Informatica®

Informatica® Cloud Data Integration

Amazon S3 V2 Connector

Informatica Cloud Data Integration Amazon S3 V2 Connector
November 2023

© Copyright Informatica LLC 2017, 2024

This software and documentation are provided only under a separate license agreement containing restrictions on use and disclosure. No part of this document may be reproduced or transmitted in any form, by any means (electronic, photocopying, recording or otherwise) without prior consent of Informatica LLC.

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation is subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License.

Informatica, the Informatica logo, Informatica Cloud, and PowerCenter are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at https://www.informatica.com/trademarks.html. Other company and product names may be trade names or trademarks of their respective owners.

Portions of this software and/or documentation are subject to copyright held by third parties. Required third party notices are included with the product.

See patents at https://www.informatica.com/legal/patents.html.

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: 2024-04-22

Table of Contents

Preface 6

Informatica Resources 6

- Informatica Documentation 6
- Informatica Intelligent Cloud Services web site 6
- Informatica Intelligent Cloud Services Communities 6
- Informatica Intelligent Cloud Services Marketplace 7
- Data Integration connector documentation 7
- Informatica Knowledge Base 7
- Informatica Intelligent Cloud Services Trust Center 7
- Informatica Global Customer Support 7

Chapter 1: Introduction to Amazon S3 V2 Connector 8

Amazon S3 V2 Connector assets 8

Introduction to Amazon S3 8

Chapter 2: Connections for Amazon S3 V2 10

Prepare for authentication 10

- Create a minimal Amazon IAM policy 10
- IAM authentication 11
- AssumeRole via EC2 role and IAM user 12
- Credential profile file authentication 14

Connect to Amazon S3 14

- Before you begin 14
- Connection details 15
- Authentication types 15
- Advanced settings 23

Private communication with Amazon S3 24

Server-side encryption with KMS 24

Client-side encryption with serverless runtime environment 25

SSE-KMS encryption for mappings in advanced mode 26

Proxy server settings 26

- Bypass the proxy server 27

Rules and guidelines for AssumeRole via IAM user 28

Rules and guidelines for AWS regions 29

Rules and guidelines for S3 compatible storage 29

Chapter 3: Amazon S3 V2 sources and targets 30

Amazon S3 V2 sources 30

- Data encryption in Amazon S3 V2 sources 31
- Source types in Amazon S3 V2 sources 32

Reading from multiple files. . . . 33
Incrementally loading files. . . . 34
Wildcard characters. . . . 35
Recursively read files from directories. . . . 36
Source partitioning. . . . 36
Reading source objects path. . . . 37
SQL ELT optimization. . . . 38
Amazon S3 V2 targets. . . . 38
Data encryption in Amazon S3 V2 targets. . . . 38
Overwriting existing files. . . . 40
Target partitioning. . . . 40
Incremental write to partition directory. . . . 41
Distribution column. . . . 41
Writing to multiple target objects. . . . 41
Object tag. . . . 42
Directory-level partitioning. . . . 43
Rules and guidelines for reading from and writing to a partition folder. . . . 45
Data compression in Amazon S3 V2 sources and targets. . . . 47
Reading a compressed flat file. . . . 48
Reading a compressed JSON file. . . . 48
Fixed-width file formats. . . . 49

Chapter 4: Mappings and mapping tasks with Amazon S3 V2. . . . 50

Amazon S3 V2 objects in mappings. . . . 50
Amazon S3 V2 sources in mappings. . . . 51
Amazon S3 V2 targets in mappings. . . . 54
Amazon S3 V2 lookups. . . . 57
File formatting options. . . . 57
Specifying a target. . . . 60
Amazon S3 V2 parameterization. . . . 62
Rules and guidelines for mappings in advanced mode. . . . 63
Mapping in advanced mode example. . . . 64

Chapter 5: Migrating a mapping. . . . 67

Use the same object path for the migrated mapping. . . . 67
Rules and guidelines for the same object path. . . . 67
Use a different object path for the migrated mapping. . . . 68
Migration options. . . . 68
Rules and guidelines. . . . 69

Chapter 6: Upgrading to Amazon S3 V2 Connector. . . . 71

Upgrade the connection type. . . . 71
Connection switching example. . . . 72

Advanced properties retained after the switch. 74

Chapter 7: Data type reference. 75

Flat file data types and transformation data types. 75

Avro Amazon S3 file data types and transformation data types. 76

Enabling Date, Decimal, and Timestamp types. 77

JSON Amazon S3 file data types and transformation data types. 78

ORC Amazon S3 file data types and transformation data types. 78

Parquet Amazon S3 file data types and transformation data types. 79

Chapter 8: Troubleshooting. 82

Troubleshooting for Amazon S3 V2 Connector. 82

Troubleshooting FAQ. 83

Index. 85

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 mappings, mapping tasks, dynamic mapping tasks, and data transfer tasks in Cloud Data Integration. Learn how to push down the transformation logic for processing to the Amazon Redshift database.

Informatica Resources

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

Informatica Documentation

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

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

Informatica Intelligent Cloud Services web site

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

Informatica Intelligent Cloud Services Communities

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

Access the Informatica Intelligent Cloud Services Community at:

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

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

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

Informatica Intelligent Cloud Services Marketplace

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

https://marketplace.informatica.com/

Data Integration connector documentation

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

Informatica Knowledge Base

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

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

Informatica Intelligent Cloud Services Trust Center

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

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

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

To subscribe, on the Informatica Intelligent Cloud Services Status page, click **SUBSCRIBE TO UPDATES**. You can choose to receive notifications sent as emails, SMS text messages, webhooks, RSS feeds, or any combination of the four.

Informatica Global Customer Support

You can contact a Global Support Center through the Informatica Network or by telephone.

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

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

CHAPTER 1

Introduction to Amazon S3 V2 Connector

You can use Amazon S3 V2 Connector to connect to Amazon S3 from Data Integration.

You can create an Amazon S3 V2 connection and use the connection in mappings or mapping tasks. You can switch the mapping to advanced mode to include transformations and functions that enable advanced functionality. The advanced cluster can be a self-service cluster, a local cluster, or hosted on Amazon Web Services. You can read and write primitive data types for Avro, Parquet, JSON and ORC files.

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 mappings in advanced mode. You can read and write hierarchical data types only for Avro, Parquet and JSON for mappings in advanced mode.

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 Connector assets

Create assets in Data Integration to integrate data using Amazon S3 V2 Connector.

When you use Amazon S3 V2 Connector, you can include the following Data Integration assets:

- Data transfer task
- Dynamic mapping task
- Mappings
- Mapping task

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

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.

CHAPTER 2

Connections for Amazon S3 V2

Create an Amazon S3 V2 connection to securely read data from or write data to Amazon S3.

You can use an Amazon S3 V2 connection to specify sources and targets in mappings and mapping tasks.

Prepare for authentication

You can configure multiple authentication types to access Amazon S3.

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

- Basic authentication requires access key and secret key values from your AWS account.
- IAM authentication requires only the folder path to the Amazon S3 objects. The EC2 role must have access to the folder.
- AssumeRole via EC2 role authentication requires you to enable the EC2 role to assume another IAM role specified in the IAM Role ARN option.
- AssumeRole via IAM user authentication requires the access key and secret key values of the IAM user and the ARN of the IAM role.
- Credential profile file authentication requires the credential profile file path and profile name.
- Federated user single sign-on authentication requires the user name and password of the federated user, IdP SSO URL, ARN of the SAML identity provider, and ARN of the IAM role assumed by the federated user. You can only use ADFS 3.0 (IDP) for SSO.

Create a minimal Amazon IAM policy

You can configure an IAM policy through the AWS console. Use AWS IAM authentication to securely control access to Amazon S3 resources.

Use the following minimum required policies for users to read data from an Amazon S3 bucket:

- GetObject
- ListBucket

Use the following minimum required policies for users to write data to an Amazon S3 bucket:

- PutObject
- GetObject
- DeleteObject

- ListBucket
- ListBucketMultipartUploads. Applicable only for mappings in advanced mode.

The following sample policy shows the minimal Amazon IAM policy to write data to an Amazon S3 bucket:

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

For bucket level access in advanced mode, you need to provide the `AllowListBucketMultipartUploads` permission at the bucket level in addition to the `ListBucketMultipartUploads` permission.

The following sample policy shows the minimal Amazon IAM policy to access the S3 bucket at the bucket level in advanced mode:

```
  {
            "Sid": "AllowListBucketMultipartUploads",
            "Action": [
                "s3:ListBucketMultipartUploads"
            ],
            "Effect": "Allow",
            "Resource": [
                "arn:aws:s3:::infa.qa.minimum.access.bucket"
            ]
        },
```

For mappings in advanced mode, you can use different AWS accounts within the same AWS region. Make sure that the Amazon IAM policy confirms access to the AWS accounts used in the mapping.

IAM authentication

You can configure IAM authentication when the Secure Agent runs on an Amazon Elastic Compute Cloud (EC2) system. If you have an EC2 instance, and do not want to specify the keys or use the IAM role ARN, then assign the minimum policy to the EC2 with access to the S3 bucket.

When you use a serverless runtime environment, you cannot configure IAM authentication.

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.

Perform the following steps to configure IAM authentication on EC2:

1. Create a minimal Amazon IAM 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 IAM 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.

AssumeRole via EC2 role and IAM user

You can configure AssumeRole via EC2 role or IAM user to connect to Amazon S3.

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

When you use AssumeRole via EC2 role or IAM user authentication, 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 EC2 role or IAM user as a trusted entity allowing the EC2 role or IAM user 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 EC2 role or 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 EC2 role or IAM user.

AssumeRole via EC2 role

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 Amazon S3 V2 connection properties.

The Amazon EC2 role can 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 before you use AssumeRole via EC2 role:

- 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.

AssumeRole via IAM user

To use AssumeRole via IAM user, 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 access to the Amazon S3 bucket when the Amazon S3 bucket is in same or 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"
        }
      }
    }
  ]
```

AssumeRole 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>" }
}
```

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

```
{
"Version":"2012-10-17", "Statement":{ "Effect":"Allow", "Action":"sts:AssumeRole",
"Resource":"arn:aws-cn: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 " }
```

Credential profile file authentication

You can provide the credentials required to establish the connection with Amazon S3 through the credential profile file .

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`.

The following sample shows a 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` are the AWS access key and secret key used as part of credentials to authenticate the user.

The `aws_session_token` is the 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.

Connect to Amazon S3

Let's configure the Amazon S3 connection properties to connect to Amazon S3.

Before you begin

Before you get started, you'll need to get information from your Amazon S3 account based on the authentication type that you want to configure.

Check out "Prepare for authentication" on page 10 to learn more about the authentication prerequisites.

Connection details

The following table describes the basic connection properties:

Property	Description
Connection Name	Name of the connection. Each connection name must be unique within the organization. Connection names can contain alphanumeric characters, spaces, and the following special characters: _ . + -, Maximum length is 255 characters.
Description	Description of the connection. Maximum length is 4000 characters.
Type	Amazon S3 V2
Use Secret Vault	Stores sensitive credentials for this connection in the secrets manager that is configured for your organization. This property appears only if secrets manager is set up for your organization. When you enable the secret vault in the connection, you can select which credentials that the Secure Agent retrieves from the secrets manager. If you don't enable this option, the credentials are stored in the repository or on a local Secure Agent, depending on how your organization is configured. For information about how to configure and use a secrets manager, see "Secrets manager configuration" in the Administrator help.
Runtime Environment	The name of the runtime environment where you want to run tasks. Select a Secure Agent, Hosted Agent, or serverless runtime environment.

Authentication types

You can configure basic, AWS Identity and Access Management (IAM), temporary security credentials, assume role for EC2, credential profile file, and federated user single sign-on authentication types to access Amazon S3.

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

Basic authentication

Basic authentication requires access key and secret key values from your AWS account.

The following table describes the basic connection properties for basic authentication:

Property	Description
Access Key	Access key to access the Amazon S3 bucket.
Secret Key	Secret key to access the Amazon S3 bucket. The secret key is associated with the access key and uniquely identifies the account.

Property	Description
Folder Path	Bucket name or complete folder path to the Amazon S3 objects. For example, `<bucket name>/<my folder name>`
Region Name	The AWS region of the bucket that you want to access. Select one of the following regions: - Africa(Cape Town) - Asia Pacific(Mumbai) - Asis Pacific(Jakarta) - Asia Pacific (Osaka) - 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(Bejing) - China(Ningxia) - EU(Ireland) - EU(Frankfurt) - EU(London) - EU(Milan) - EU(Paris) - EU(Stockholm) - South America(Sao Paulo) - Middle East(Bahrain) - Middle East(UAE) - US East(N. Virginia) - US East(Ohio) - US ISO East - US ISOB East(Ohio) - US ISO West - US West(N. California) - US West(Oregon) Default is US East(N. Virginia).

IAM authentication

IAM authentication requires only the folder path to the Amazon S3 objects. The EC2 role must have access to the folder. :

The following table describes the basic connection properties for AWS IAM authentication:

Property	Description
Folder Path	Bucket name or complete folder path to the Amazon S3 objects. For example, `<bucket name>/<my folder name>`
Region Name	The AWS region of the bucket that you want to access. Select one of the following regions: - Africa(Cape Town) - Asia Pacific(Mumbai) - Asis Pacific(Jakarta) - Asia Pacific (Osaka) - 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(Bejing) - China(Ningxia) - EU(Ireland) - EU(Frankfurt) - EU(London) - EU(Milan) - EU(Paris) - EU(Stockholm) - South America(Sao Paulo) - Middle East(Bahrain) - Middle East(UAE) - US East(N. Virginia) - US East(Ohio) - US ISO East - US ISOB East(Ohio) - US ISO West - US West(N. California) - US West(Oregon) Default is US East(N. Virginia).

AssumeRole via EC2 role authentication

AssumeRole via EC2 role authentication requires you to enable the EC2 role to assume another IAM role specified in the IAM Role ARN option:

The following table describes the basic connection properties for AssumeRole via EC2 role authentication:

Property	Description
Folder Path	Bucket name or complete folder path to the Amazon S3 objects. For example, `<bucket name>/<my folder name>`
Region Name	The AWS region of the bucket that you want to access. Select one of the following regions: - Africa(Cape Town) - Asia Pacific(Mumbai) - Asis Pacific(Jakarta) - Asia Pacific (Osaka) - 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(Bejing) - China(Ningxia) - EU(Ireland) - EU(Frankfurt) - EU(London) - EU(Milan) - EU(Paris) - EU(Stockholm) - South America(Sao Paulo) - Middle East(Bahrain) - Middle East(UAE) - US East(N. Virginia) - US East(Ohio) - US ISO East - US ISOB East(Ohio) - US ISO West - US West(N. California) - US West(Oregon) Default is US East(N. Virginia).
IAM Role ARN	The Amazon Resource Name (ARN) of the AWS Identity and Access Management (IAM) role assumed by the user to use the dynamically generated temporary security credentials. Enter the ARN value if you want to use the temporary security credentials to access AWS resources. **Note:** Even if you remove the IAM role that enables the agent to access the Amazon S3 bucket and create a connection, the test connection is successful. For more information about how to get the ARN of the IAM role, see the AWS documentation.
External ID	Provides a more secure access to the Amazon S3 bucket when the Amazon S3 bucket is in a different AWS account.
Use EC2 Role to Assume Role	Enables the EC2 role to assume another IAM role specified in the IAM Role ARN option. By default, this property is not selected. **Note:** The EC2 role must have a policy attached with permissions to assume an IAM role from the same or different account.

AssumeRole via IAM user authentication

AssumeRole via IAM user authentication requires the access key and secret key values of the IAM user and the ARN of the IAM role.

The following table describes the basic connection properties for AssumeRole via IAM user authentication:

Property	Description
Access Key	Access key to access the Amazon S3 bucket.
Secret Key	Secret key to access the Amazon S3 bucket. The secret key is associated with the access key and uniquely identifies the account.
Folder Path	Bucket name or complete folder path to the Amazon S3 objects. For example, `<bucket name>/<my folder name>`
Region Name	The AWS region of the bucket that you want to access. Select one of the following regions: - Africa(Cape Town) - Asia Pacific(Mumbai) - Asis Pacific(Jakarta) - Asia Pacific (Osaka) - 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(Bejing) - China(Ningxia) - EU(Ireland) - EU(Frankfurt) - EU(London) - EU(Milan) - EU(Paris) - EU(Stockholm) - South America(Sao Paulo) - Middle East(Bahrain) - Middle East(UAE) - US East(N. Virginia) - US East(Ohio) - US ISO East - US ISOB East(Ohio) - US ISO West - US West(N. California) - US West(Oregon) Default is US East(N. Virginia).

Property	Description
IAM Role ARN	The Amazon Resource Name (ARN) of the AWS Identity and Access Management (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 agent to access the Amazon S3 bucket and create a connection, the test connection is successful. For more information about how to get the ARN of the IAM role, see the AWS documentation.
External ID	Provides a more secure access to the Amazon S3 bucket when the Amazon S3 bucket is in a different AWS account.

Credential profile file authentication

Credential profile file authentication requires the credential profile file path and profile name.

The following table describes the basic connection properties for credential profile file authentication:

Property	Description
Folder Path	Bucket name or complete folder path to the Amazon S3 objects. For example, `<bucket name>/<my folder name>`
Region Name	The AWS region of the bucket that you want to access. Select one of the following regions: - Africa(Cape Town) - Asia Pacific(Mumbai) - Asis Pacific(Jakarta) - Asia Pacific (Osaka) - 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(Bejing) - China(Ningxia) - EU(Ireland) - EU(Frankfurt) - EU(London) - EU(Milan) - EU(Paris) - EU(Stockholm) - South America(Sao Paulo) - Middle East(Bahrain) - Middle East(UAE) - US East(N. Virginia) - US East(Ohio) - US ISO East - US ISOB East(Ohio) - US ISO West - US West(N. California) - US West(Oregon) Default is US East(N. Virginia).
Other Authentication Type[1]	Use the Credential Profile File Authentication to access the Amazon S3 credentials from a credential file that contains the access key and secret key. Select one the following authentication types: - NONE - Credential Profile File Authentication Enter the credential profile file path and profile name to connect to 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[1]	The credential profile file path. If you don't enter the credential profile path, the Secure Agent uses the credential profile file present in the following default location in your home directory: `~/.aws/credentials`

Property	Description
Profile Name[1]	Name of the profile in the credential profile file used to get the credentials. If you don't enter the profile name, the credentials from the default profile in the credential profile file are used.
[1] Applies only to mappings.	

Federated SSO authentication

Federated user single sign-on authentication requires the user name and password of the federated user, IdP SSO URL, ARN of the SAML identity provider, and ARN of the IAM role assumed by the federated user. You can only use ADFS 3.0 (IDP) for SSO.

The following table describes the basic connection properties for federated single sign-on authentication:

Property	Description
Folder Path	Bucket name or complete folder path to the Amazon S3 objects. For example, `<bucket name>/<my folder name>`
Region Name	The AWS region of the bucket that you want to access. Select one of the following regions: - Africa(Cape Town) - Asia Pacific(Mumbai) - Asis Pacific(Jakarta) - Asia Pacific (Osaka) - 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(Bejing) - China(Ningxia) - EU(Ireland) - EU(Frankfurt) - EU(London) - EU(Milan) - EU(Paris) - EU(Stockholm) - South America(Sao Paulo) - Middle East(Bahrain) - Middle East(UAE) - US East(N. Virginia) - US East(Ohio) - US ISO East - US ISOB East(Ohio) - US ISO West - US West(N. California) - US West(Oregon) Default is US East(N. Virginia).

Property	Description
Federated SSO IdP[1]	SAML 2.0-enabled identity provider for the federated user single sign-on to use with the AWS account. You can only use ADFS 3.0 (IDP) for SSO. Select **None** if you don't want to use federated user single sign-on. **Note:** Federated user single sign-on is not applicable to mappings in advanced mode.
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.
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.

Advanced settings

The following table describes the advanced connection properties:

Property	Description
S3 Account Type	The type of the Amazon S3 account. Select from the following options: - Amazon S3 Storage. Enables you to use the Amazon S3 services. - S3 Compatible Storage. Enables you to use the endpoint for a third-party storage provider such as Scality RING or MinIO. Default is Amazon S3 storage.
REST Endpoint	The S3 storage endpoint required for S3 compatible storage. Enter the S3 storage endpoint in HTTP or HTTPs format. For example, http://s3.isv.scality.com.
S3 VPC Endpoint Type[1]	The VPC endpoint type for Amazon S3. You can enable private communication with Amazon S3 by selecting a VPC endpoint. Select one of the following options: - None - Gateway Endpoint - Interface Endpoint Default is None.
Endpoint DNS Name for Amazon S3[1]	The DNS name for the Amazon S3 interface endpoint. Enter the DNS name in the following format: `bucket.<DNS name of the interface endpoint>`
STS VPC Endpoint Type[1]	The VPC endpoint type for AWS STS. Applicable when you select the S3 VPC interface endpoint. When you select **IAM Role ARN** or **Federated SSO IdP**, configure the STS VPC endpoint.

Property	Description
Endpoint DNS Name for AWS STS[1]	The DNS name for the AWS STS interface endpoint.
KMS VPC Endpoint Type[1]	Applicable when you select the interface endpoint. The VPC endpoint type for the AWS KMS. When you select **Customer Master Key ID**, configure the KMS VPC endpoint.
Endpoint DNS Name for AWS KMS[1]	The DNS name for the AWS KMS interface endpoint.
Master Symmetric Key	A 256-bit AES encryption key in the Base64 format when you use client-side encryption. You can generate a key using a third-party tool.
Customer Master Key ID	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 available for mappings in advanced mode. You must generate the customer master key for the same region where the 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.
[1] Applies only to mappings.	

Private communication with Amazon S3

You can enable private communication with Amazon S3 by configuring a gateway endpoint or interface endpoint on AWS console and in the Amazon S3 V2 connection.

You can configure Amazon S3 V2 Connector to establish private communication with Amazon S3 without exposing your traffic to the public internet. To access Amazon S3, ensure that the Secure Agent is a part of the subnet in the AWS Virtual Private Cloud (VPC). AWS S3 VPC endpoint enables an S3 request to be routed to the Amazon S3 service, without having to connect a subnet to an internet gateway. You can create an interface endpoint or a gateway endpoint.

For more information, see
Configuring private communication with Amazon S3 using the Amazon S3 V2 Connector.

Server-side encryption with KMS

To use the customer master key managed by AWS Key Management Service (AWS KMS) and enable the encryption with KMS, you need to 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": ["arn:aws:kms:region:account:key/<KMS_key>"]
    }
  ]
   }
```

When you configure KMS and access an Amazon S3 endpoint in the China region, use the following sample policy:

```
{
"Version":"2012-10-17", "Statement":[{ "Effect":"Allow", "Action":
[ "kms:GenerateDataKey", "kms:DescribeKey", "kms:Encrypt", "kms:Decrypt",
"kms:ReEncrypt*" ], "Resource": ["arn:aws-cn:kms:region:account:key/<KMS_key>"]
    }
  ]
   }
```

Client-side encryption with serverless runtime environment

You can use the serverless runtime environment with Amazon S3 V2 Connector to configure client-side encryption.

Before you configure client-side encryption using the serverless runtime environment, you must configure the .yml serverless configuration file.

Configure the .yml serverless configuration file

Perform the following steps to configure the .yml serverless configuration file in the serverless runtime environment so that Amazon S3 V2 Connector can use client-side encryption:

1. Copy the following code snippet to a text editor:

```
version: 1
agent:
  agentAutoApply:
    general:
    jdk:
         security:
         policyJars:
           - local_policy.jar
        - US_export_policy.jar
```

2. Ensure that the syntax and indentations are valid, and then save the file as `serverlessUserAgentConfig.yml` in the following AWS or Azure location:

`<Supplementary file location>/serverless_agent_config`

When the .yml file runs, the policy jars are copied from the AWS or Azure location to the serverless agent directory.

3. After you update the .yml configuration file, redeploy the serverless runtime environment.

Specify the master symmetric key in the connection properties and the client-side encryption type in the advanced source and target properties.

SSE-KMS encryption for mappings in advanced mode

To enable encryption with KMS, create an AWS Key Management Service (AWS KMS) policy and an AWS KMS-managed customer master key.

To use SSE-KMS encryption for mappings in advanced mode, perform one of the following tasks:

- To use the credentials from the `~/.aws/credentials` location, create the master instance profile and the worker instance profile in AWS, attach the KMS policy to the worker profile, and specify the profiles in the cluster configuration.
- Use the Secure Agent on Amazon EC2, create the master instance profile and the worker instance profile in AWS, and attach the KMS policy to the worker profile.
- Use the Secure Agent on Amazon EC2, use the default IAM role, and attach the KMS policy to the Secure Agent role.

Proxy server settings

If your organization uses an outgoing proxy server to connect to the Internet, the Secure Agent connects to Informatica Intelligent Cloud Services through the proxy server.

You can configure the Secure Agent to use the proxy server on Windows and Linux.

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

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

To configure the proxy settings for the Secure Agent, perform the following tasks:

- For instructions, see "Configure the proxy settings on Windows" or "Configure the proxy settings on Linux" in *Getting Started* in the Data Integration help .
- Configure the JVM options for the DTM in the Secure Agent properties. For instructions, see the Proxy server settings Knowledge Base article.
- Configure the proxy server properties in the proxy.ini file.

When you use a serverless runtime environment, you cannot use a proxy server to connect to Informatica Intelligent Cloud Services.

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.

Bypass the proxy server

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.

However, if you want to exclude certain IP addresses and host names from the proxy, you can bypass the proxy. Set the *InfaAgent.NonProxyHost* property in the `proxy.ini` file and the *-Dhttp.nonProxyHosts* property in the JVM options of the Secure Agent properties and include the IP addresses and host names that you want to exclude.

The following table shows the proxy setting for configuring proxy using the `proxy.ini` file and the JVM options.

Proxy configuration	Proxy Flag Setting

Proxy.ini	InfaAgent.NonProxyHost=localhost|**<your_bucket_name>.s3.**|127.*|[\:\:1]
	For example, to bypass a single S3 Bucket `iam.qa.bucket`, use the following proxy setting:
	`InfaAgent.NonProxyHost=localhost|`**`iam.qa.bucket.s3.`**`|127.*|[\:\:1]`
	To bypass all S3 buckets, use the following proxy setting:
	`InfaAgent.NonProxyHost=localhost|`**`*.s3.*`**`|127.*|[\:\:1]`

JVM option	-Dhttp.nonProxyHosts=localhost|**<your_bucket_name>.s3.**|127.* |[\:\:1]
	For example, to bypass a single S3 Bucket, `iam.qa.bucket`, use the following proxy setting:
	`-Dhttp.nonProxyHosts=localhost|`**`iam.qa.bucket.s3.`**`|127.*|[\:\:1]`
	To bypass all S3 buckets, use the following proxy setting:
	`-Dhttp.nonProxyHosts=localhost|`**`*.s3.*`**`|127.*|[\:\:1]`

Bypass the proxy server in advanced mode

To bypass the proxy server, you must update the `NonProxyHost` in `proxy.ini` file. You can set the property in the agent core path to configure the `NonProxyHost` in the advanced cluster configuration.

To bypass the proxy at the Amazon S3 endpoint, perform the following steps:

1. Edit the `proxy.ini` file and set the property in the `NonProxyHost` with the cluster region.
2. Enter the appropriate region name in the property in the following format:

   ```
   InfaAgent.NonProxyHost=localhost|127.*|[\:\:1]|
   169.254.169.254|.<REGION_NAME>.elb.amazonaws.com|*.<REGION_NAME>.elb.amazonaws.com
   ```

   For the US West region, the following example shows how you can update the non-proxy host for a specific region in the `proxy.ini` file:

   ```
   InfaAgent.NonProxyHost=localhost|127.*|[\:\:1]|169.254.169.254|.us-
   west-2.elb.amazonaws.com|*.us-west-2.elb.amazonaws.com|s3.us-west-2.amazonaws.com|
   *.s3.us-west-2.amazonaws.com|s3.amazonaws.com
   ```

3. After you edit the `proxy.ini` file, you must set the property `ccs.enable.storage.proxy.settings` to false in the runtime properties of the advanced cluster. Perform the following steps to set the property:
   a. Go to the **Administrator** service.
   b. In the **Advanced Clusters** page, select the name of the configuration that you want to edit from the list of advanced configurations.

c. Set the property `ccs.enable.storage.proxy.settings` to false and save the cluster configuration in the **Runtime Properties** for the particular cluster.
The following image shows the cluster runtime properties:



Rules and guidelines for AssumeRole via IAM user

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 S3 V2 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 mappings in advanced mode, 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. 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.

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.
- 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).

Rules and guidelines for S3 compatible storage

Consider the following rules and guidelines when you configure S3 compatible storage in an Amazon S3 V2 connection:

- You can only configure basic authentication when you use S3 compatible storage.
- You cannot configure SSE-KMS encryption for the Scality RING S3 compatible storage. You cannot configure SSE and SSE-KMS encryption for MinIO S3 compatible storage.
- You cannot configure SQL ELT optimization to load data from Amazon S3 sources to Amazon Redshift.

CHAPTER 3

Amazon S3 V2 sources and targets

You can configure the Amazon S3 V2 sources and targets to read from and write to Amazon S3.

Amazon S3 V2 sources

You can use an Amazon S3 V2 object as a source in a mapping or amapping 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 encryption type supported for various file types:

Encryption Type	Avro File	Binary File[1]	Flat	JSON File[2]	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]Doesn't apply to mappings in advanced mode.
[2]Applies to mappings in advanced mode.

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 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. Copy the `local_policy.jar` and `US_export_policy.jar` files from the following directory:
   `<Secure Agent installation directory>/jdk/jre/lib/security/policy/unlimited/`
3. Paste the files in the following directory:
   `<Secure Agent installation directory>/jdk/jre/lib/security/`
4. Restart the Secure Agent.

Server-side encryption for Amazon S3 V2 sources

Server-side encryption is a technique to encrypt data using Amazon S3-managed encryption keys. Server-side encryption with KMS is a technique to encrypt data using the AWS KMS-managed customer master key.

Server-side encryption

To read a server-side encrypted file, select the encrypted file in the Amazon S3 V2 source.

Server-side encryption with KMS

To read a server-side encrypted file with KMS, specify the AWS KMS-managed customer master key in the **Customer Master Key ID** connection property and select the encrypted file in the Amazon S3 V2 source.

Note: You do not need to specify the encryption type in the advanced source properties.

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.

   Note: For Informatica Encryption in the advanced cluster, you must install the Secure Agent on the Amazon EC2 machine.

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 run time. When you select the **Source Type** option as **Directory**, the value of **File Name** is honored only when you use wildcard characters to specify the folder path or file name, or recursively read files from directories.

For the 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 parsed only when you recursively read files from directories.

Reading from multiple files

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

You can use the following types of manifest files:

- Custom manifest file
- Amazon Redshift manifest file

Custom manifest file

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/dir3/file_1.csv",
                "dir1/dir2/dir3/file_2.csv",
                "dir1/file_3.csv"
            ]
        },
        {
            "URIPrefixes": [
                "dir1/dir2/dir3/",
                "dir1/dir2/dir4/"
            ]
        },
        {
            "WildcardURIs": [
                "dir1/dir2/dir3/*.csv"
            ]
        }
    ]
}
```

The custom manifest file contains the following tags:

- URIs. Specify the path for the files relative to the bucket name.
- URIPrefixes. Specify the path for the directory relative to the bucket name.

- WildcardURIs. Specify an asterisk (*) wildcard in the file name, which are of flat format type, to fetch files from the Amazon S3 bucket. Specify the asterisk (*) wildcard to fetch all the files or only the files that match the name pattern.

You can specify URIs, URIPrefixes, WildcardURIs, or all sections within fileLocations in the `.manifest` file.

You cannot use the wildcard characters to specify folder names. For example, { "WildcardURIs": ["multiread_wildcard/dir1*/", "multiread_wildcard/*/"] }.

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.

Amazon Redshift manifest file

You can use an Amazon Redshift manifest file created by the UNLOAD command to read multiple flat files from Amazon S3. All flat files must have the same metadata and must be available in the same Amazon S3 bucket.

Create a `.manifest` file and list all the source files with the URL that includes the bucket name and full object path for the file. You must specify the `.manifest` file name in the following format: `<file_name>.manifest`.

For example, the Amazon Redshift manifest file contains source files in the following format:

```
{
    "entries": [
            {"url": "s3://mybucket-alpha/2013-10-04-custdata", "mandatory":true},
            {"url": "s3://mybucket-alpha/2013-10-05-custdata", "mandatory":true},
            {"url": "s3://mybucket-beta/2013-10-04-custdata", "mandatory":true},
            {"url": "s3://mybucket-beta/2013-10-05-custdata", "mandatory":true},
    ]
}
```

The Redshift manifest file format contains the following tags:

url

The url tag consists of the source file in the following format:

```
"url": "<endpoint name>://<folder path>/<filename>", "mandatory":<value>
```

mandatory

Amazon S3 V2 Connector uses the `mandatory` tag to determine whether to continue reading the files in the `.manifest` file or not, based on the following scenarios:

- If the value of `mandatory` tag is `true`, and the S3 bucket does not have the specified source file, Amazon S3 V2 Connector does not read the rest of the files as well in the `.manifest` file. The mapping task fails.
- If the value of `mandatory` tag is `false`, and the S3 bucket does not have the specified file, Amazon S3 V2 Connector continues to read the rest of the files in the `.manifest` file in a sequence.
- If the `.manifest` file does not contain any files, the mapping task fails.

By default, the value of `mandatory` tag is `false`.

Incrementally loading files

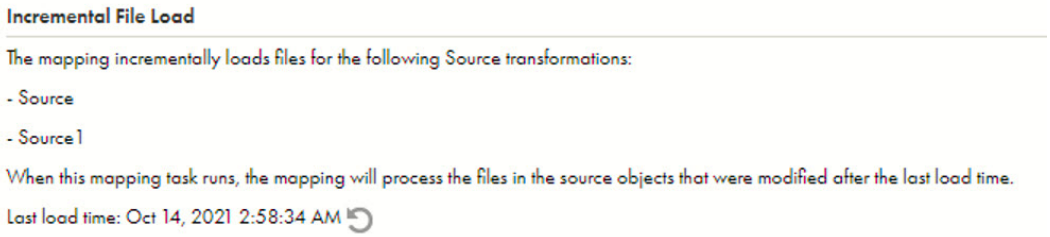
You can incrementally load source files in a directory to read and process only the files that have changed since the last time the mapping task ran.

You can incrementally load files only from mappings in advanced mode. Ensure that all of the source files exist in the same Cloud environment.

To incrementally load source files, select **Incremental File Load** and **Directory** as the source type in the advanced read options of the Amazon S3 V2 data object.

When you incrementally load files from Amazon S3, the job loads files that have changed from the last load time to 15 minutes before the job started running. For example, if you run a job at 2:00 p.m, the job loads files changed before 1:45 p.m. The 15-minute buffer ensures that the job loads only complete files, since uploading objects on Amazon S3 can take a few minutes to complete.

When you configure a mapping task, the **Incremental File Load** section lists the Source transformations that incrementally loads files and the time that the last job completed loading the files. By default, the next job that runs checks for files modified after the last load time.



You can also override the load time that the mapping uses to look for changed files in the specified source directory. You can reset the incremental file load settings to perform a full load of all the changed files in the directory, or you can configure a time that the mapping uses to look for changed files.

A mapping in advanced mode that incrementally load a directory that contains complex file formats such as Parquet and Avro fails if there are no new or changed files in the source since the last run.

For more information on incremental loading, see Reprocessing incrementally-loaded source files in *Tasks* in the Data Integration documentation.

Wildcard characters

When you run a mapping in advanced mode to read data from an Avro, flat, JSON, ORC, or Parquet file, you can use the `?` and `*` wildcard characters to specify the folder path or file name.

To use wildcard characters for the folder path or file name, select the **Allow Wildcard Characters** option in the advanced read properties of the Amazon S3 V2 data object.

? (Question mark)

The question mark character (?) allows one occurrence of a character. For example, if you enter the source file name as `a?b.txt`, the Secure Agent reads data from files with the following names:

- `a1b.txt`
- `a2b.txt`
- `aab.txt`

*** (Asterisk)**

The asterisk mark character (*) allows zero or more than one occurrence of a character. For example, if you enter the source file name as `a*b.txt`, the Secure Agent reads data from files with the following names:

- `aab.txt`
- `a1b.txt`
- `ab.txt`

- `abc11b.txt`

You can use 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.*

If you specify `abc*.txt`, the Secure Agent reads all the file names starting with the term `abc` and ending with the `.txt` file extension. If you specify `abc.*`, the Secure Agent reads all the file names starting with the term `abc` regardless of the extension.

Rules and guidelines for wildcard characters

Consider the following rules and guidelines when you use wildcard characters to specify the folder path or file name:

- You cannot specify wildcard characters in a bucket name.
- When you specify wildcard characters in a folder path and the Amazon S3 bucket does not contain folders matching the name pattern, the mapping fails.
- When you specify wildcard characters in a file name and the Amazon S3 bucket does not contain files matching the name pattern, the mapping fails.
- When you use wildcard characters in a folder path for a mapping in advanced mode, the Secure Agent reads data from the folders and the files that match the name pattern.

Recursively read files from directories

You can read objects stored in subdirectories in Amazon S3 V2 mappings in advanced mode. You can use recursive read for flat, Avro, JSON, ORC, and Parquet files. The files that you read using recursive read must have the same metadata.

To enable recursive read, select the source type as **Directory** in the advanced source properties. Enable the **Recursive Directory Read** advanced source property to read objects stored in subdirectories.

You can also use recursive read when you **specify wildcard characters** in a folder path or file name. For example, you can use a wildcard character to recursively read files in the following ways:

- Folder path is **/abc*/**. Returns all files within any folder or subfolder that has a pattern `abc` at the starting of the folder name.
- Folder path is **/abc*/** and file name is **myfile***. Returns all files that have a pattern `myfile` at the starting of the file name within a folder or subfolder and has a pattern `abc` at the starting of the folder name.

Source partitioning

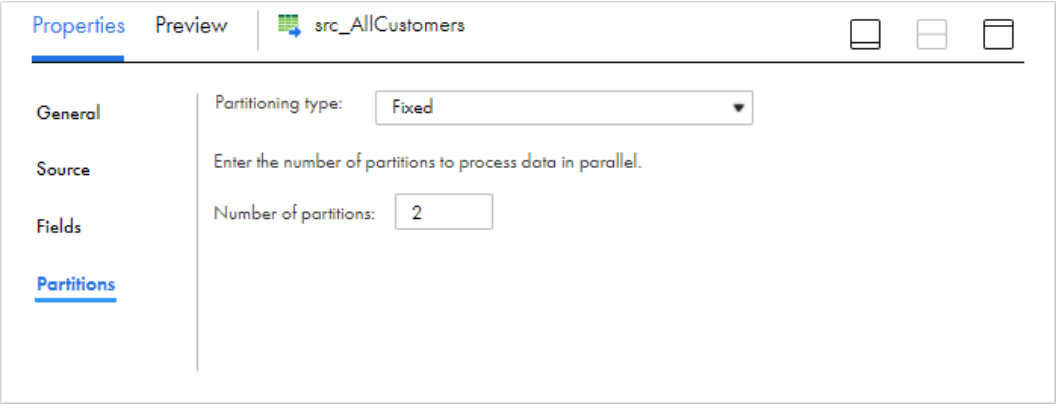
You can configure fixed partitioning to optimize the mapping performance at run time when you read data from flat, Avro, ORC, or Parquet files. You can configure fixed partitioning only on mappings.

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

Enable partitioning when you configure the Source transformation in the Mapping Designer.

On the **Partitions** tab for the Source transformation, you select fixed partitioning and enter the number of partitions based on the amount of data that you want to read. By default, the value of the **Number of partitions** field is one.

The following image shows the configured partitioning:



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. Applicable only to mappings.
- ORC
- Parquet

Note: Avoid using FileName as the column name in the source data. FileName is a reserved keyword. The name is case sensitive.

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

Feature	Mapping	Mappings in advanced mode
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`.

SQL ELT optimization

You can enable full SQL ELT optimization when you want to load data from Amazon S3 sources to your data warehouse in Amazon Redshift. While loading the data to Amazon Redshift, you can transform the data as per your data warehouse model and requirements. When you enable full SQL ELT optimization on a mapping task, the mapping logic is pushed to the AWS environment to leverage AWS commands. You cannot configure SQL ELT optimization for a task based on a mapping configured in advanced mode.

For more information on SQL ELT optimization, see the help for Amazon Redshift V2 Connector. If your use case involves loading data to any other supported cloud data warehouse, see the connector help for the applicable cloud data warehouse.

Amazon S3 V2 targets

You can use an Amazon S3 V2 object as a target in a mapping or amapping 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.

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 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 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.

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 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.

Note: For Informatica Encryption in the advanced cluster, you must install the Secure Agent on the Amazon EC2 machine.

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

Encryption Type	Avro File	Binary File[1]	Flat	JSON File[2]	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]Doesn't apply to mappings in advanced mode.
[2]Applies only to mappings in advanced mode.

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:

- When you use Informatica encryption in a mapping, the `_SUCCESS` file is not generated in the target directory for mappings in advanced mode.
- When you use Informatica encryption in a mapping, you cannot encrypt more than 1000 files.

To understand how to enable Informatica encryption in the AWS console and Data Integration, see Configuring Informatica Encryption for Mappings in Advanced Mode in Amazon S3 V2 Connector.

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 as the file that you overwrite, and creates a new file with the same file name in the target directory.

The Secure Agent also deletes all files and folders that have the same prefix as the file names created at design time. This is applicable for Avro, ORC, and Parquet files.

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. You can configure target partitioning only on mappings.

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.

Incremental write to partition directory

You can write to a partition directory incrementally for a mapping in advanced mode and append data to the partition directory.

When you do not select the **Overwrite File(s) If Exists** option in the Amazon S3 V2 target advanced properties, you can write to a partition directory incrementally and append data to the partition directory.

When you create a target at runtime and the file name ends with `/`, the Secure Agent appends the incoming source partitions to the existing parent directory and does not add a time stamp to the target file name. For an existing target, the Secure Agent appends the incoming source partitions to the existing parent directory and does not add a time stamp to the target file name.

You can override the folder path for a target that you create at runtime. If you specify a file name ending with `/` and override the folder path, the Secure Agent considers the path as a folder. The file name is not appended with the folder path.

You cannot override the file name for both an existing target or for a target that you create at runtime.

Distribution column

You can write multiple flat files to Amazon S3 target from a single source in a mapping. 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 the following different names for null values:

  - For mappings: `_EMPTY_`

  - For mappings in advanced mode: `_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. You can use an object tag for flat files.

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",
"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.

Directory-level partitioning

You can read from and write to partition columns when you use mappings in advanced mode.

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.

You can read data from or write data to partition columns for the following file formats:

- Avro
- Parquet
- ORC
- JSON

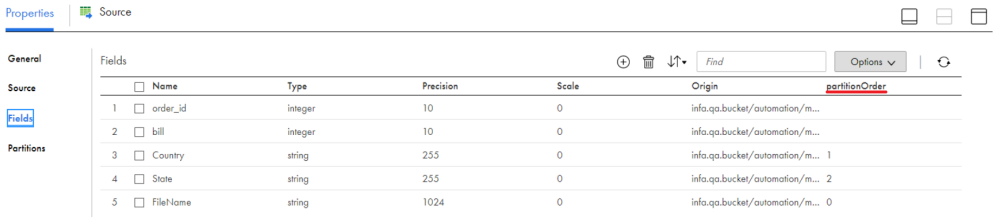
Reading from partition columns

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.



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:

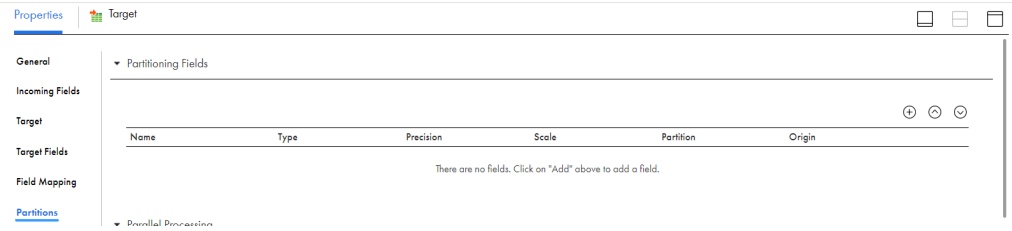


The **partitionOrder** column specifies whether a column is partitioned.
In the above image, 2 partition columns are present. the partition order values 1 and 2 signify the order in which the `Country` and `State` fields were selected for partitioning. The FileName field has 0 as the partition order.

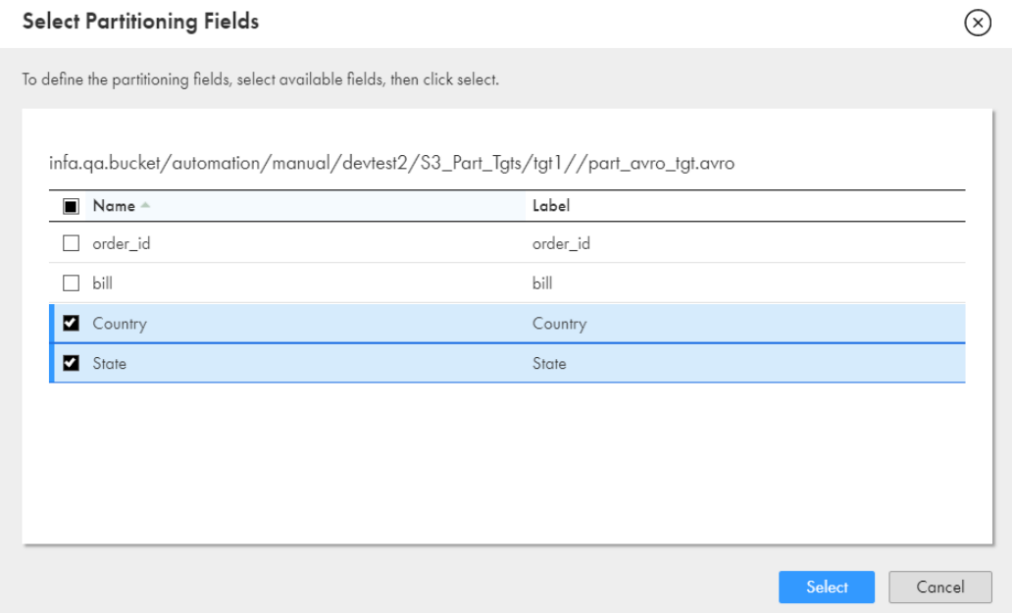
Writing to partition columns

Perform the following steps to write to partition columns:

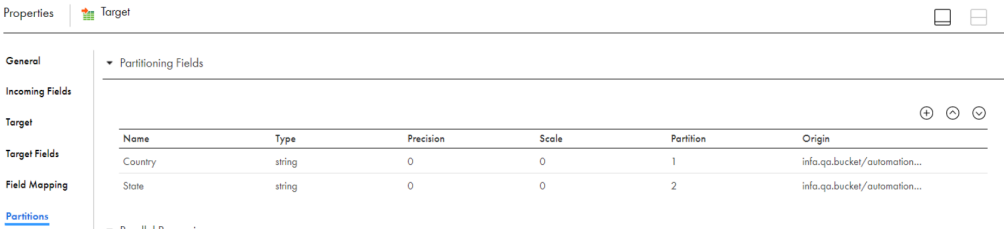
1. Click the ⊕ icon in the **Partitions** tab to add the partition columns for a target. The following image shows how you can add the partition columns:



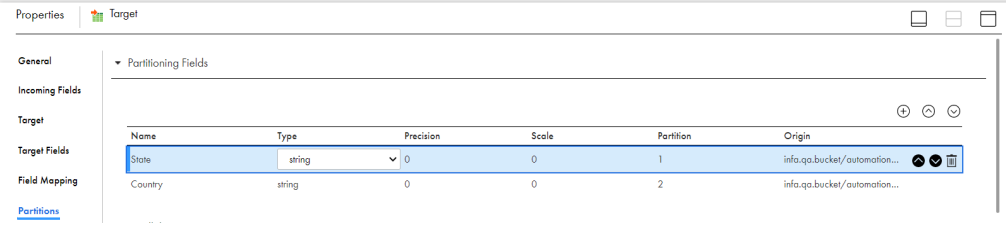
2. In the Partitions tab, select the partitioning fields from the list of available fields.



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



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



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 push down 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 pass a value with special characters in a partition column, the value gets encoded. For example, `@#$#$%%?*` gets encoded as `@%23$%23$%25%25%3F%2A`.
- 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 columns of hierarchical data type as partition columns.
- 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. You cannot add partition fields and arrange the partition columns in an order for an existing target.
- 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.

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]	Flat	JSON File [2]	ORC File	Parquet File
None	Yes	No	Yes	Yes	Yes	Yes
Bzip2	No	No	No	Yes	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

Note: Binary File support doesn't apply to mappings in advanced mode, while the JSON File support applies only to mappings in advanced mode.

The following table lists the supported target compression formats:

Compression format	Avro File	Binary File[1]	Flat	JSON File[2]	ORC File	Parquet File
None	Yes	No	Yes	Yes	Yes	Yes
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]Doesn't apply to mappings in advanced mode. [2]Applies only to mappings in advanced mode.						

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 data from 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 data from a flat file.

1. Select the required compressed flat file.
2. Navigate to **Formatting Options** property field.
3. Select the **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.

Reading a compressed JSON file

When you run a mapping to read data from a compressed JSON file, you must upload a schema file and select `Bzip2` as the compression format. Use the **.BZ2** file name extension when you use the `Bzip2` compression format to read a JSON file.

1. Select the required compressed JSON 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 JSON file:

```
{"Field1":"<string>","Field2":"<string>","Field3":<integer>}
```

   Use a row that has data for all the columns as the JSON schema.

4. Select **Compression Format** as **Bzip2** from the advanced source properties.

Fixed-width file formats

You can use a fixed-width flat file as a source or target in mappings and mapping tasks.

When you configure a Source transformation or Target transformation and select the fixed-width flat file type, you must select the most appropriate fixed-width file format to use based on the data in the fixed-width flat file.

Consider the following rules and guidelines for a fixed-width flat file:

- You cannot use a fixed-width flat file as a source or target for mappings in advanced mode and data transfer tasks.
- When you create a flat file target at runtime and select a fixed width file format, the Secure Agent ignores the fixed-width column boundaries that you specified for the fixed-width flat file format and applies the additional fixed width attributes for the new target object.
- When you write a column of Numeric data type from a fixed-width flat file source to an empty fixed-width flat file target that uses the same fixed-width file format, the Secure Agent appends a null character to the value in the Numeric column in the target.
- When you create a fixed-width file format, the sample file uses different characters as the new line symbol, depending on the operating system on which the Secure Agent is installed. Also, ensure that the source files use the same symbol.
  - For a Linux machine, use the `\n` character as the new line symbol for the sample file.
  - For a Windows machine, use the `\r\n` character as the new line symbol for the sample file.
- When you use a fixed-width flat file as a source or target, you cannot edit the metadata for the fields.
- When you read data of the datetime data type, you can read the date value only upto milliseconds.
- When you read data of the double data type from a fixed-width file and write the data to a Parquet or Avro file, the double data type is mapped to the decimal data type in target. Hence, the data is written incorrectly.
  To write the data correctly to the target, edit the metadata in the Target transformation and change the decimal data type to double.
- When you create a fixed-width file format, ensure that the sample flat file only uses UTF-8 character set encoding.

CHAPTER 4

Mappings and mapping tasks with Amazon S3 V2

When you create a mapping, you can configure Source, Target, and Lookup transformations to represent an Amazon S3 V2 object.

Use the Mapping Designer in Data Integration to add the Source, Target, or Lookup transformations in the mapping canvas and configure the Amazon S3 source, target, and lookup properties.

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

Amazon S3 V2 objects in mappings

You can define source and target properties to run mappings and mapping tasks using an Amazon S3 V2 connection. You can also parameterize the connection and objects. You can use a parameter file in the task properties to overwrite the connection and object properties at runtime. Additionally, you can also specify file formatting options for Amazon S3 V2 objects.

You can use Amazon S3 V2 objects in mappings configured in advanced mode.

Amazon S3 V2 sources 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.

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. If you want to overwrite the parameter at runtime, select the **Allow parameter to be overridden at run time** option when you create a parameter. When the task runs, the agent uses the parameters from the file that you specify in the task advanced session properties.
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. When you select an object, you can also select a `.manifest` file object when you want to read from multiple files.
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 **Allow parameter to be overridden at run time** option when you create a parameter. When the task runs, the agent uses the parameters from the file that you specify in the task advanced session properties.
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[1] - Flat - Avro - ORC - Parquet - JSON[2] - Discover Structure[2] Default is **None**. If you select **None** as the format type, the Secure Agent reads data from Amazon S3 files in binary format. 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, see "File formatting options" on page 57.

Property	Description
Intelligent Structure Model[2]	Applies to Discover Structure format type. Determines the underlying patterns in a sample file and auto-generates a model for files with the same data and structure. Select one of the following options to associate a model with the transformation: - Select. Select an existing model. - New. Create a new model. Select **Design New** to create the model. Select **Auto-generate from sample file for Intelligent Structure Discovery** to generate a model based on sample input that you select. Select one of the following options to validate the XML source object against an XML-based hierarchical schema: - Source object doesn't require validation. - Source object requires validation against a hierarchical schema. Select to validate the XML source object against an existing or a new hierarchical schema. When you create a mapping task, on the **Runtime Options** tab, you configure how Data Integration handles the schema mismatch. You can choose to skip the mismatched files and continue to run the task or stop the task when the task encounters the first file that does not match. For more information, see *Components*.
[1]Doesn't apply to mappings in advanced mode. [2]Applies only to mappings in advanced mode.	

The following table describes the Amazon S3 V2 source advanced properties:

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, see "Source types in Amazon S3 V2 sources" on page 32.
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.
Incremental File Load[2]	Indicates whether you want to incrementally load files when you use a directory as the source for a mapping in advanced mode. When you incrementally load files, the mapping task reads and processes only files in the directory that have changed since the mapping task last ran. For more information, see "Incrementally loading files" on page 34.

Property	Description
Allow Wildcard Characters[2]	Indicates whether you want to use wildcard characters for the directory source type. If you select this option, you can use the question mark (?) and asterisk (*) wildcard characters in the folder path or file name. For more information, see "Wildcard characters" on page 35.
Enable Recursive Read[2]	Indicates whether you want to read flat, Avro, JSON, ORC, or Parquet files recursively from the specified folder and its subfolders and files. Applicable when you select the directory source type. For more information, see "Recursively read files from directories" on page 36.
Encryption Type	Method you want to use to decrypt data. You can select one of the following encryption types: - None - Informatica encryption Default is **None**. **Note:** You cannot select client-side encryption, server-side encryption, and server-side encryption with KMS encryption types.
Staging Directory[1]	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 source property does not apply to Avro, ORC, and Parquet files.
Hadoop Performance Tuning Options	This property is not applicable for Amazon S3 V2 Connector.
Compression Format	Decompresses data when you read data from Amazon S3. You can choose to decompress the data in the following formats: - None - Bzip2[2] - Gzip - Lzo Default is **None**. You can decompress data for a mapping in advanced mode 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, see "Data compression in Amazon S3 V2 sources and targets" on page 47.
Download Part Size[1]	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.

Property	Description
Multiple Download Threshold[1]	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.

[1]Doesn't apply to mappings in advanced mode.
[2]Applies only to mappings in advanced mode.

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. If you want to overwrite the parameter at runtime, select the **Allow parameter to be overridden at run time** option when you create a parameter. When the task runs, the agent uses the parameters from the file that you specify in the task advanced session properties.
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. When you create an object at runtime, enter a name and the path for the target object.
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 **Allow parameter to be overridden at run time** option when you create a parameter. When the task runs, the agent uses the parameters from the file that you specify in the task advanced session properties.
Create Target	Creates a target. Enter a name and path for the target object. You can use parameters defined in a parameter file in the target name. For more information, see "Rules and guidelines for creating a target" on page 61.

Property	Description
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 [1] - Flat - Avro - ORC - Parquet - JSON[2] Default is **None**. If you select **None** is as the format type, the Secure Agent writes data to Amazon S3 files in binary format. Open the **Formatting Options** dialog box to define the format of the file. For more information about format options, see "File formatting options" on page 57.
Operation	Type of the target operation. You can perform only insert operation on an Amazon S3 V2 target.
[1]Doesn't apply to mappings in advanced mode. [2]Applies only to mappings in advanced mode.	

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

Property	Description
Overwrite File(s) If Exists	Overwrites an existing target file. Default is true. For more information about overwriting the existing files, see "Overwriting existing files" on page 40.
Folder Path	Bucket name or folder path where you want to write the Amazon S3 target file. The path that you enter here overrides the path specified for the target configured to create at runtime. 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.

Property	Description
Encryption Type	Method you want to use to encrypt data. Select one of the following encryption types: - None - Client Side Encryption[1] - Server Side Encryption - Server Side Encryption with KMS - Informatica Encryption Default is **None**. For more information about the encryption type, see "Data encryption in Amazon S3 V2 targets" on page 38.
Staging Directory[1]	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 target property does not apply to Avro, ORC, and Parquet files.
File Merge	This property is not applicable for Amazon S3 V2 Connector.
Hadoop Performance Tuning Options	This property is not applicable for Amazon S3 V2 Connector.
Compression Format	Compresses data when you write data to Amazon S3. You can compress the data in the following formats: - None - Bzip2[2] - Deflate - Gzip - Lzo - Snappy - Zlib Default is None. **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 47.
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 42.

Property	Description
TransferManager Thread Pool Size[1]	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[1]	Determines 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.
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[1]	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.
[1]Doesn't apply to mappings in advanced mode. [2]Applies only to mappings in advanced mode.	

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*.

File formatting options

When you select the format of an Amazon S3 file, you can configure the formatting options.

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

Property	Description
Schema Source	The schema of the source or target file. You can select one of the following options to specify a schema: - Read from data file. Imports the schema from the file in Amazon S3. - Import from Schema File. Imports schema from a schema definition file in your local machine.
Schema File	Upload a schema definition file. You cannot upload a schema file when you create a target at runtime.

The following table describes the formatting options for flat files:

Property	Description
Read from data file	Imports the schema from the file in Amazon S3. If you select **Read from data file** and use the JSON[2] file format, 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.
Import from schema file	Imports schema from a schema definition file in your local machine. If you select **Import from schema file**, you can select **Schema File** to upload a schema file. You cannot upload a schema file when you select the **Create Target** option to write data to Amazon S3.
Flat File Type	The type of flat file. Select one of the following options: - Delimited. Reads a flat file that contains column delimiters. - Fixed Width. Reads a flat file with fields that have a fixed length. You must select the file format in the **Fixed Width File Format** option. If you do not have a fixed-width file format, click **New** > **Components** > **Fixed Width File Format** to create one.
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.
Escape Char	Character immediately preceding a column delimiter character embedded in an unquoted string, or immediately preceding the quote character in a quoted string. You can specify a character or `\<decimal value>`. When you specify `\<decimal value>`, the agent considers the ASCII character for the decimal value as the escape character. For example, if you specify `\64`, the agent considers the ASCII character `@`. To ignore the escape character, specify `\0`.
Qualifier	Quote character that defines the boundaries of data. You can set the 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. Applies qualifier to data that has a delimiter value in the data. Otherwise, the Secure Agent does not apply the qualifier when writing data to the target. - All. Applies the qualifier to all non-empty columns. Default mode is minimal.
Code Page	Select the code page that the 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.

Property	Description
Disable escape char when a qualifier is set	Check to disable the escape character when a qualifier value is already set.
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. 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 when you preview the source data and at runtime for the mapping. Default is 1.
First Data Row[1]	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[1]	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 41.
Max Rows To Preview	Not applicable to Amazon S3 V2 Connector.
Row Delimiter	Character used to separate rows of data. You can set values as \r, \n, and \r\n.

[1] Doesn't apply to mapping in advanced mode.

[2]Applies only to mappings in advanced mode.

The following table describes the formatting options for JSON files:

Property	Description
Data elements to sample[1]	Specify the number of rows to read to find the best match to populate the metadata.
Memory available to process data[1]	The memory that the parser uses to read the JSON sample schema and process it. The default value is 2 MB. If the file size is more than 2 MB, you might encounter an error. Set the value to the file size that you want to read.
Read multiple-line JSON files	Not applicable.

[1]Applies only to mappings in advanced mode.

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 for mappings in advanced mode.
- 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 flat 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 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"}]}
```

- 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
```

- You cannot select the **Read multiple-line JSON files** checkbox in the formatting options, as Amazon S3 V2 does not support the feature.
- When you run a mapping with an Amazon S3 V2 source, with source columns having values of Parquet-datetime data types, then the timestamp values that are mapped to the target do not appear in UTC format if you do not enable full SQL ELT optimization.

Specifying a target

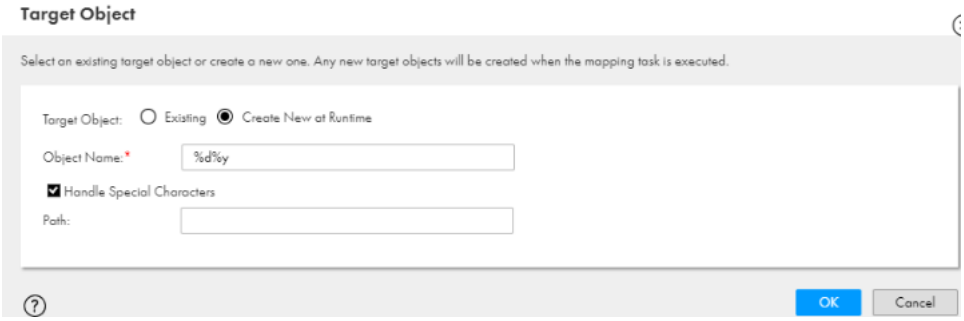
You can use an existing target or create a target to hold the results of a mapping. If you choose to create the target, the Secure Agent creates the target when you run the task.

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:



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

Rules and guidelines for creating a target

Consider the following rules and guidelines when you use the **Create Target** property:

- 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.
- When you write an Avro, ORC, or Parquet file using **Create Target** and the target directory has a file name that contains a colon (:), the mapping fails.
- 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.
- When you configure the precision of a string column of a JSON file in the source and select **Create Target**, the default precision of the string column is retained.
- When you select the file or directory source type, select **Create New at Runtime**, and run the mapping, the tomcat log shows the following exception even if the mapping succeeds:
  `Internal error. Encountered an error because invalid path element [Partition] was encountered. Contact Informatica Global Customer Support.`
- When you run a mapping on the advanced mode with an Amazon S3 source, and create a new target object by selecting **Create New at Runtime**, the mapping is successful, but the special characters in the source are replaced by an underscore (_).
- When you write data to a flat file created at runtime, the target flat file contains a blank line at the end of the file.

Rules and guidelines for the path in target object

Consider the following rules and guidelines when you specify a path in the **Create Target** property:

- 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`.
- Do not specify a bucket name in the path.
  - For mappings, if you specify a bucket name in the path, the bucket name is considered as a folder in the folder path. For example, if you specify the path as `<bucket_name>/<path_name>`, `<bucket_name>` is considered as a folder.
  - For mappings in advanced mode, if you specify a bucket name in the path, the bucket name is ignored.

Amazon S3 V2 parameterization

You can parameterize both Source and Target objects using input parameter and data in advanced properties using in-out parameter.

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.

If you configure a mapping with following criteria, the mapping fails:

- Parameterized source and target
- Allow parameter to be overridden at run time checkbox is selected
- Source object is selected within a folder during the mapping task creation

Parameterization using timestamp

You can append time stamp information to the file name to show when the file is created. You can use parameterization using timestamp when you create a mapping to write a file of flat format type.

You cannot parameterize using timestamp in mappings in advanced mode.

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.

Special Character	Description
%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. You can use a parameter file 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 **Runtime Options** tab of the mapping task.
6. Save and run the mapping task.

Rules and guidelines for mappings in advanced mode

Consider the following guidelines when you create a mapping in advanced mode:

- When you run a mapping in advanced mode, 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 a mapping in advanced mode 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 mapping in advanced mode 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 configure a read operation to read from a file, the file name should not begin with _ or ..
- When you configure 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 read 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.
- When you read data from a JSON file that contains unicode characters, data inconsistencies might occur.
- 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.
- The staging directory source and target advanced property is not applicable. However, you must specify a staging directory on Amazon S3 in advanced configurations. For more information, see *Administrator*.
- When you set the qualifier mode to Minimal and use an escape character, the characters are not escaped and quoted in the target. Set the qualifier mode to All.
- When you set the qualifier mode to All and do not specify a value for the qualifier, \0 (NUL) is considered as the qualifier.
- When you parameterize a flat source file, the FileName field appears in the source fields. As a workaround, add a rule in the incoming fields of the target to exclude the FileName field.
- If the table or schema names contain a backward slash (/), the mapping fails.
- If the JSON data that you read from a source fails to align with the source schema defined in the schema definition file, the data written to the target appears corrupted.

Mapping in advanced mode example

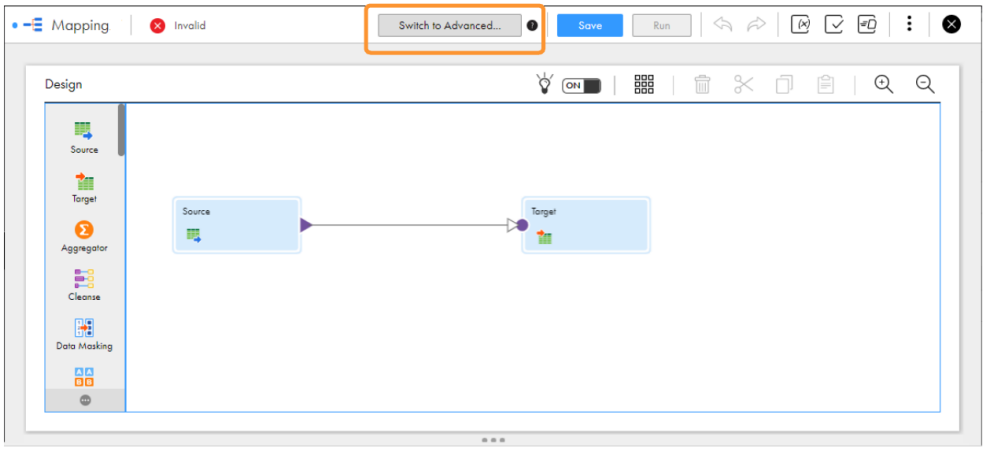
You work for one of the largest community college that maintains millions of records in their ongoing student database. The college has more than 10,000 faculty members teaching at 45 campuses and 700 locations

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 a mapping that runs in advanced mode to achieve faster performance when you read all the purchase records from Amazon S3 and write the records to an Amazon Redshift target.

1. In Data Integration, click **New** > **Mappings** > **Mapping**.
2. In the Mapping Designer, click **Switch to Advanced**.

   The following image shows the **Switch to Advanced** button in the Mapping Designer.



3. In the **Switch to Advanced** dialog box, click **Switch to Advanced**.

   The Mapping Designer updates the mapping canvas to display the transformations and functions that are available in advanced mode.
4. Enter a name, location, and description for the mapping.
5. On the Source transformation, specify a name and description in the general properties.
6. 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.
7. On the **Fields** tab, map the source fields to the target fields.
8. On the Target transformation, specify a name and description in the general properties.
9. 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.

10. Map the source and target.
11. Click **Save** > **Run** to validate the mapping.

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

CHAPTER 5

Migrating a mapping

You can configure a connection and mapping in one environment and then migrate and run the mapping in another environment. You can also migrate mappings configured in advanced mode.

After the migration, you can change the connection properties from the Administrator service, but you do not need to modify the mapping. Data Integration uses the configured runtime attributes from the earlier environment to run the mapping successfully in the new environment.

Consider a scenario where you develop a mapping in the development organization (Org 1) and you then migrate and run the mapping in the production organization (Org 2). After you migrate, you might want to use the same or a different connection endpoint or object path in Org 2. Based on your requirement, follow the guidelines in this section before you plan the migration.

Use the same object path for the migrated mapping

If you want the migrated mapping in Org 2 to use the same object path as in Org 1, you must maintain the same folder path and file name in the Amazon S3 account for Org 2.

For example, if you have two different accounts, Account1 used for Org 1 and Account2 used for Org 2, the folder path and the file name must be the same in both the accounts:

Account1: folder1/filename1

Account2: folder1/filename1

In this scenario, you do not need to override the folder path and file name in the advanced properties.

Rules and guidelines for the same object path

Consider the following rules and guidelines when you use the same object path for the migrated mapping:

- When you want to specify the bucket name in the folder path in a connection, you can specify only the bucket name in the folder path. You cannot specify both the bucket name and the folder name, for example, `bucket1/folder1`. Otherwise, the mapping fails with an error.
- When you migrate a mapping that has a bucket name in the folder path, only the bucket name changes. For example, if a mapping has a source object `bucket1/folder1/source1`, and is imported to Org2 where the connection has folder path `bucket2`, the data is read from `bucket2/folder1/source1`.
- For a mapping in advanced mode, when you specify the folder path in the connection in Org 1 as `bucket1/folder1`, and migrate the mapping to Org 2 with the folder path as `bucket2`, the mapping writes data to the folder `bucket2/bucket1/folder1`.
  A mapping writes data to the folder `bucket2/folder1`.

Use a different object path for the migrated mapping

After you migrate the mapping, you can use a different object path to run the mapping from the new environment.

In this scenario, before you migrate the mapping, you can change the object metadata, runtime attributes, or the connection attributes to reflect the object path in the migrated environment. You do not have to edit or update the mapping in the new environment.

As a rule, when you specify the folder path and file name in the advanced properties, connection, or object properties, Data Integration honors the attributes in the following order of precedence:

1. **Runtime advanced attributes**. The advanced properties such as the folder path and file name in the Source, Target, or Lookup transformation in a mapping.
2. **Connection attributes**. The folder path attribute in the connection properties.
3. **Object metadata**. The object selected in the Source, Target, or Lookup transformation in a mapping.

Migration options

When you migrate, you can choose from one of the following options to update the object path:

Option 1. Update the connection properties to reference the new object

When you import the mapping into Org 2, in the **Review Connections** section, you can change the existing connection to map to the connection that has access to the specified folder path and file name in Org 2.

Option 2. Override the properties from the advanced properties

Before the migration, specify the required folder path and file name for the object from Org 2 in the advanced properties of the Org 1 mapping.

After the migration, when you run the mapping, the Secure Agent uses the configured advanced parameters to override the object specified in the mapping imported from Org 1.

Option 3. Parameterize the properties in the mapping

You can choose to parameterize the advanced attributes, such as the folder path and file name before the migration. You can configure input parameters, in-out parameters, and parameter files in the mapping. After you migrate the mapping, do not edit or update the mapping. If you have used in-out parameters for the advanced attributes such as for the folder path and file name, you can update these from the parameter file.

Parameterizing only the advanced properties, but not the object in the mapping

If you want to parameterize only the advanced properties and use them at runtime, select a placeholder object in the object properties in the mapping and then specify an override to this placeholder object from the advanced properties. Ensure that the placeholder object contains the same metadata as the corresponding table that you specify as an override. When you run the mapping, the value specified in the advanced property overrides the placeholder object.

Parameterizing both the object and the advanced properties

If you want to keep both the Amazon S3 object type and the advanced fields parameterized, you must leave the **Allow parameter to be overridden at runtime** option unselected in the input parameter window while adding the parameters, and then select the required object at the task level. When you run the task, the values specified in the advanced properties take precedence.

Parameterization rules

Consider the following rules to parameterize the object and advanced properties:

- Parameterization is not applicable for mappings that use the **Create Target** option.
- Parameterization is not applicable for mappings in the advanced mode in the migration use case.
- You cannot enable the **Allow parameter to be overridden at run time** check box to parameterize the connection or object.

Rules and guidelines

Consider the following rules and guidelines when you use the same or a different object path for the migrated mapping :

- The following table lists the transformation, object type, and the fields in the advanced properties of a mapping that you can retain when you migrate to the new environment:

Transformations	Object Type	Advanced Fields
Source	Single object	Folder path and file name
Lookup	Single object	
Target	Single object	

- After you migrate the mapping from Org1 to Org2, you must not edit the mapping.
- For an existing target object in Org2, you must specify the folder path and file name in the advanced properties. If the object does not exist in Org2 or the connection does not have access to the object that you used in Org1, the mapping fails with a 403 error.
- Ensure that you specify a valid folder and file name in the advanced properties. Otherwise, the mapping fails.
- When you use the FileName field in a target object, override the folder path and file name, and migrate the mapping to Org2, the mapping fails with a 403 error. This issue occurs if the connection in Org2 does not have access to the bucket, object, and folder path that you used in Org1.
- When you specify wildcard characters in the folder path advanced property, all folders that match the name pattern must contain files with the same metadata and formatting options.
- When you recursively read files from subdirectories, the folder path advanced property that you override must contain a file in the directory.
  For example, if you specify the folder path as `<bucket1>/<dir1>`, ensure that a file is present in the directory, `<dir1>`: `<bucket1>/<dir1>/<file1>`.

  If the folder path advanced property contains files only in the subdirectories and no file in the directory, the mapping fails.
- You cannot dynamically refresh the data object schema at runtime. You must maintain the same metadata for the table selected in the source, target, or lookup transformations and the corresponding advanced field overrides as schema change handling is not supported.
- You can configure all encryption types, except Informatica encryption, in a mapping that you migrate to the new environment.
- When you select the source type as Directory, delete any auto-generated transactional files, for example, `.tmp`, for the mapping to run successfully. The `_SUCCESS` and `.CRC` files are ignored.

- If you do not provide the bucket name and specify the folder path in the advanced property 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 the advanced 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 the `<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 the advanced property, the Secure Agent writes the file in the `<my_bucket2>/<dir2>` folder path.
- Ensure that you specify a valid override folder path in the advanced properties for a mapping in advanced mode. If the override folder path pattern does not match with the actual folder path, the mapping fails.
- When the source object is enabled for recursive read in the mapping, after the migration, make sure you have a file in the parent directory in the overridden folder path specified for the source object, or manually reselect the object. Otherwise, the mapping fails.

CHAPTER 6

Upgrading to Amazon S3 V2 Connector

If you are accessing Amazon S3 using the Amazon S3 V1 connection, you can upgrade to the newer Amazon S3 V2 Connector. You can replace the source or target connection type in existing mappings and mapping tasks that use the Amazon S3 connection with the Amazon S3 V2 connection.

Upgrade the connection type

After you replace the connection in an existing mapping, the object selected previously is not retained. You must reimport the Amazon S3 object. The configured advanced source, target, and lookup properties in the fields that are common between the two connectors are retained in the new connector. You also have the option to retain the configured field mappings from the old connector.

You can run the mapping successfully using the configured values from the old connector. You can additionally configure features that the enhanced Amazon S3 V2 Connector offers.

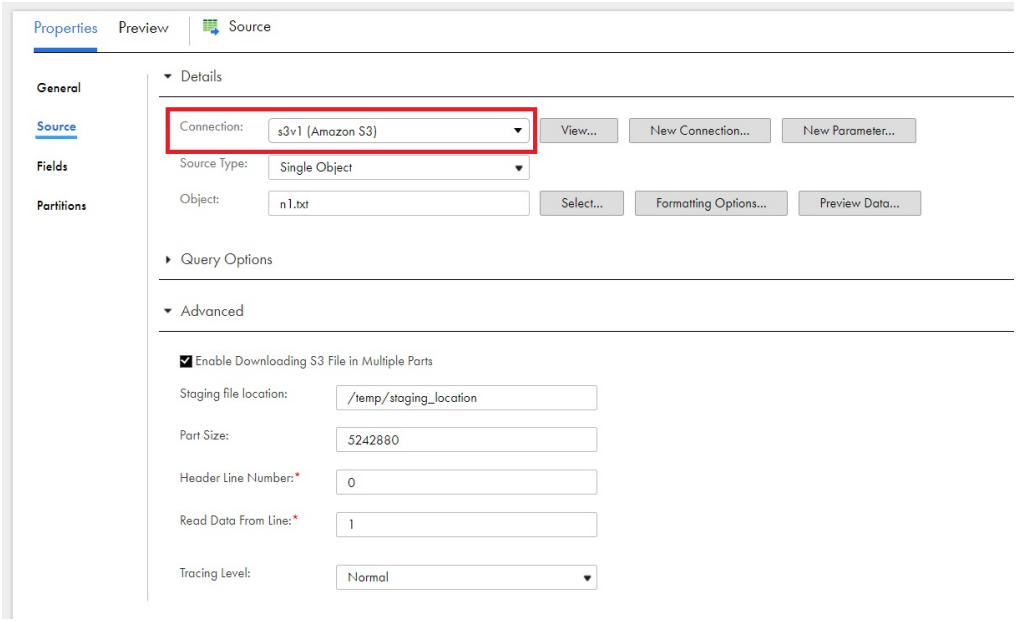
Note: If you are using the Amazon S3 V1 connection in mappings to read from or write data to Amazon S3, Informatica recommends you to use the Amazon S3 V2 connection to make use of the features that the enhanced connector offers. To get the license for Amazon S3 V2 Connector, contact Global Customer support.

Connection switching example

You want to upgrade your existing Amazon S3 mapping that uses the Amazon S3 V1 connection to the Amazon S3 V2 connection.

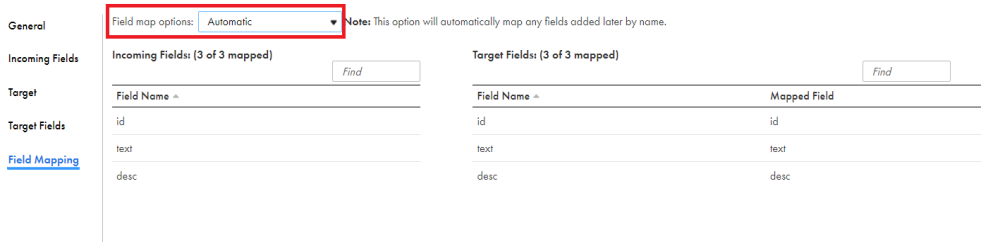
1. Open the existing Amazon S3 V1 mapping that you want to upgrade to Amazon S3 V2.

   The following image shows an existing mapping that uses the Amazon S3 V1 connection and contains the configured advanced properties in the Source transformation:



2. To retain the mapped fields from the field mapping when you switch the connection, on the **Field Mapping** tab, choose from the following **Field Map Options** menu in the Amazon S3 V1 mapping:

   - To retain the fields automatically mapped after the switch, select **Automatic**.

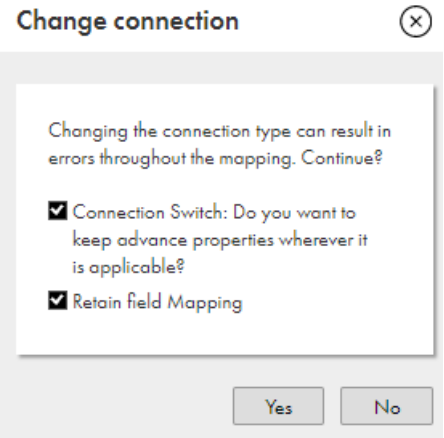


   - To manually map the retained fields after the switch, select **Manual**.

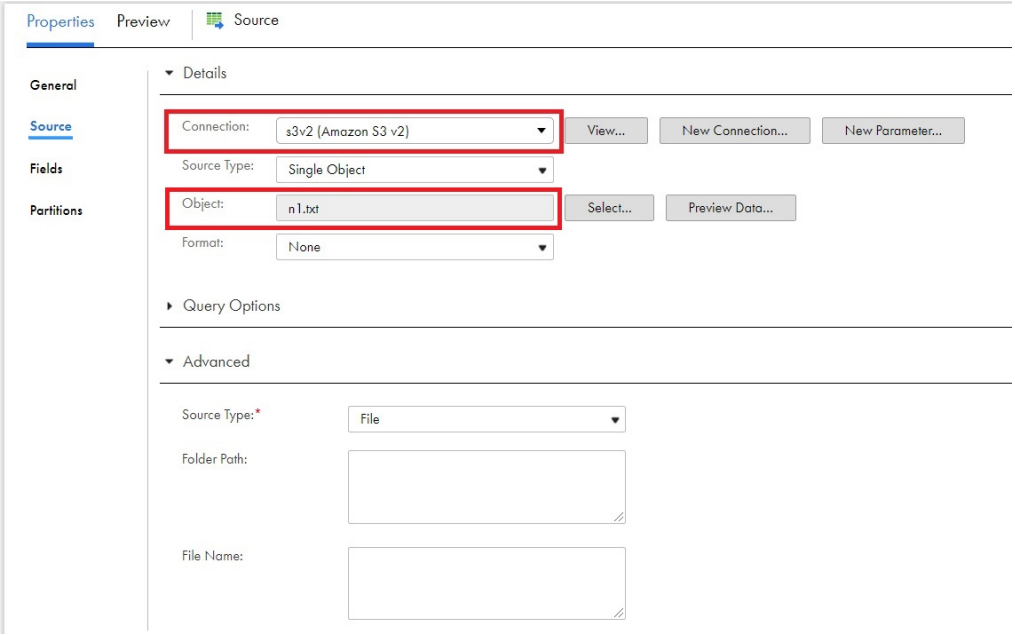
     Note: When you select manual, after switching the connection, you have the option to automap the retained fields using the previous mapping.

3. To switch the connection, in the **Connection** field, change the connection from Amazon S3 V1 to Amazon S3 V2.
4. In the **Change Connection** dialog box, select the following properties, and click **Yes**:
   - **Connection switch**. Switches to the connection that you select.
   - **Retain field mapping**. Retains the configured field mappings from Amazon S3 V1.

The following image shows the options that you must select:

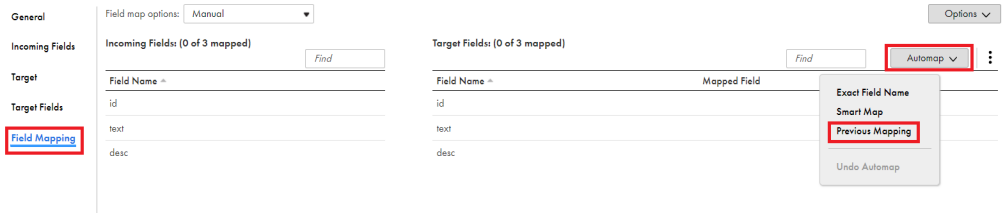


5. Use the same object path in the mapping as Amazon S3 V1. The following image shows the switched connection with the same object path as Amazon S3 V1:



The configured source advanced properties from the Amazon S3 V1 mapping reflect in the Source transformation.

6. If you had selected **Manual** on the **Field Mapping** tab in the Amazon S3 V1 mapping and you want to reflect the field mappings in Amazon S3 V2, on the **Field Mapping** tab, select **Automap**, and then select **Previous Mapping**.



Note: If you had selected **Automatic** in Amazon S3 V1, you do not have to perform this task.

7. Click **Save**.

Advanced properties retained after the switch

When you replace the source or target connection type in existing mappings or mapping tasks that use the Amazon S3 connection with the Amazon S3 V2 connection, you have the option to retain the configured advanced properties.

The following table lists the configured advanced source, lookup, and target properties from Amazon S3 V1 that are retained in Amazon S3 V2 mappings:

Properties	Amazon S3 V1 properties retained in Amazon S3 V2
Source and Lookup	- Part Size - Staging Directory
Target	- Part Size - Staging Directory - TransferManager Thread Pool Size - Object Tags - Merge Partition Files - Folder Path - Compression Format

CHAPTER 7

Data type reference

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[1]	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 18 and 28 digits. Scale 0 to 28. If you specify a precision less than 18 or 28 digits, 18 or 28 is considered as the precision. For mappings in advanced mode- Precision 18, 28, and 38 digits. Scale 0 to 38.
Double	Double	Precision 15
Float	Double	Precision 15
Int	Integer	-2,147,483,648 to 2,147,483,647 Precision 10, scale 0
Long	Bigint	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision 19, scale 0
Map[1]	Map	Unlimited number of characters
Null	Integer	-2,147,483,648 to 2,147,483,647 Precision 10, scale 0
Record[1]	Struct	Unlimited number of characters
String	String	-1 to 104,857,600 characters

Avro Amazon S3 File Data Type	Transformation Data Type	Range and Description
Union[1]	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.
[1]Applies only to mappings in advanced mode.		

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

Timestamp Data type	Mapping	Mapping in advanced mode
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.
- When you use the **Create Target** option and modify the precision of an Avro data type in the upstream transformation, the precision is not honored.
  To resolve the issue with the precision, create a target object with the default precision and subsequently edit the precision.

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.

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 for mappings in advanced mode.

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

ORC Amazon S3 File Data Type	Transformation Data Type	Range and Description
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 mappings- Precision 18 and 28 digits. Scale 0 to 28. If you specify a precision less than 18 or 28 digits, 18 or 28 is considered as the precision. For mappings in advanced mode- Precision 18, 28, and 38 digits. Scale 0 to 38.
Double	Double	Precision 15
Float	Double	Precision 15

Parquet Amazon S3 File Data Type	Transformation Data Type	Range and Description
Int32	Integer	-2,147,483,648 to +2,147,483,647
Int64	Bigint	-9,223,372,036,854,775,808 to +9,223,372,036,854,775,807 8-byte signed integer
Int96	Binary	12-byte signed integer
Map[1]	Map	Unlimited number of characters.
Struct[1]	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)[1]	Array	Unlimited number of characters.
[1]Applies only to mappings in advanced mode.		

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.

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	Mappings in advanced mode
Timestamp_micros	Yes	No
Timestamp_millis	Yes	No
Time_millis	Yes	No
Time_micros	Yes	No

Timestamp Data type	Mapping	Mappings in advanced mode
int96	Yes	Yes
Date	Yes	Yes

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

- Timestamp_nanos
- Time_nanos
- Timestamp_tz

CHAPTER 8

Troubleshooting

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 a large number 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 Server.
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 Server.
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 Server.
5. Enter the **Name** as **FFParserRetainNullString** and the **Value** as **true**.
6. Click **Save**.

RedirectToSessionLog custom property

When you select a source or a target Amazon S3 V2 file and configure the Parquet file format option, a large number of log files might be generated. To resolve this issue and disable the logs, perform the following tasks and configure the custom property `RedirectToSessionLog`:

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 **DTM** for the Data Integration Server.
5. Enter the **Name** as **RedirectToSessionLog** and the **Value** as **No**.
6. Click **Save**.

Empty struct data in JSON file

If a JSON file has a field with an empty struct data, the Secure Agent ignores the field and reads the remaining fields during metadata read.

For example, if the JSON file has the following data in the first row: `{"id":123,"address":{}}`, the `address` field is ignored and does not appear in the **Fields** tab. If the JSON file has values for the `address` field in the consecutive row, you can use the **Data elements to sample** property to fetch this field.

Amazon S3 bucket does not exist or the user does not have permission to access the bucket

Do not modify the time on the machine that hosts the Secure Agent. The time on the Secure Agent must be correct as per the time zone. Otherwise, the mapping fails with an exception.

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://docs.informatica.com/integration-cloud/cloud-data-integration-connectors/h2l/1207-frequently-asked-questions-for-amazon-s3-v2-connector/abstract.html.

How can I configure AWS IAM authentication for Amazon S3 V2 Connector?

For information about configuring AWS IAM authentication, see https://docs.informatica.com/integration-cloud/cloud-data-integration-connectors/h2l/1199-configuring-iam-authentication-for-amazon-s3-and-amazon-s3-/abstract.html.

How can I grant folder-level and object-level access to the users?

For information about granting and restricting user access within the Amazon S3 bucket, see https://docs.informatica.com/integration-cloud/data-integration-connectors/h2l/1199-configuring-iam-authentication-for-amazon-s3-and-amazon-s3-/configuring-iam-authentication-for-amazon-s3-and-amazon-s3-v2-co/create-a-minimal-amazon-s3-bucket-policy/grant-folder-level-and-object-level-access.html

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://docs.informatica.com/integration-cloud/cloud-data-integration-connectors/h2l/1271-reading-a-json-file-using-an-amazon-s3-v2-connector/abstract.html.

Can I run a mapping in advanced mode to read or write multi-level JSON files?

Yes. For more information about reading or writing multi-level JSON files from a mapping in advanced mode, see https://kb.informatica.com/faq/7/Pages/23/574694.aspx?myk=574694.

INDEX

A

Amazon S3
  specifying targets 60
Amazon S3 and transformation
  data types 75
Amazon S3 Connector
  introduction 8
Amazon S3 V2
  handling multiple files 33
  directory source 32
  formatting options 57
  lookups 57
  Source transformation 51
  sources 30
  Sources in mappings 51
  supported assets 8
  targets 38
Amazon S3 V2 connection
  overview 10
Amazon S3 V2 Connector
  overview 8
Amazon S3 V2 sources
  client-side encryption 31
Amazon S3 V2 target
  mappings 54

C

Cloud Application Integration community
  URL 6
Cloud Developer community
  URL 6
create target
  adding time stamps 62
  target file parameterization 62
creating an mapping
  rules and guidelines 63

D

data compression
  sources and targets 47
data encryption
  Informatica encryption 32
  sources 31
  targets 38
Data Integration community
  URL 6
data type reference
  overview 75
distribution column 41

I

incrementally load files
  overview 34
Informatica Global Customer Support
  contact information 7
Informatica Intelligent Cloud Services
  web site 6

J

Java heap size
  configuration 82
JSON Amazon S3 file data types
  transformation data types 78

M

maintenance outages 7
mapping in advanced mode
  example 65
mappings
  Amazon S3 V2 Source properties 51

O

object tag
  rules and guidelines 43
object tags 42, 43
ORC file data types
  transformation data types 78
overwriting
  existing files 40

P

parameterization through a parameter file 63
partitioning
  Amazon S3 V2 sources 36
Partitioning
  Amazon S3 V2 targets 40

R

reading compressed flat file 48
reading compressed JSON file 48
rules and guidelines
  setting formatting options 60

S

source
FileName field 37
Source transformation
Amazon S3 V2 properties 51
Sources
Amazon S3 V2 in mappings 51
status
Informatica Intelligent Cloud Services 7
system status 7

T

target
FileName field 41
troubleshooting
Amazon S3 V2 Connector 83
trust site
description 7

U

upgrade notifications 7

W

web site 6
wildcard character
overview 35