign-on user.

Rules and guidelines for AWS regions

Consider the following rules and guidelines when you configure the region name of the bucket in the connection properties.

- When you change the runtime environment of an existing connection, the region is changed to the default region `US East (N. Virginia)`. Select the region manually to change the default region.
- You can use the regions with spaces for an Amazon S3 bucket by copying the latest AWS SDK at the following location on the Secure Agent machine:

  ```
  <Secure agent installation directory>/downloads/<package.AmazonS3V2.zip>/package/s3/
  thirdparty/infa.amazon.s3
  ```
- When you edit an existing connection, you see duplicate entries for regions. Use the regions that contain spaces because these regions are populated from AWS SDK. For example, use `US West (Oregon)` instead of `US West(Oregon)`.
- When you read from or write to complex files, you can only use Cloudera 6.1 distribution to use the regions with spaces.

Configuring proxy settings

If your organization uses an outgoing proxy server to connect to the internet, the Secure Agent connects to Informatica Intelligent Cloud Services through the proxy server.
Contact your network administrator for the correct proxy settings.

Consider the following rules when you configure a proxy server:

- You can only use an unauthenticated proxy server to connect to Informatica Intelligent Cloud Services.

- When you use a serverless runtime environment, you cannot use a proxy server to connect to Informatica Intelligent Cloud Services.
- Proxy settings do not apply to elastic mappings.
- You cannot use an unauthenticated proxy server when you read from complex files.

Configuring proxy settings on Windows

To configure the proxy server settings for the Secure Agent on a Windows machine, you can configure the proxy server settings through the Secure Agent or the JVM options of the Secure Agent.

Configuring proxy settings through Secure Agent Manager

To configure the proxy server settings through the Secure Agent Manager, perform the following steps:

1. Click **Start** > **All Programs** > **Informatica Cloud Secure Agent** > **Informatica Cloud Secure Agent** to launch the Secure Agent Manager.

   The Secure Agent Manager displays the Secure Agent status.
2. Click **Proxy** in the Secure Agent Manager page.
3. Click **Use a Proxy Server** to enter proxy server settings.
4. Configure the following proxy server details:

Field	Description
Proxy Host	Required. Host name of the outgoing proxy server that the Secure Agent uses.
Proxy Port	Required. Port number of the outgoing proxy server.

5. Click **OK**.

   The Secure Agent Manager restarts the Secure Agent to apply the settings.

Configuring the proxy settings through JVMOptions

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

Parameter	Description
-Dhttps.proxyHost=	Host name of the outgoing HTTPS proxy server.
-Dhttps.proxyPort=	Port number of the outgoing HTTPS proxy server.

For example,

```
JVMOption1=-Dhttps.proxyHost=<proxy_server_hostname>
```

```
JVMOption2=-Dhttps.proxyPort=<proxy_port>
```

6. Click **Save**.

   The Secure Agent restarts to apply the settings.

Configuring proxy settings on Linux

The Secure Agent installer configures the proxy server settings for the Secure Agent based on settings configured in the browser. You can update the proxy server settings defined for the Secure Agent from the command line.

To configure the proxy server settings for the Secure Agent on a Linux machine, use a shell command that updates the `proxy.ini` file. Contact the network administrator to determine the proxy settings.

1. Navigate to the following directory:

   ```
   <Secure Agent installation directory>/apps/agentcore
   ```

2. Update the `proxy.ini` file.
   - To update the `proxy.ini` file for an unauthenticated proxy, enter the following command:

     ```
     consoleAgentManager.bat configureProxy <proxy host> <proxy port>
     ```

3. Restart the Secure Agent.

CHAPTER 3

Amazon S3 V2 sources and targets

This chapter includes the following topics:

- Amazon S3 V2 sources, 22
- Amazon S3 V2 targets, 27
- Data compression in Amazon S3 V2 sources and targets, 33

Amazon S3 V2 sources

You can use an Amazon S3 V2 object as a source in a mapping, elastic mapping, or mapping task.

When you configure the advanced source properties, 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.

The following table lists the Amazon S3 V2 source features that you can use in mappings:

Feature	Mapping	Elastic Mapping
Assume role	Yes	Yes
Credential profile file authentication	Yes	No
Federated user single sign-on	Yes	No
Client-side encryption	Yes	No
Informatica encryption	Yes	Yes
File source type	Yes	Yes
Directory source type	Yes **Note:** When you run a mapping that reads data from a directory, the Secure Agent creates a single file in the target.	Yes **Note:** When you run an elastic mapping that reads data from a directory, the Secure Agent creates multiple files in the target.

Feature	Mapping	Elastic Mapping
Reading multiple files	Yes	No
Partitioning	Yes	You can use directory-level partitioning.
Data compression	Yes	Yes
FileName field	Yes (Avro, ORC, Parquet, and Binary)	Yes (Avro, ORC, and Parquet)
Read hierarchical data	No	Yes (Avro, JSON, and Parquet)
Discover Structure file format	No	Yes

Data encryption in Amazon S3 V2 sources

You can decrypt data when you read binary and flat file sources 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 decrypts the data by using the master symmetric key or customer master key. When you use a serverless runtime environment, you cannot configure client-side encryption for Amazon S3 V2 sources.

When you generate a client-side encrypted file using a third-party tool, metadata for the encrypted file is generated. To read an encrypted file from Amazon S3, you must upload the encrypted file and the metadata for the encrypted file to the Amazon S3 bucket.

You require the following keys in the metadata when you upload the encrypted file:

- Content-Type
- x-amz-meta-x-amz-key
- x-amz-meta-x-amz-unencrypted-content-length
- x-amz-meta-x-amz-matdesc
- x-amz-meta-x-amz-iv

Reading a client-side encrypted file

Perform the following tasks to read a client-side encrypted file:

1. Provide the master symmetric key when you create an Amazon S3 V2 connection.
   Ensure that you provide a 256-bit AES encryption key in Base64 format.
2. Download the `local_policy.jar` and `US_export_policy.jar` files for your JAVA environment from the Oracle website.
3. Replace the existing `local_policy.jar` and `US_export_policy.jar` files in the following directory: `<Secure Agent installation directory>apps\jdk\1.8.0_202\jre\lib\security\policy\unlimited`.
4. Enable the client-side encryption in the mapping.

Informatica encryption for Amazon S3 V2 sources

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

Informatica encryption is applicable only when you run mappings on the Secure Agent machine. To read a source file that is encrypted using the Informatica crypto libraries, perform the following tasks:

1. Ensure that the organization administrator has permission to Informatica crypto libraries license when you create an Amazon S3 V2 connection.
2. Select **Informatica Encryption** as the encryption type in the advanced source properties.

When you read an Informatica encrypted source file and select the **Informatica Encryption** as the encryption type, the data preview fails.

To preview the data successfully, select a dummy source file that contains same metadata present in the Informatica encrypted source file that you want to read. Enter the file name of the Informatica encrypted source file in the **File Name** advanced source property to override the file name of the dummy source file. Then, select **Informatica Encryption** as the encryption type in the advanced source property.

Note: When you use Informatica encryption in a mapping, you cannot decrypt more than 1000 files.

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 and 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.

To read multiple flat 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, which are of flat format type, 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 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.

Note: If you configure partitioning for an Amazon S3 V2 source in a mapping to read from a manifest file, compressed `.gz` file, or a read directory file, the Secure Agent ignores the partition. However, the task runs successfully.

Reading source objects path

When you import source objects, the Secure Agent appends a FileName field to the imported source object. The FileName field stores the absolute path of the source file from which the Secure Agent reads the data at run time.

For example, a directory contains a number of files and each file contains multiple records that you want to read. You select the directory as source type in the Amazon S3 V2 source advanced properties. When you run the mapping, the Secure Agent reads each record and stores the absolute path of the respective source file in the FileName field.

The FileName field is applicable to the following file formats:

- Avro
- Binary
- ORC
- Parquet

When you use the FileName field in a source object, the Secure Agent reads file names and directory names differently for mappings and elastic mappings.

Feature	Mapping	Elastic Mapping
File name	xyz.amazonaws.com/aa.bb.bucket/1024/characterscheckfor1024	s3a://<bucket_name>/customer.avro
Directory name	<absolute path of the file including the file name>	s3a://<bucket_name>/avro/<directory_name>/<file_name>

Note: The FileName field in a source object uses the format with -, by default. For example, s3-us-west-2.amazonaws.com/<bucket_name>/automation/customer.avro.

To change the format for the FileName field to use ., set the JVM option changeS3EndpointForFileNamePort = true. For example, s3.us-west-2.amazonaws.com/<bucket_name>/automation/customer.avro.

Pushdown optimization

You can enable full pushdown optimization when you want to load data from Amazon S3 sources to your data warehouse in Amazon Redshift. Before loading the data to Amazon Redshift, you can transform the data as per your data warehouse model and requirements. When you enable full pushdown on a mapping task, the mapping logic is pushed to the AWS environment to leverage AWS commands. For more information, see the help for Amazon Redshift V2 Connector.

Amazon S3 V2 targets

You can use an Amazon S3 V2 object as a target in a mapping, elastic 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.

The following table lists the Amazon S3 V2 target features that you can use in mappings:

Feature	Mapping	Elastic Mapping
Assume role	Yes	Yes
Credential profile file authentication	Yes	No
Federated user single sign-on	Yes	No
Server-side encryption	Yes	Yes
Server-side encryption with KMS	Yes*	Yes
Client-side encryption	Yes*	No
Informatica encryption	Yes	Yes

Feature	Mapping	Elastic Mapping
Partitioning	Yes	You can use directory-level partitioning.
Distribution column	Yes	No
Object tag	Yes	Yes
Data compression	Yes	Yes
Target file parameterization	Yes	No
FileName field	Yes	Yes
Write hierarchical data	No	Yes (Avro, JSON, and Parquet), only when **CreateTarget** is used.

Data encryption in Amazon S3 V2 targets

To protect data, you can encrypt the Amazon S3 files when you write the files to the target. Do not use the master symmetric key and customer master key together.

Select the type of the encryption in the **Encryption Type** field under the Amazon S3 V2 advanced target properties.

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 to the target.

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 to the binary or flat file 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.

Client-side encryption

Select **Client Side Encryption** as the encryption type if you want the Secure Agent to encrypt the data when you write to the binary or flat file 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.

When you use a serverless runtime environment, you cannot configure client-side encryption for Amazon S3 V2 targets.

Informatica encryption

Select **Informatica Encryption** as the encryption type if you want to encrypt the data using Informatica crypto libraries when you write to a binary or flat file target. Informatica encryption is applicable only when you run mappings on the Secure Agent machine.

To encrypt a file using Informatica Encryption method, perform the following tasks:

1. Ensure that the organization administrator has permission to Informatica crypto libraries when you create an Amazon S3 V2 connection.
2. Select **Informatica Encryption** as the encryption type in the advanced target properties.

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

Encryption Type	Avro File	Binary File	Delimited	JSON File[1]	ORC File	Parquet File
Client-side encryption	No	Yes	Yes	No	No	No
Server-side encryption	Yes	Yes	Yes	Yes	Yes	Yes
Server-side encryption with KMS	Yes	Yes	Yes	Yes	Yes	Yes
Informatica encryption	No	Yes	Yes	No	No	No

[1]. Applies only to elastic mappings.

Rules and guidelines for data encryption in Amazon S3 V2 targets

Consider the following rules and guidelines when you configure data encryption in Amazon S3 V2 targets:

- Amazon S3 V2 Connector supports reading from or writing to complex files such as Avro, Parquet and ORC only on Cloudera 6.1 distribution for the server-side encryption with KMS.
- When you write an Avro, ORC, or Parquet file to an Amazon S3 target, you cannot encrypt the file using the client-side encryption.
- When you use Informatica encryption in a mapping, the `_SUCCESS` file is not generated in the target directory for elastic mappings.
- When you use Informatica encryption in a mapping, you cannot encrypt more than 1000 files.

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 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 flat files to Amazon S3 target from a single source. Configure the **Distribution Column** option in the advanced target properties.

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`

Writing to multiple target objects

When you import target objects, the Secure Agent appends a FileName field to the imported target object. When you map the FileName field in the target object to an incoming field, the Secure Agent creates the folder structure and the target files based on the FileName field. For example:

Syntax:

```
<tgt_filename_folder>/<tgt_filename=incoming_value_folder>/part_file
```

Sample:

```
emp_tgt.parquet/emp_tgt.parquet=128000/part-0000-e9ca8-6af-efd43-455c-8709.c000.parquet
```

The FileName field is applicable to the following file formats:

- Avro
- ORC
- Parquet

Consider the following guidelines when using the target FileName field in mappings:

- Do not map the source object FileName field to the target object FileName field. If you map the FileName field in the target object to an incoming field, the Secure Agent does not create directory structure as expected.
- When you use the FileName field in a target object, the Secure Agent creates folders with different names for null values for mappings and elastic mappings:
  - For mappings: `_EMPTY_`
  - For elastic mappings: `_HIVE_DEFAULT_PARTITION_`
- When you map a date type incoming field to the FileName field in the target object, the Secure Agent creates a nested folder structure based on the incoming date value for target objects.
- When you map an incoming field to the FileName field in the target object, the mapping runs successfully for the first time. At subsequent runs, the mapping fails with the following error:
  `Operation failed: Index: 0, Size: 0.`

  To successfully rerun the mapping, use a dummy target file at design time and override the dummy target file in advanced target properties.

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.

The following table lists the supported source compression formats:

Compression format	Avro File	Binary File [1]	Delimited	JSON File [2]	ORC File	Parquet File
None	Yes	No	Yes	Yes	Yes	Yes
Bzip2	No	No	No	No	No	No
Deflate	Yes	No	No	No	No	No
Gzip	No	No	Yes	No	No	Yes
Lzo	No	No	No	No	No	No
Snappy	Yes	No	No	No	Yes	Yes
Zlib	No	No	No	No	Yes	No

[1]. *Applies only to mappings.*
[2]. *Applies only to elastic mappings.*

The following table lists the supported target compression formats:

Compression format	Avro File	Binary File [1]	Delimited	JSON File [2]	ORC File	Parquet File
None	Yes	No	Yes	Yes	Yes	Yes
Binary	No	No	No	No	No	No
Bzip2	No	No	No	Yes	No	No
Deflate	Yes	No	No	Yes	No	No
Gzip	No	No	Yes	Yes	No	Yes
Lzo	No	No	No	No	No	No
Snappy	Yes	No	No	Yes	Yes	Yes
Zlib	No	No	No	No	Yes	No

[1]. *Applies only to mappings.*
[2]. *Applies only to elastic mappings.*

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

For the Avro, ORC and Parquet file formats, the support for the following compression formats are implicit even though these compression formats do not appear in the **Compression Format** option under the advanced source property:

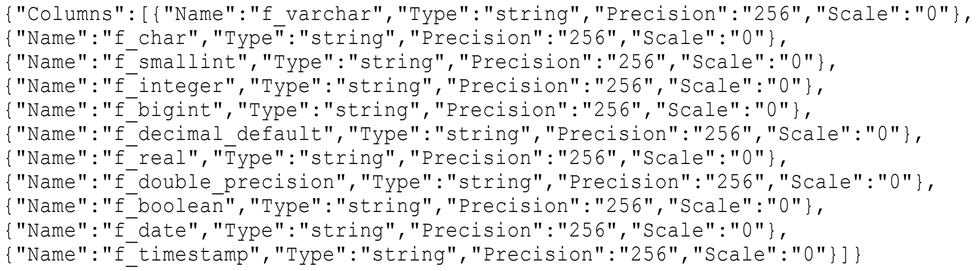
Compression format	Avro File	ORC File	Parquet File
Deflate	Yes	No	No
Snappy	Yes	Yes	Yes
Zlib	No	Yes	No

Reading a compressed flat file

When you run a mapping to read a compressed flat file, you must upload a schema file and select `Gzip` as the compression format. Use the **.GZ** file name extension when you use the `Gzip` compression format to read a flat file.

1. Select the required compressed flat 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.

CHAPTER 4

Mappings and mapping tasks with Amazon S3 V2

This chapter includes the following topics:

- Amazon S3 V2 objects in mappings, 35
- Amazon S3 V2 file formatting options, 42
- Rules and guidelines for creating an elastic mapping, 44
- Amazon S3 V2 objects in mapping tasks, 45
- Specifying a target, 50
- Amazon S3 V2 target file parameterization, 50
- Directory-level partitioning, 52
- Elastic mapping example, 54

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. Select a source connection or click **New Parameter** to define a new parameter for the source connection.
Source Type	Source type. Select one of the following types: - Single Object - Parameter: Select Parameter to define the source type when you configure the mapping task.
Object	Name of the source object.
Parameter	Select an existing parameter for the source object or click **New Parameter** to define a new parameter for the source object. The Parameter property appears only if you select Parameter as the source type. If you want to overwrite the parameter at runtime, select the **Overwrite Parameter** option.
Format	Specifies the file format that the Amazon S3 V2 Connector uses to read data from Amazon S3. You can select the following file format types: - None - Delimited - Avro - ORC - Parquet - JSON - Discover Structure You can use the JSON format only in elastic mappings. Default is **None**. If you select **None** is as the format type, the Secure Agent reads data from Amazon S3 files in binary format. You cannot read binary files in elastic mappings. Discover Structure format is applicable only to elastic mappings. You cannot use parameterized sources when you select the discover structure format. Open the **Formatting Options** dialog box to define the format of the file. For more information about format options, see "Amazon S3 V2 file formatting options" on page 42.
Intelligent Structure Model	Applicable to Discover Structure format type. Select the intelligent structure model. For more information, see *Components*.

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

Note: The properties appear based on the type of mapping you create.

Property	Description
Source Type	Type of the 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 24.
Folder Path	Overwrites the bucket name or folder path of 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	Overwrites the Amazon S3 source file name.
Encryption Type	Method you want to use to decrypt data. You can select one of the following encryption types: - Informatica Encryption - None Default is **None**. For more information about Informatica Encryption, see "Informatica encryption for Amazon S3 V2 sources " on page 24.
Staging Directory	Path of the local 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. Default staging directory is the `/temp` directory on the machine that hosts the Secure Agent. 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` The temporary files are created within the new directory. The staging directory in the source property does not apply to elastic mappings. However, you must specify a staging directory on Amazon S3 in elastic configurations. For more information, see *Administrator*.
Hadoop Performance Tuning Options	This property is not applicable for Amazon S3 V2 Connector.

Property	Description
Compression Format	Decompresses data when you read data from Amazon S3. You can choose to decompress the data in the following formats: - None - Bzip2 - Gzip - Lzo Default is **None**. You can decompress data in an elastic mapping if the mapping reads data from a JSON file in Bzip2 format. **Note:** Amazon S3 V2 Connector does not support the Lzo compression format even though the option appears in this property. For more information about the compression format, see "Data compression in Amazon S3 V2 sources and targets" on page 33.
Download Part Size	Downloads the part size of an Amazon S3 object in bytes. Default is 5 MB. Use this property when you run a mapping to read a file of flat format type.
Multiple Download Threshold	Minimum threshold size to download an Amazon S3 object in multiple parts. To download the object in multiple parts in parallel, ensure that the file size of an Amazon S3 object is greater than the value you specify in this property. Default is 10 MB.
Temporary Credential Duration	The time duration during which an IAM user can use the dynamically generated temporarily credentials to access the AWS resource. Enter the time duration in seconds. Default is 900 seconds. If you require more than 900 seconds, you can set the time duration maximum up to 12 hours in the AWS console and then enter the same time duration 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. Select a target connection or click **New Parameter** to define a new parameter for the target connection.
Target Type	Target type. Select one of the following types: - Single Object - Parameter: Select **Parameter** to define the target type when you configure the mapping task.
Object	Name of the target object. You can select an existing object or create an object at runtime.

Property	Description
Parameter	Select an existing parameter for the source object or click **New Parameter** to define a new parameter for the target object. The Parameter property appears only if you select Parameter as the target type. If you want to overwrite the parameter at runtime, select the **Overwrite Parameter** option.
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. You can use parameters defined in a parameter file in the target name. 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. 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:** If you specify a target name that already exists, you do not get a warning message. However, the Secure Agent overwrites the existing target with the same file name. When you write an Avro, ORC, or Parquet file using the **Create Target** option, you cannot provide a Null data type.
Format	Specifies the file format that the Amazon S3 V2 Connector uses to write data Amazon S3. You can select the following file format types: - None - Delimited - Avro - ORC - Parquet - JSON You can use the JSON format only in elastic mappings. Default is **None**. If you select **None** is as the format type, the Secure Agent writes data to Amazon S3 files in binary format. You cannot write binary files in elastic mappings. Open the **Formatting Options** dialog box to define the format of the file. For more information about format options, see "Amazon S3 V2 file formatting options" on page 42.
Operation	Type of 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:

Note: The properties appear based on the type of mapping you create.

Property	Description
Overwrite File(s) If Exists	Overwrites an existing target file. Default is true. For more information about overwriting the existing files, see "Overwriting existing files" on page 29.
Folder Path	Bucket name or folder path where you want to write 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	Creates a new file name or overwrites an existing target file name.
Encryption Type	Method you want to use to encrypt data. Select one of the following encryption types: - None - Client Side Encryption - Server Side Encryption - Server Side Encryption with KMS - Informatica Encryption Default is **None**. You can use only **Server Side Encryption** in an elastic mapping. For more information about the encryption type, see "Data encryption in Amazon S3 V2 targets" on page 28.
Staging Directory	Enter the path of the local 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. Default staging directory is the `/temp` directory on the machine that hosts the Secure Agent. 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` The temporary files are created within the new directory. The staging directory in the target property does not apply to elastic mappings. However, you must specify a staging directory on Amazon S3 in elastic configurations. For more information, see *Administrator*.
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 - Bzip2 - Deflate - Gzip - Lzo - Snappy - Zlib Default is None. You can compress data in an elastic mapping if the mapping writes data from a JSON file in Bzip2 format. **Note:** Amazon S3 V2 Connector does not support the Lzo compression format even though the option appears in this property. For more information about the compression format, see "Data compression in Amazon S3 V2 sources and targets" on page 33.
Object Tags	The key value pairs to 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. Use this property when you run a mapping to write a file of flat format type. For more information about the object tags, see "Object tag" on page 31.
TransferManager Thread Pool Size	The number of threads to write data in parallel. Default is 10. Use this property when you run a mapping to write a file of flat format type. 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 **TransferManager Thread Pool Size** to greater than 50, the value reverts to 50.
Merge Partition Files	This property is not applicable for Amazon S3 V2 Connector.
Temporary Credential Duration	The time duration during which an IAM user can use the dynamically generated temporarily credentials to access the AWS resource. Enter the time duration in seconds. Default is 900 seconds. If you require more than 900 seconds, you can set the time duration maximum up to 12 hours in the AWS console and then enter the same time duration in this property.
Part Size	Uploads the part size of an Amazon S3 object in bytes. Default is 5 MB. 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.

Note: When you read from or write to the ORC files on Cloudera 6.1 distribution, the unicode characters are parsed incorrectly.

When you create a mapping and the column name in the Amazon S3 source or target object contains special characters, the Secure Agent replaces the special characters with an underscore `(_)` and the mapping fails.

Amazon S3 V2 lookups

You can use Amazon S3 V2 objects in a connected and an unconnected cached Lookup transformation.

For more information about the Lookup transformation, see *Transformations*.

Amazon S3 V2 file formatting options

When you select a source or a target Amazon S3 V2 file, you can configure format options.

Note: The properties appear based on the type of mapping you create.

Schema Source

You must specify the schema of the source or target file. You can select one of the following options to specify a schema:

- **Read from data file**. Amazon S3 V2 Connector imports the schema from the file in Amazon S3.
- **Import from Schema File**. Imports schema from a schema definition file in your local machine.

Read from data file

If you select **Read from data file**, you can select one of the following options:

- **Data elements to sample**. The number of rows to read from the metadata.
- **Memory available to process data**. The memory that the parser uses to read the JSON sample schema and process it. You can increase the parser memory. Default is 2 MB.

Applicable only to elastic mappings.

Schema File

You can upload a schema file.

You cannot upload a schema file when you select the **Create Target** option to write data to Amazon S3.

If you select **Flat** as the **Format Type**, you must configure the following options:

Delimiter

Character used to separate columns of data. You can configure parameters such as comma, tab, colon, semicolon, or others. To set a tab as a delimiter, you must type the tab character in any text editor. Then, copy and paste the tab character in the **Delimiter** field.

If you specify a multibyte character as a delimiter in the source object, the mapping fails.

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 data. You can set qualifier as single quote or double quote.

Qualifier Mode

Specify the qualifier behavior for the target object. You can select one of the following options:

- **Minimal**. Default mode. Applies qualifier to data that have a delimiter value or a special character present in the data. Otherwise, the Secure Agent does not apply the qualifier when writing data to the target.
- **All**. Applies qualifier to all 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 file that does not have a header. Default is 1.

To read data from a file with no header, specify the value of the **Header Line Number** field as 0. To read data from a file with a header, set the value of the **Header Line Number** field to a value that is greater than or equal to one.

This property is applicable during runtime and data preview to read a file.

First Data Row

Specify the line number from where you want the Secure Agent to read data. You must enter a value that is greater or equal to one.

To read data from the header, the value of the **Header Line Number** and the **First Data Row** fields should be the same. Default is 1.

This property is applicable during runtime and data preview to read a file. This property is applicable during data preview to write a file.

Target Header

Select whether you want to write data to a target that contains a header or without a header in the flat file. You can select **With Header** or **Without Header** options.

This property is not applicable when you read data from a Amazon S3 source.

Distribution Column

Specify the name of the column that is used to create multiple target files during run time.

This property is not applicable when you read data from a Amazon S3 source. For more information about the distribution column, see “Distribution column” on page 30.

escapeCharacterDataRetained

Not applicable to Amazon S3 V2 Connector.

maxRowsToPreview

Not applicable to Amazon S3 V2 Connector.

rowDelimiter

Character used to separate rows of data. You can set values as `\r\n`, `\n`, and `\r`.

Rules and guidelines for setting formatting options

You must set the appropriate formatting options when you select the Amazon S3 file format types.

Use the following guidelines when you select the format types and set the formatting options:

- You can use JSON format only in elastic mappings.
- 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.
- If you select an Avro, JSON, ORC, 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, JSON, ORC, or Parquet 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.
- If you select the Delimited format type and select **Import from schema file** as the value of the **Schema Source** formatting option, you can only upload a schema file in the JSON format.
  The following sample shows a schema file for a delimited file:

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

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

  ```
  invalid character encapsulated
  ```

Rules and guidelines for creating an elastic mapping

Consider the following guidelines when you create an elastic mapping:

- When you run an elastic mapping, a folder of the following format is created in the target and multiple target files are generated within the folder: `<target_foldername>.<file_extension>`
- When an elastic mapping writes data to an Amazon S3 file, the file is replaced with a folder and the target file is generated inside the folder. If you create another elastic mapping that references the same Amazon S3 file, the file cannot be found and the mapping fails with the following error message:

  ```
  Operation failed: Index: 0, Size: 0.
  ```

- When you run an elastic mapping with a Lookup transformation and select the Report Error option to report an error on multiple matches, the data from the unmatched columns is written to the target without displaying an error message.
- When you run an elastic mapping that reads data from a JSON file and if one of the rows contains an incorrect boolean value, all rows are rejected and a null value is written to the target.
- Do not run an elastic mapping to read data from a JSON file that contains unicode characters to avoid data inconsistencies.
- In an elastic mapping, if you select data types that Amazon S3 V2 Connector does not support, the mapping might either fail or reject the rows.
- When there are empty array elements such as `"{"Elements":[]}"` as part of the JSON sample data file for metadata resolution, the JSON parser fails with the following error:

  ```
  [SDK_APP_COM_20000] error [Array must contain at least 1 element for projection].
  ```

  Provide the entire schema as a sample data row without any empty array for metadata resolution. You can use the `Import from schema file` option to upload this file.
- When there are empty struct elements such as `"{"Elements":[]}"` as part of the JSON sample data file for metadata resolution, the JSON parser fails with the following error:

  ```
  Struct must contain at least one key :: fields
  ```

  Provide the entire schema as a sample data row without any empty structs for metadata resolution. You can use the `Import from schema file` option to upload this file.
- The JSON parser interprets the data type for an array element using the first value from the array. For example, if the first value is an integer and subsequently contains long values, the metadata is resolved as an integer. During runtime, the entire row is dropped because the long value cannot fit into the DTM buffer.

  Provide the entire schema as a sample data row with the first array or struct elements containing the data types or sub-fields that are required for the metadata resolution. You can use the `Import from schema file` option to upload the file.
- There is a limit of 2 MB to the first row size used to interpret metadata from a JSON file. If the first row in the data file is larger than 2 MB, the JSON parser fails.

  Decrease the sample data row size by removing the additional tags from the struct and array elements. The JSON parser only requires the first element within struct and array elements. Provide the data types or sub-fields that are required for metadata resolution in the first element. You can use the `Import from schema file` option to upload the file.

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:

Note: The properties appear based on the type of mapping you create.

Property	Description
Source Type	Type of the 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 24.
Folder Path	Overwrites the bucket name or folder path of 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	Overwrites the Amazon S3 source file name.
Encryption Type	Method you want to use to decrypt data. You can select one of the following encryption types: - Informatica Encryption - None Default is **None**. For more information about Informatica Encryption, see "Informatica encryption for Amazon S3 V2 sources " on page 24.
Staging Directory	Path of the local 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. Default staging directory is the `/temp` directory on the machine that hosts the Secure Agent. 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` The temporary files are created within the new directory. The staging directory in the source property does not apply to elastic mappings. However, you must specify a staging directory on Amazon S3 in elastic configurations. For more information, see *Administrator*.
Hadoop Performance Tuning Options	This property is not applicable for Amazon S3 V2 Connector.

Property	Description
Compression Format	Decompresses data when you read data from Amazon S3. You can choose to decompress the data in the following formats: - None - Bzip2 - Gzip - Lzo Default is **None**. You can decompress data in an elastic mapping if the mapping reads data from a JSON file in Bzip2 format. **Note:** Amazon S3 V2 Connector does not support the Lzo compression format even though the option appears in this property. For more information about the compression format, see "Data compression in Amazon S3 V2 sources and targets" on page 33.
Download Part Size	Downloads the part size of an Amazon S3 object in bytes. Default is 5 MB. Use this property when you run a mapping to read a file of flat format type.
Multiple Download Threshold	Minimum threshold size to download an Amazon S3 object in multiple parts. To download the object in multiple parts in parallel, ensure that the file size of an Amazon S3 object is greater than the value you specify in this property. Default is 10 MB.
Temporary Credential Duration	The time duration during which an IAM user can use the dynamically generated temporarily credentials to access the AWS resource. Enter the time duration in seconds. Default is 900 seconds. If you require more than 900 seconds, you can set the time duration maximum up to 12 hours in the AWS console and then enter the same time duration 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:

Note: The properties appear based on the type of mapping you create.

Property	Description
Overwrite File(s) If Exists	Overwrites an existing target file. Default is true. For more information about overwriting the existing files, see "Overwriting existing files" on page 29.
Folder Path	Bucket name or folder path where you want to write 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	Creates a new file name or overwrites an existing target file name.
Encryption Type	Method you want to use to encrypt data. Select one of the following encryption types: - None - Client Side Encryption - Server Side Encryption - Server Side Encryption with KMS - Informatica Encryption Default is **None**. You can use only **Server Side Encryption** in an elastic mapping. For more information about the encryption type, see "Data encryption in Amazon S3 V2 targets" on page 28.
Staging Directory	Enter the path of the local 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. Default staging directory is the `/temp` directory on the machine that hosts the Secure Agent. 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` The temporary files are created within the new directory. The staging directory in the target property does not apply to elastic mappings. However, you must specify a staging directory on Amazon S3 in elastic configurations. For more information, see *Administrator*.
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 - Bzip2 - Deflate - Gzip - Lzo - Snappy - Zlib Default is None. You can compress data in an elastic mapping if the mapping writes data from a JSON file in Bzip2 format. **Note:** Amazon S3 V2 Connector does not support the Lzo compression format even though the option appears in this property. For more information about the compression format, see "Data compression in Amazon S3 V2 sources and targets" on page 33.
Object Tags	The key value pairs to 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. Use this property when you run a mapping to write a file of flat format type. For more information about the object tags, see "Object tag" on page 31.
TransferManager Thread Pool Size	The number of threads to write data in parallel. Default is 10. Use this property when you run a mapping to write a file of flat format type. 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 **TransferManager Thread Pool Size** to greater than 50, the value reverts to 50.
Merge Partition Files	This property is not applicable for Amazon S3 V2 Connector.
Temporary Credential Duration	The time duration during which an IAM user can use the dynamically generated temporarily credentials to access the AWS resource. Enter the time duration in seconds. Default is 900 seconds. If you require more than 900 seconds, you can set the time duration maximum up to 12 hours in the AWS console and then enter the same time duration in this property.
Part Size	Uploads the part size of an Amazon S3 object in bytes. Default is 5 MB. 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.

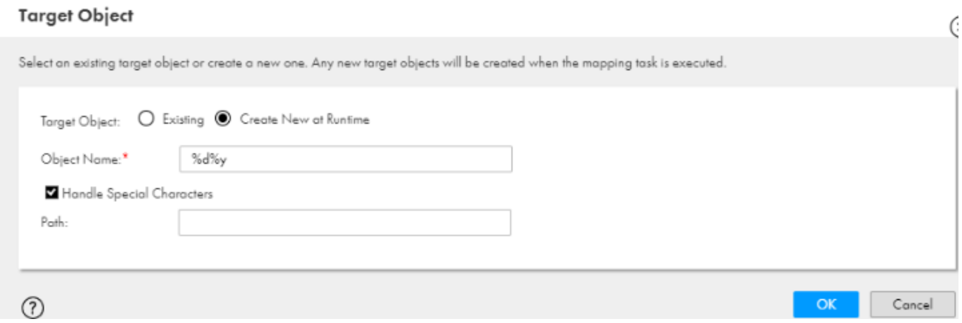
Specifying a target

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`.
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:



Target Object

Select an existing target object or create a new one. Any new target objects will be created when the mapping task is executed.

Target Object: ○ Existing ● Create New at Runtime

Object Name:* %d%y

☑ Handle Special Characters

Path:

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.

Does not apply to elastic mappings.

Parameterization using timestamp

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

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

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 handle any special characters in the `%[mod]` format included in the file name. You can use the STRFTIME function formats in a mapping.

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

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.

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

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.

Directory-level partitioning

You can read from and write to partition columns when you use elastic mappings.

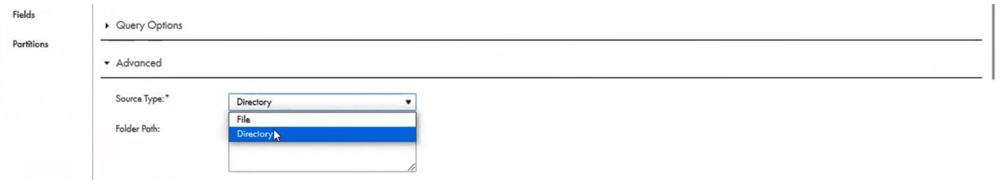
You can organize tables or data sets into partitions for grouping same type of data together based on a column or partition key. You can select one or more partition columns in a table or data set.

To read from partition columns, select a partition directory and identify the partition columns. To write to partition columns, you can add partition columns from the list of fields and change the partition order, if required.

Reading from partition columns

Perform the following steps to read data from partition columns:

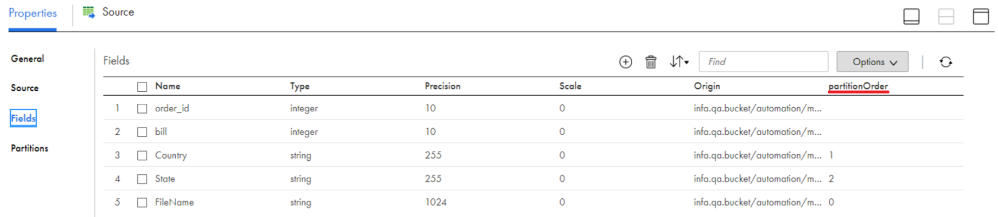
1. Select a directory from the list of source objects.
2. Select the Source Type as **Directory** in the Advanced Source Properties.



Source Type:* Directory

Folder Path:

3. In the Fields tab, you can view the number of partitions. The **partitionOrder** column appears for the list of partitioned fields, as shown in the following image:



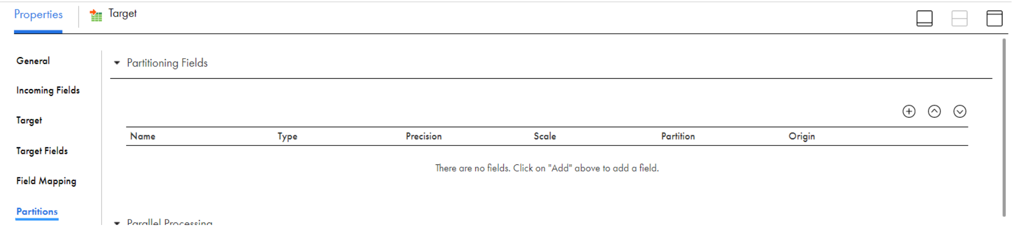
	Name	Type	Precision	Scale	Origin	partitionOrder
1	order_id	integer	10	0	infa.qa.bucket/automation/m...	
2	bill	integer	10	0	infa.qa.bucket/automation/m...	
3	Country	string	255	0	infa.qa.bucket/automation/m...	1
4	State	string	255	0	infa.qa.bucket/automation/m...	2
5	FileName	string	1024	0	infa.qa.bucket/automation/m...	0

The **partitionOrder** column specifies whether a column is partitioned.
In the above image, 2 partition columns are present. the partition order values 1 and 2 signify the order in which the `Country` and `State` fields were selected for partitioning. The FileName field has 0 as the partition order.

Writing to partition columns

Perform the following steps to write to partition columns:

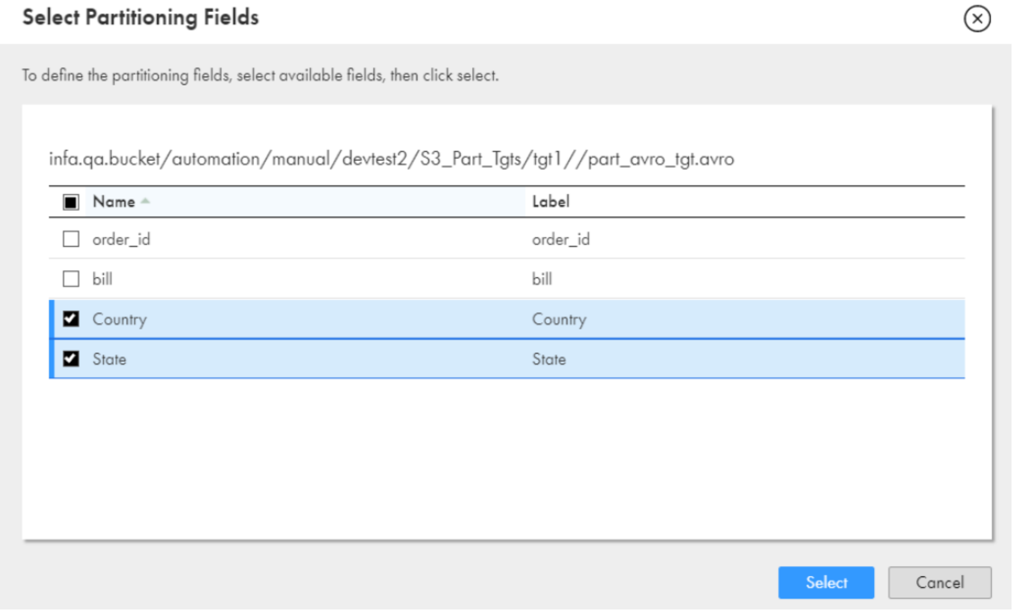
1. Click the ⊕ icon in the **Partitions** tab to add the partition columns for a target. The following image shows how you can add the partition columns:



Name	Type	Precision	Scale	Partition	Origin

There are no fields. Click on "Add" above to add a field.

2. In the Partitions tab, select the partitioning fields from the list of available fields.



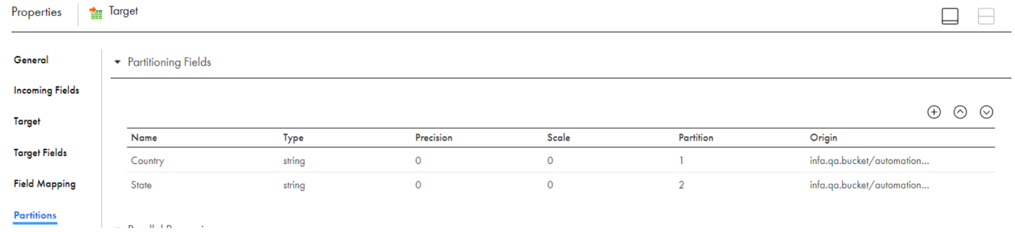
Select Partitioning Fields

To define the partitioning fields, select available fields, then click select.

infa.qa.bucket/automation/manual/devtest2/S3_Part_Tgts/tgt1//part_avro_tgt.avro

■ Name	Label
□ order_id	order_id
□ bill	bill
☑ Country	Country
☑ State	State

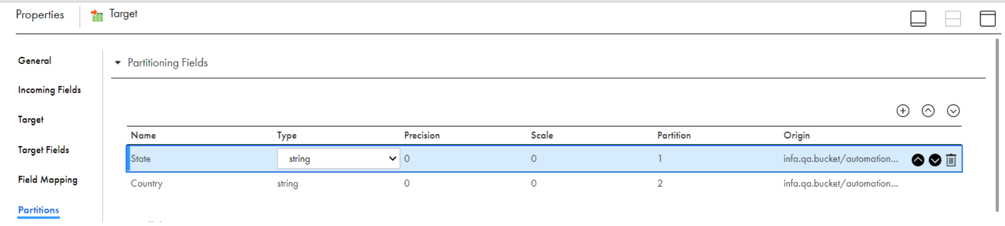
3. Click **Select**.
   The Partitions tab shows the partition columns that you selected:



Partitioning Fields

Name	Type	Precision	Scale	Partition	Origin
Country	string	0	0	1	infa.qa.bucket/automation...
State	string	0	0	2	infa.qa.bucket/automation...

Note: You can change the partition order using the up and down arrows as shown in the following image:



Partitioning Fields

Name	Type	Precision	Scale	Partition	Origin
State	string	0	0	1	infa.qa.bucket/automation...
Country	string	0	0	2	infa.qa.bucket/automation...

Rules and guidelines for reading from and writing to a partition folder

Consider the following rules and guidelines when you read from and write to a partition folder:

- You must import a directory that contains only partition folders and select the source type as **Directory** in the advanced source property.
- If you import a partition directory that does not have data, a validation error is encountered.
- If you import a partition directory that contains only files but no partition folders, a validation error is encountered.
- If you import a partition directory that has a partition folder but no files in the partition folder, a validation error is encountered.

- You can read data from or write data to partition folders with Avro, Parquet, and Orc files.
- The FileName field has 0 as the partition order.
- The partitioned directory that you select cannot have a partitioned column named FileName. The name is case insensitive.
- When you import an existing target object or create a new target object with a partition directory, the FileName field does not appear for the target objects. The FileName field appears only when you import the source objects.
- You can pushdown a Filter transformation on a partition column for an Amazon S3 source.
- When you pass a timestamp value in a partition column, the value gets encoded. For example, `03:26:01` gets encoded as `03%3A26%3A01`.
- When you import a directory that has a partition folder, the data type for the partition column is imported as a String.
- You cannot edit the data type for a partition column.
- You cannot use the **Edit Metadata** option with partition columns.
- You cannot use the **View Schema** option for a partition directory at source and target side.
- You cannot use the **Import from Schema File** option for partition directory at source because the schema file does not have information for partition columns.
- You cannot use the **Data Preview** option with partition columns.
- You cannot select the partition columns in a mapping task if the target object is parameterized.
- For **Create Target**, you can add partition fields and arrange the partition columns in an order.
- At **Create Target**, the **Label** column in the **Partitions** tab denotes the partition column name.
- When you import an Amazon S3 object that has partition columns, the partition fields are listed at the end of the list.
- If a partition column contains data that has more than 255 characters, the data is truncated and only 255 characters are written in the partition column.
- If a partition column name contains more than 74 characters, the name is truncated and only 74 characters are written in the partition column name.
- The value of the partition directory file path formed using the combination of the partition column name and the target file within the partition directory must not exceed 1024 characters. Otherwise, the mapping will fail.

Elastic mapping example

You work for one of the largest community college that maintains millions of records in their ongoing student database. The college has more than 10,000 faculty members teaching at 45 campuses and 700 locations across the globe. The college has a very large IT infrastructure and about 15 TB of information gets downloaded on daily basis from the Internet.

To avoid performance, scalability, and high cost challenges, the college plans to port its entire data from its operational data stores to Amazon S3 within a short span of time. Create an elastic mapping that runs on the elastic cluster to achieve faster performance when you read data from the operational data stores and write data to the Amazon S3 target.

1. In Data Integration, click **New** > **Mappings** > **Elastic Mapping** > **Create**.

   The **New Mapping** dialog box appears.

2. Enter a name, location, and description for the mapping.
3. On the Source transformation, specify a name and description in the general properties.
4. On the **Source** tab, perform the following steps to provide the source details to read data from the source:
   a. In the **Connection** field, select the required source connection.
   b. In the **Source Type** field, select the type of the source.
   c. In the **Object** field, select the required object.
   d. In the **Advanced Properties** section, provide the appropriate values.
5. On the **Fields** tab, map the source fields to the target fields.
6. On the Target transformation, specify a name and description in the general properties.
7. On the **Target** tab, perform the following steps to provide the target details to write data to the Amazon S3 target:
   a. In the **Connection** field, select the Amazon S3 V2 target connection.
   b. In the **Target Type** field, select the type of the target.
   c. In the **Object** field, select the required object.
   d. In the **Operation** field, select the required operation.
   e. In the **Advanced Properties** section, provide appropriate values for the advanced target properties.
8. Map the source and target.
9. Click **Save** > **Run** to validate the mapping.

   In Monitor, you can monitor the status of the logs after you run the task.

CHAPTER 5

Data type reference

This chapter includes the following topics:

- Data type reference overview, 56
- Flat file data types and transformation data types, 57
- Avro Amazon S3 file data types and transformation data types, 57
- JSON Amazon S3 file data types and transformation data types, 59
- ORC Amazon S3 file data types and transformation data types, 60
- Parquet Amazon S3 file data types and transformation data types, 61

Data type reference overview

Data Integration uses the following data types in 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.

Flat file data types 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
NUMBER	Decimal	Precision from 1 through 28 digits, scale from 0 through 28 digits
STRING	String	1 to 104,857,600 characters
NSTRING	Text	1 to 104,857,600 characters

Note: The NUMBER and NSTRING data types are supported only when you import the flat file by using **Import from Schema File** option.

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:

Avro Amazon S3 File Data Type	Transformation Data Type	Range and Description
Array	Array	Unlimited number of characters
Boolean	Integer	TRUE (1) or FALSE (0)
Date	Date/Time	January 1, 0001 to December 31, 9999.
Decimal	Decimal	For mappings: Precision 1 to 28 digits, scale 0 to 28. For elastic mappings: Precision 1 to 38 digits, scale 0 to 38.
Double	Double	Precision 15
Float	Double	Precision 15

Avro Amazon S3 File Data Type	Transformation Data Type	Range and Description
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
Map	Map	Unlimited number of characters
Null	Integer	-2,147,483,648 to 2,147,483,647 Precision 10, scale 0
Record	Struct	Unlimited number of characters
String	String	-1 to 104,857,600 characters
Union	Corresponding data type in a union of ["primitive_type|complex_type", "null"] or ["null", "primitive_type|complex_type"].	Dependent on primitive or complex data type.

The following table lists the Timestamp data type support for Avro file formats:

Timestamp Data type	Mapping	Elastic Mapping
Timestamp_micros	Yes	Yes
Timestamp_millis	Yes	No
Time_millis	Yes	No
Time_micros	Yes	No

Rules and guidelines for Avro data types

Consider the following rules and guidelines when you use Avro Amazon S3 file data types and transformation data types:

- When you run a mapping to write data to an Amazon S3 target, you cannot use the fixed type of the Avro data type.
- Specify the correct precision and scale in the source file. Otherwise, the decimal point is shifted when you write the source data to a target.
- The source file must have a timestamp value greater than or equal to 1900-01-01T00:00:00Z. Otherwise, the mapping fails with an error.

Enabling Date, Decimal, and Timestamp types

Perform the following steps to use the Date, Decimal, and Timestamp types for Avro data types before you run a mapping:

1. In Administrator, select the Secure Agent listed on the **Runtime Environments** tab.
2. Click **Edit**.
3. In the **System Configuration Details** section, select **Data Integration Service** as the service.
4. Edit the **INFA_DEBUG** property, and enter *-DEnableNewAvroDataTypes=true*.

   Note: To enter multiple flags, separate the flags with spaces.
5. Click **Save**.
6. Restart the Secure Agent.

Perform the following steps to use the Date, Decimal, and Timestamp types for Avro data types before you run an elastic mapping:

1. In Administrator, select **Elastic Clusters**.
2. Go to the existing configuration or create a new configuration.
3. In the **Advanced Configuration** section, go to **Custom Properties**.
4. Enter *-DEnableNewAvroDataTypes=true*.
5. Click **Save**.

JSON Amazon S3 file data types and transformation data types

JSON Amazon S3 file data types map to transformation data types that the Secure Agent uses to move data across platforms. You can use JSON file data types only in elastic mappings.

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

JSON Amazon S3 File Data Type	Transformation Data Type	Range and Description
Array	Array	Unlimited number of characters
Double	Double	Precision 15
Integer	Integer	-2,147,483,648 to 2,147,483,647 Precision of 10, scale of 0
Object	Struct	Unlimited number of characters
String	String	-1 to 104,857,600 characters

ORC Amazon S3 file data types and transformation data types

ORC 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 ORC Amazon S3 file data types that the Secure Agent supports and the corresponding transformation data types:

ORC Amazon S3 File Data Type	Transformation Data Type	Range and Description
BigInt	BigInt	-9223372036854775808 to 9,223,372,036,854,775,807
Boolean	Integer	TRUE (1) or FALSE (0)
Char	String	1 to 104,857,600 characters
Date	Date/Time	Jan 1, 1753 A.D. to Dec 31, 4712 A.D. (precision to microsecond)
Double	Double	Precision of 15 digits
Float	Double	Precision of 15 digits
Integer	Integer	-2,147,483,648 to 2,147,483,647
SmallInt	Integer	-32,768 to 32,767
String	String	1 to 104,857,600 characters
Timestamp	Date/Time	1 to 19 characters Precision 19 to 26, scale 0 to 6
TinyInt	Integer	-128 to 127
Varchar	String	1 to 104,857,600 characters

Parquet 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:

Parquet Amazon S3 File Data Type	Transformation Data Type	Range and Description
Boolean	Integer	TRUE (1) or FALSE (0)
Date	Date/Time	January 1,0001 to December 31,9999
Decimal	Decimal	For Mapping- Precision 1 to 28 digits, scale 0 to 28. For Elastic mapping- Precision 1 to 38 digits, scale 0 to 38.
Double	Double	Precision 15
Float	Double	Precision 15
Int32	Integer	-2,147,483,648 to +2,147,483,647
Int64	Bigint	-9,223,372,036,854,775,808 to +9,223,372,036,854,775,807 8-byte signed integer
Int96	Binary	12-byte signed integer
Map	Map	Unlimited number of characters.
Struct	Struct	Unlimited number of characters.
String	String	-1 to 104,857,600 characters
Time	Date/Time	Time of the day. Precision to microsecond.
Timestamp	Date/Time	January 1,0001 00:00:00 to December 31,9999 23:59:59.997997. Precision to microsecond.
group(LIST)	Array	Unlimited number of characters.

Note: Specify the correct precision and scale in the source file. Otherwise, the decimal point is shifted when you write the source data to a target.

Amazon S3 V2 Connector supports the Parquet file type **Date**, **Decimal**, **Time**, and **Timestamp** only on Cloudera 6.1 distribution.

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

Parquet Timestamp Data Type Support

The following table lists the Timestamp data type support for Parquet file format:

Timestamp Data type	Mapping	Elastic Mapping
Timestamp_micros	Yes	No
Timestamp_millis	Yes	No
Time_millis	Yes	No
Time_micros	Yes	No
int96	Yes	Yes
Date	Yes	Yes

Unsupported Parquet Data Types

The Secure Agent does not support the following Parquet data types:

- Timestamp_nanos
- Time_nanos
- Timestamp_tz

CHAPTER 6

Troubleshooting

This chapter includes the following topics:

- Troubleshooting overview, 63
- Troubleshooting for Amazon S3 V2 Connector, 63
- Troubleshooting FAQ, 64

Troubleshooting overview

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

Troubleshooting for Amazon S3 V2 Connector

Java heap size configuration

This section describes the errors that you might encounter if the JVM options in the Secure Agent is not configured accordingly to read large amount of files.

"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 **Administrator** > **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.

"[ERROR] java.lang.OutOfMemoryError: Java heap space" occurs when you run a mapping to write a file of size 1.4 GB or higher and select Informatica Encryption as the encryption type.

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 **Administrator** > **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 **-Xmx8046m**.
6. Restart the Secure Agent manually.

FFParserRetainNullString custom property

When you read from a csv file that contains a string named `null`, the task does not write any data to the target. To resolve this issue, perform the following tasks and configure the custom property `FFParserRetainNullString`:

1. Select **Administrator** > **Runtime Environments**.
2. On the **Runtime Environments** page, select the Secure Agent.
3. In the upper-right corner, click **Edit**.
4. In the **Custom Configuration Details** section, select the **Type** as **Tomcat JRE** for the Data Integration Service.
5. Enter the **Name** as **FFParserRetainNullString** and the **Value** as **true**.
6. Click **Save**.

Troubleshooting FAQ

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

How to solve the following error that occurs when you use the Create Target option and do not set the formatting options in the target property of a mapping: "com.informatica.powercenter.sdk.SDKException:Error during binary data write: java.lang.String cannot be cast to [B"

For information about the issue, see https://kb.informatica.com/solution/23/Pages/69/568954.aspx?myk=568954

How can I read a JSON file using Amazon S3 V2 Connector?

For information about reading a JSON file using Amazon S3 V2 Connector, see https://kb.informatica.com/h2l/HowTo%20Library/1/1271-HowtoReadaJSONFileUsingAmazonS3V2Connector-H2L.pdf

Can I run an elastic mapping to read or write multi-level JSON files?

Yes. For more information about running an elastic mapping to read or write multi-level JSON files, see https://kb.informatica.com/faq/7/Pages/23/574694.aspx?myk=574694.

INDEX

A

administration
- IAM authentication 9
- minimal Amazon S3 bucket policy 9

Amazon S3
- specifying targets 50

Amazon S3 and transformation
- data types 57

Amazon S3 Connector
- introduction 8

Amazon S3 V2
- handling multiple files 25
- connection properties 16
- directory source 24
- file formatting options 42
- lookups 42
- sources 22
- supported object types 8
- supported task types 8
- targets 27

Amazon S3 V2 connection
- overview 15

Amazon S3 V2 Connector
- administration 8
- overview 7

Amazon S3 V2 objects
- mapping tasks 45
- mappings 35

Amazon S3 V2 sources
- client-side encryption 23
- mapping tasks 45
- mappings 35

Amazon S3 V2 target
- mappings 38

Amazon S3 V2 targets
- mapping tasks 47

C

Cloud Application Integration community
- URL 5

Cloud Developer community
- URL 5

connections
- Amazon S3 V2 16

create target
- adding time stamps 50
- target file parameterization 50

creating an elastic mapping
- rules and guidelines 12, 44

D

data compression
- sources and targets 33

data encryption
- Informatica encryption 24
- sources 23
- targets 28

Data Integration community
- URL 5

data type reference
- overview 56

distribution column 30

E

elastic mapping
- example 54

I

Informatica Global Customer Support
- contact information 6

Informatica Intelligent Cloud Services
- web site 5

J

Java heap size
- configuration 63

JSON Amazon S3 file data types
- transformation data types 59

L

Linux
- configuring proxy settings 21

M

maintenance outages 6

O

object tag
- rules and guidelines 32

object tags 31, 32

ORC file data types
- transformation data types 60

overwriting
    existing files 29

P

parameterization through a parameter file 51
partitioning
    Amazon S3 V2 sources 26
Partitioning
    Amazon S3 V2 targets 30
proxy settings
    configuring on Linux 21
    configuring on Windows 20
    JVMOptions 20

R

reading compressed flat file 34
rules and guidelines
    setting formatting options 44

S

source
    FileName field 26
status
    Informatica Intelligent Cloud Services 6
system status 6

T

target
    FileName field 31
temporary security credentials
    policy 11
troubleshooting
    Amazon S3 V2 Connector 64
trust site
    description 6

U

upgrade notifications 6

W

web site 5
Windows
    configuring proxy settings 20