



Informatica® Data Integration - Free & PayGo

Amazon S3 V2 Connector

© Copyright Informatica LLC 2017, 2023

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: 2023-04-04

Table of Contents

Preface	5
Informatica Resources.	5
Informatica Documentation.	5
Informatica Intelligent Cloud Services web site.	5
Informatica Intelligent Cloud Services Communities.	5
Informatica Intelligent Cloud Services Marketplace.	5
Data Integration connector documentation.	6
Informatica Knowledge Base.	6
Informatica Intelligent Cloud Services Trust Center.	6
Informatica Global Customer Support.	6
 Chapter 1: Introduction to Amazon S3 V2 Connector.....	 7
Amazon S3 V2 Connector assets.	7
Introduction to Amazon S3.	7
Authentication methods.	8
Administration of Amazon S3 V2 Connector.	8
Create a minimal Amazon IAM policy.	8
IAM authentication.	9
Temporary security credentials using AssumeRole	9
Credential profile file authentication.	12
Private communication with Amazon S3.	13
 Chapter 2: Amazon S3 V2 connections.....	 14
Amazon S3 V2 connection properties.	14
Rules and guidelines for AWS regions.	18
Rules and guidelines for S3 compatible storage.	18
Configuring proxy settings.	18
Configuring proxy settings on Windows.	19
Configuring proxy settings on Linux.	19
Configuring the proxy settings through JVMOptions.	19
Bypass the proxy server.	20
 Chapter 3: Amazon S3 V2 sources.....	 21
Amazon S3 V2 sources.	21
Data encryption in Amazon S3 V2 sources	21
Source types in Amazon S3 V2 sources.	22
Reading from multiple files.	23
Reading source objects path.	25
Pushdown optimization.	25
Data compression in Amazon S3 V2 sources.	26

Reading a compressed flat file.	26
Reading a compressed JSON file.	27
Chapter 4: Mappings and mapping tasks with Amazon S3 V2.....	28
Amazon S3 V2 objects in mappings.	28
Amazon S3 V2 sources in mappings.	28
Amazon S3 V2 lookups.	30
File formatting options.	30
Chapter 5: Data type reference.....	32
Flat file data types and transformation data types.	32
Avro Amazon S3 file data types and transformation data types.	33
ORC Amazon S3 file data types and transformation data types.	34
Parquet Amazon S3 file data types and transformation data types.	35
Chapter 6: Troubleshooting.....	37
Troubleshooting for Amazon S3 V2 Connector.	37
Troubleshooting FAQ.	37
Index.	39

Preface

Use *Amazon S3 V2 Connector* to learn how to read from Amazon S3 by using Data Integration. Learn to create a connection and configure mappings, mapping tasks, and 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, go to <https://status.informatica.com/> and click **SUBSCRIBE TO UPDATES**. You can then choose to receive notifications sent as emails, SMS text messages, webhooks, RSS feeds, or any combination of the four.

Informatica Global Customer Support

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

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

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

CHAPTER 1

Introduction to Amazon S3 V2 Connector

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 read primitive data types for Avro, Parquet, JSON and ORC files.

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

You cannot use Amazon S3 V2 Connector to read 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
- 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 read data from a source from a list of configured source connections. You can accomplish the tasks using the AWS Management Console web interface.

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

Authentication methods

You can access Amazon S3 private and public buckets only by configuring a valid Amazon S3 authentication.

Amazon S3 V2 Connector supports the following authentication methods:

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

Administration of Amazon S3 V2 Connector

As a user, you can use Amazon S3 V2 Connector after the organization administrator creates a minimal Amazon Identity and Access Management (IAM) policy for Amazon S3 V2 Connector. Apart from this, the organization administrator performs the following tasks:

Prerequisites for client-side and server-side encryption

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

Prerequisites for Informatica encryption

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

Prerequisites for temporary security credentials via AssumeRole

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

Create 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 successfully read data from an Amazon S3 bucket:

- `GetObject`
- `ListBucket`

IAM authentication

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

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

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

Perform the following steps to configure IAM authentication on EC2:

1. Create 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.

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

Temporary security credentials using AssumeRole

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

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

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

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

External ID

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

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

The external ID must be a string.

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

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

Temporary security credentials policy

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

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

IAM user

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

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

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

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 " }
```

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

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

- GenerateDataKey
- DescribeKey
- Encrypt
- Decrypt
- ReEncrypt

Sample policy:

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

Temporary security credentials using AssumeRole for EC2

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

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

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

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

For more information about the policies for EC2 roles and IAM roles, see [“Temporary security credentials policy” on page 10](#).

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

Rules and guidelines for using the temporary security credentials

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

- The IAM user or IAM role that requests for the temporary security credentials must not have access to any AWS resources.

- Only authenticated IAM users or IAM roles can request for the temporary security credentials from the AWS Security Token Service (AWS STS).
- Before you run a task, ensure that you have enough time to use the temporary security credentials for running the task. You cannot extend the time duration of the temporary security credentials for an ongoing task. For example, when you read from Amazon Redshift and if the temporary security credentials expire, you cannot extend the time duration of the temporary security credentials that causes the task to fail.
- After the temporary security credentials expire, AWS does not authorize the IAM users or IAM roles to access the resources using the credentials. You must request for new temporary security credentials before the previous temporary security credentials expire in a mapping.
- 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 a complex file such as Avro, ORC, or Parquet file depends on the Hadoop distribution in your environment. However, to read data from a flat file using the temporary security credentials, no Hadoop distribution is required by Amazon S3 V2 Connector.
- In a mapping, if you configure two or more Amazon S3 data sources for the same Amazon S3 bucket with different IAM roles, either of the IAM roles must be able to access the other data source as well.
- In a mapping, if you configure one Amazon S3 data source with user credentials and the other Amazon S3 data source with an IAM role, consider the following rules:
 - The user credentials for the first data source must also be able to assume the IAM role of the second Amazon S3 data source.
 - The IAM role that you configured for the second data source must also have access to the first Amazon S3 data source.

Credential profile file authentication

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

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

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

Consider the following rules for a credential profile file:

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

`~/.aws/credentials`

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

A sample credential profile file:

```
[default]
aws_access_key_id = 1233333
aws_secret_access_key = abcabcabc
```

```
[test-profile]

aws_access_key_id = 1233333

aws_secret_access_key = abcabcabc

aws_session_token = jahaheieomdrftflmlioerp
```

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

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

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.](#)

CHAPTER 2

Amazon S3 V2 connections

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

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

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

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

Amazon S3 V2 connection properties

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

The following table describes the Amazon S3 V2 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	The Amazon S3 V2 connection type.
Runtime Environment	Name of the runtime environment where you want to run the tasks. You can specify a Secure Agent or a Hosted Agent.
Access Key	Access key to access the Amazon S3 bucket. Enter the access key value based on the following authentication methods: <ul style="list-style-type: none">- Basic authentication. Enter the actual access key value.- IAM authentication.- Temporary security credentials using assume role. Enter the secret access key of an IAM user with no permissions to access Amazon S3 bucket.- Assume role for EC2.- Credential profile file authentication.- Federated user single sign-on. Don't enter the secret access key value.

Property	Description
Secret Key	<p>Secret access key to access the Amazon S3 bucket. The secret key is associated with the access key and uniquely identifies the account.</p> <p>Enter the secret access key value based on the following authentication methods:</p> <ul style="list-style-type: none"> - Basic authentication. Enter the actual access secret value. - IAM authentication. - Temporary security credentials using assume role. Enter access secret of an IAM user with no permissions to access Amazon S3 bucket. - Assume role for EC2. - Credential profile file authentication. - Federated user single sign-on. Don't enter the access secret value.
IAM Role ARN	<p>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.</p> <p>Enter the value of this property if you want to use the temporary security credentials to access the AWS resources.</p> <p>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.</p> <p>For more information about how to get the ARN of the IAM role, see the AWS documentation.</p>
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	<p>Enables the EC2 role to assume another IAM role specified in the IAM Role ARN option.</p> <p>Note: The EC2 role must have a policy attached with a permission to assume an IAM role from the same or different account.</p> <p>By default, the Use EC2 Role to Assume Role check box is not selected.</p>
Folder Path	<p>Bucket name or complete folder path to the Amazon S3 objects.</p> <p>Don't use a slash at the end of the folder path. For example, <bucket name>/<my folder name>.</p>
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	<p>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.</p> <p>You must generate the customer master key for the same region where the Amazon S3 bucket resides.</p> <p>You can specify the following master keys:</p> <ul style="list-style-type: none"> - 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.
S3 Account Type	<p>The type of the Amazon S3 account.</p> <p>Select from the following options:</p> <ul style="list-style-type: none"> - 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 Scalify RING or MinIO. <p>Default is Amazon S3 storage.</p>

Property	Description
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 .
Region Name	The AWS region of the bucket that you want to access. Select one of the following regions: <ul style="list-style-type: none"> - Africa(Cape Town) - Asia Pacific(Mumbai) - Asia 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(Beijing) - 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).
Federated SSO IdP	SAML 2.0-enabled identity provider for the federated user single sign-on to use with the AWS account. Amazon S3 V2 connector supports only the ADFS 3.0 identity provider. Select None if you don't want to use federated user single sign-on.
Other Authentication Type	Select one the following authentication types: <ul style="list-style-type: none"> - NONE - Credential Profile File Authentication Select the Credential Profile File Authentication option to access the Amazon S3 credentials from a credential file that contains the access key and secret key.
Credential Profile File Path	Specifies 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: <code>~/.aws/credentials</code>

Property	Description
Profile Name	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.
S3 VPC Endpoint Type	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 VPC endpoint types: <ul style="list-style-type: none"> - None - Gateway Endpoint - Interface Endpoint Default is None.
Endpoint DNS Name for Amazon S3	The DNS name for the Amazon S3 interface endpoint. Enter the DNS name in the following format: <code>bucket.<DNS name of the interface endpoint></code>
STS VPC Endpoint Type	Applicable when you select the S3 VPC interface endpoint. The VPC endpoint type for AWS STS. When you select IAM Role ARN or Federated SSO IdP , configure the STS VPC endpoint.
Endpoint DNS Name for AWS STS service	The DNS name for the AWS STS interface endpoint.
KMS VPC Endpoint Type	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 service	The DNS name for the AWS KMS interface endpoint.

Federated user single sign-on connection properties

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

Property	Description
Federated User Name	User name of the federated user to access the AWS account through the identity provider.
Federated User Password	Password for the federated user to access the AWS account through the identity provider.
IdP SSO URL	Single sign-on URL of the identity provider for AWS.
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.

Rules and guidelines for AWS regions

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

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

```
<Secure agent installation directory>/downloads/<package.AmazonS3V2.zip>/package/s3/thirdparty/infra.amazon.s3
```
- When you edit an existing connection, you see duplicate entries for regions. Use the regions that contain spaces because these regions are populated from AWS SDK. For example, use `US West (Oregon)` instead of `US West(Oregon)`.

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 Scalify RING S3 compatible storage. You cannot configure SSE and SSE-KMS encryption for MinIO S3 compatible storage.
- You cannot configure pushdown optimization to load data from Amazon S3 sources to Amazon Redshift.

Configuring proxy settings

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

You can set the proxy details in both the `proxy.ini` file and the JVM options. Use the following methods to configure the proxy settings:

- **Proxy.ini file.**
 - To use the proxy server at design time for mappings on a Windows system, pass the proxy values to the `proxy.ini` file on using the Secure Agent Manager. The values are updated in the `proxy.ini` file.
 - Set the values directly in the `proxy.ini` file to enable the proxy server in on a Linux system.
- **JVM options.** To use the proxy server at run time, set the proxy server details on Windows or Linux through the JVM options.

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

Configuring proxy settings on Windows

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

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

The Secure Agent Manager displays the Secure Agent status.

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

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

5. Click **OK**.

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

Configuring proxy settings on Linux

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

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

1. Navigate to the following directory:

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

2. Update the unauthenticated proxy details directly in the `proxy.ini` file, or run the following command to update the proxy details in the `proxy.ini` file:

```
./consoleAgentManager.sh configureProxy <proxy host> <proxy port>
```

3. Restart the Secure Agent.

Configuring the proxy settings through JVMOptions

You can configure the proxy server settings through JVMOptions for the Secure Agent on a Windows or Linux machine.

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

- Add the following parameters in any **JVMOption** field and specify appropriate values for each parameter:

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

For example,

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

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

6. Click **Save**.

The Secure Agent restarts to apply the settings.

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

CHAPTER 3

Amazon S3 V2 sources

You can configure the Amazon S3 V2 sources to read from 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	Flat	JSON File	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

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.

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.

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 .

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 . 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"
      ]
    }
  ],
  "settings":
  {
    "stopOnFail": "true"
  }
}

```

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`.

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.

Feature	Mapping
File name	<code>xyz.amazonaws.com/aa.bb.bucket/1024/characterscheckfor1024</code>
Directory name	<code><absolute path of the file including the file name></code>

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

To change the format for the `FileName` field to use `.`, set the JVM option `changes3EndpointForFileNamePort` = `true`. For example, `s3.us-west-2.amazonaws.com/<bucket_name>/automation/customer.avro`.

Pushdown optimization

You can enable full pushdown optimization when you want to load data from Amazon S3 sources to your data warehouse in Amazon Redshift. While loading the data to Amazon Redshift, you can transform the data as per your data warehouse model and requirements. When you enable full pushdown on a mapping task, the

mapping logic is pushed to the AWS environment to leverage AWS commands. Full pushdown optimization is enabled by default in mapping tasks.

For more information on pushdown 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.

Data compression in Amazon S3 V2 sources

You can decompress data when you read data from Amazon S3 .

The following table lists the supported source compression formats:

Compression format	Avro File	Binary File	Flat	JSON File	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

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

CHAPTER 4

Mappings and mapping tasks with Amazon S3 V2

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

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

Amazon S3 V2 objects in mappings

You can configure a source object to read data from Amazon S3 by specifying the source properties and advanced source properties. You can perform lookup tasks and specify file formatting options in the Amazon S3 V2 mappings. You can parameterize the connection and objects. You can also use a parameter file in the task properties to overwrite the parameters at runtime.

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: <ul style="list-style-type: none">- Single Object- Parameter. Select Parameter to define the source type when you configure the mapping task.

Property	Description
Object	Name of the source object. When you select an object, you can also select a <code>.manifest</code> 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 Overwrite Parameter option.
Format	Specifies the file format that the Amazon S3 V2 Connector uses to read data from Amazon S3. You can select the following file format types: <ul style="list-style-type: none"> - None - Flat - Avro - ORC - Parquet 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 30 .

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: <ul style="list-style-type: none"> - File - Directory Default is File . For more information, see "Source types in Amazon S3 V2 sources" on page 22 .
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 <code><bucket_name>/<folder_name></code> format. If you do not provide the bucket name and specify the folder path starting with a slash (/) in the <code></folder_name></code> format, the folder path appends with the folder path that you specified in the connection properties. For example, if you specify the <code></dir2></code> folder path in this property and <code><my_bucket1>/<dir1></code> folder path in the connection property, the folder path appends with the folder path that you specified in the connection properties in <code><my_bucket1>/<dir1>/<dir2></code> format. If you specify the <code><my_bucket1>/<dir1></code> folder path in the connection property and <code><my_bucket2>/<dir2></code> folder path in this property, the Secure Agent reads the file in the <code><my_bucket2>/<dir2></code> folder path that you specify in this property.
File Name	Overwrites the Amazon S3 source file name.
Encryption Type	Method you want to use to decrypt data. Default is None . Note: You cannot select client-side encryption, server-side encryption, and server-side encryption with KMS encryption types.

Property	Description
Staging Directory	Path of the local staging directory. Default staging directory is the <code>/temp</code> directory on the machine that hosts the Secure Agent. 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: <ul style="list-style-type: none"> - None - Lzo Default is None . Note: Amazon S3 V2 Connector does not support the Lzo compression format even though the option appears in this property.
Download Part Size	Downloads the part size of an Amazon S3 object in bytes. Default is 5 MB. Use this property when you read a file of flat format type.
Multiple Download Threshold	Minimum threshold size to download an Amazon S3 object in multiple parts. To download the object in multiple parts in parallel, ensure that the file size of an Amazon S3 object is greater than the value you specify in this property. Default is 10 MB.
Temporary Credential Duration	The time duration during which an IAM user can use the dynamically generated temporarily credentials to access the AWS resource. Enter the time duration in seconds. Default is 900 seconds. If you require more than 900 seconds, you can set the time duration maximum up to 12 hours in the AWS console and then enter the same time duration in this property.
Tracing Level	This property is not applicable for Amazon S3 V2 Connector.

Amazon S3 V2 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 flat files:

Property	Description
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. If you select Import from schema file , you can select Schema File to upload a schema file.

Property	Description
Flat File Type	The type of flat file. Default is Delimited.
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. Do not specify a multibyte character as a delimiter in the source object.
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 <code>\<decimal value></code> . When you specify <code>\<decimal value></code> , the agent considers the ASCII character for the decimal value as the escape character. For example, if you specify <code>\64</code> , the agent considers the ASCII character @. To ignore the escape character, specify <code>\0</code> .
Qualifier	Quote character that defines the boundaries of data. You can set the qualifier as single quote or double quote.
Code Page	Select the code page that the agent must use to read data. Amazon S3 V2 Connector supports the following code pages: <ul style="list-style-type: none"> - MS Windows Latin 1. Select for ISO 8859-1 Western European data. - UTF-8. Select for Unicode and non-Unicode data. - Shift-JIS. Select for double-byte character data. - ISO 8859-15 Latin 9 (Western European). - ISO 8859-2 Eastern European. - ISO 8859-3 Southeast European. - ISO 8859-5 Cyrillic. - ISO 8859-9 Latin 5 (Turkish). - IBM EBCDIC International Latin-1.
Header Line Number	Specify the line number that you want to use as the header when you read data from Amazon S3. You can also read a file that does not have a header. 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. Default is 1.
First Data Row	Specify the line number from where you want the Secure Agent to read data. You must enter a value that is greater or equal to one. To read data from the header, the value of the Header Line Number and the First Data Row fields should be the same. Default is 1.
Is Escape Character Data Retained	Not applicable to Amazon S3 V2 Connector.
Max Rows To Preview	Not applicable to Amazon S3 V2 Connector.
Row Delimiter	Not applicable to Amazon S3 V2 Connector.

CHAPTER 5

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 transformation 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.

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
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.
Double	Double	Precision 15
Float	Double	Precision 15
Int	Integer	-2,147,483,648 to 2,147,483,647 Precision 10, scale 0
Long	Bigint	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision 19, scale 0
Null	Integer	-2,147,483,648 to 2,147,483,647 Precision 10, scale 0
String	String	-1 to 104,857,600 characters

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

Timestamp Data type	Mapping
Timestamp_micros	Yes
Timestamp_millis	Yes
Time_millis	Yes
Time_micros	Yes

ORC Amazon S3 file data types and transformation data types

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

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

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

Parquet Amazon S3 file data types and transformation data types

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

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

Parquet Amazon S3 File Data Type	Transformation Data Type	Range and Description
Boolean	Integer	TRUE (1) or FALSE (0)
Date	Date/Time	January 1,0001 to December 31,9999
Decimal	Decimal	For 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.
Double	Double	Precision 15
Float	Double	Precision 15
Int32	Integer	-2,147,483,648 to +2,147,483,647
Int64	Bigint	-9,223,372,036,854,775,808 to +9,223,372,036,854,775,807 8-byte signed integer
Int96	Binary	12-byte signed integer
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.

The Parquet schema that you specify to read 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
Timestamp_micros	Yes
Timestamp_millis	Yes
Time_millis	Yes
Time_micros	Yes
int96	Yes
Date	Yes

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

- Timestamp_nanos
- Time_nanos
- Timestamp_tz

CHAPTER 6

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.

RedirectToSessionLog custom property

When you select a source 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**.

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/cloud-data-integration-connectors/h2l/1199-configuring-iam-authentication-for-amazon-s3-and-amazon-s3-/abstract.html>

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>.

INDEX

A

- administration
 - IAM authentication [9](#)
 - minimal Amazon IAM policy [8](#)
- Amazon S3 and transformation
 - data types [32](#)
- Amazon S3 Connector
 - introduction [7](#)
- Amazon S3 V2
 - handling multiple files [23](#)
 - connection properties [14](#)
 - directory source [22](#)
 - formatting options [30](#)
 - lookups [30](#)
 - Source transformation [28](#)
 - sources [21](#)
 - Sources in mappings [28](#)
 - supported assets [7](#)
- Amazon S3 V2 connection
 - overview [14](#)
- Amazon S3 V2 Connector
 - administration [8](#)
 - overview [7](#)
- Amazon S3 V2 sources
 - client-side encryption [21](#)

C

- Cloud Application Integration community
 - URL [5](#)
- Cloud Developer community
 - URL [5](#)
- connections
 - Amazon S3 V2 [14](#)

D

- data compression [26](#)
- data encryption
 - sources [21](#)
- Data Integration community
 - URL [5](#)
- data type reference
 - overview [32](#)

I

- Informatica Global Customer Support
 - contact information [6](#)
- Informatica Intelligent Cloud Services
 - web site [5](#)

J

- Java heap size
 - configuration [37](#)

L

- Linux
 - configuring proxy settings [19](#)

M

- maintenance outages [6](#)
- mappings
 - Amazon S3 V2 Source properties [28](#)

O

- ORC file data types
 - transformation data types [34](#)

P

- proxy settings
 - configuring on Linux [19](#)
 - configuring on Windows [19](#)
 - JVMOptions [19](#)

R

- reading compressed flat file [26](#)
- reading compressed JSON file [27](#)

S

- source
 - FileName field [25](#)
- Source transformation
 - Amazon S3 V2 properties [28](#)
- Sources
 - Amazon S3 V2 in mappings [28](#)
- status
 - Informatica Intelligent Cloud Services [6](#)
- system status [6](#)

T

- temporary security credentials
 - policy [10](#)

troubleshooting
 Amazon S3 V2 Connector [37](#)
trust site
 description [6](#)

U

upgrade notifications [6](#)

W

web site [5](#)
Windows
 configuring proxy settings [19](#)