AWS S3 Bucket Connectivity with IICS CDI

Shubham Goel
Informatica Global Customer Support

Informatica

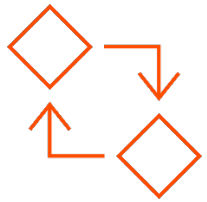
Housekeeping Tips

- Today's Webinar is scheduled for 1 hour
- The session will include a webcast and then your questions will be answered live at the end of the presentation
- All dial-in participants will be muted to enable the speakers to present without interruption
- Questions can be submitted to "All Panelists" via the Q&A option and we will respond at the end of the presentation
- The webinar is being recorded and will be available to view on our INFASupport YouTube channel and Success Portal. The link will be emailed as well.
- Please take time to complete the post-webinar survey and provide your feedback and suggestions for upcoming topics.

Feature Rich Success Portal

Bootstrap trial and POC Customers

Enriched Customer Onboarding experience



Product Learning Paths and Weekly Expert Sessions

Informatica Concierge with Chatbot integrations

Tailored training and content recommendations

Informatica

More Information

Success Portal

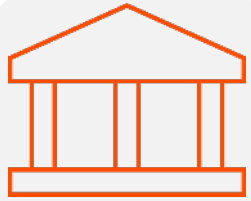
https://success.informatica.com

Communities & Support

https://network.informatica.com

Documentation

https://docs.informatica.com



University

https://www.informatica.com/in/services-and-training/informatica-university.html

Informatica

Safe Harbor

The information being provided today is for informational purposes only. The development, release, and timing of any Informatica product or functionality described today remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision.

Statements made today are based on currently available information, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products or functionality in the future.

Agenda

- Introduction.
- Basics of AWS S3.
- Prerequisites.
- Minimal S3 Bucket Policy.
- Amazon S3 V2 Connection properties.
- Authentication & Encryption.
- Designing Integration.
- Demo – AWS S3 V2 .
- Q&A.

Basics of AWS S3

Amazon Simple Storage Service, also known as Amazon S3, is a highly scalable, fast, and durable object-level storage of any data type provided by Amazon as part of their Amazon Web Services (AWS).

S3 bucket allows users to upload files, videos, and documents like you were to upload files, videos, and documents to popular cloud storage products like Dropbox and Google Drive.
This makes Amazon S3 very flexible and platform independent.

Some major advantages of using Amazon S3 include durability, security, and reliability.

AWS S3 redundantly stores your data across multiple devices spanning in AZs (Availability Zones) in an S3 Region.

Prerequisites

You can use Amazon S3 V2 Connector after the organization administrator performs the following tasks:

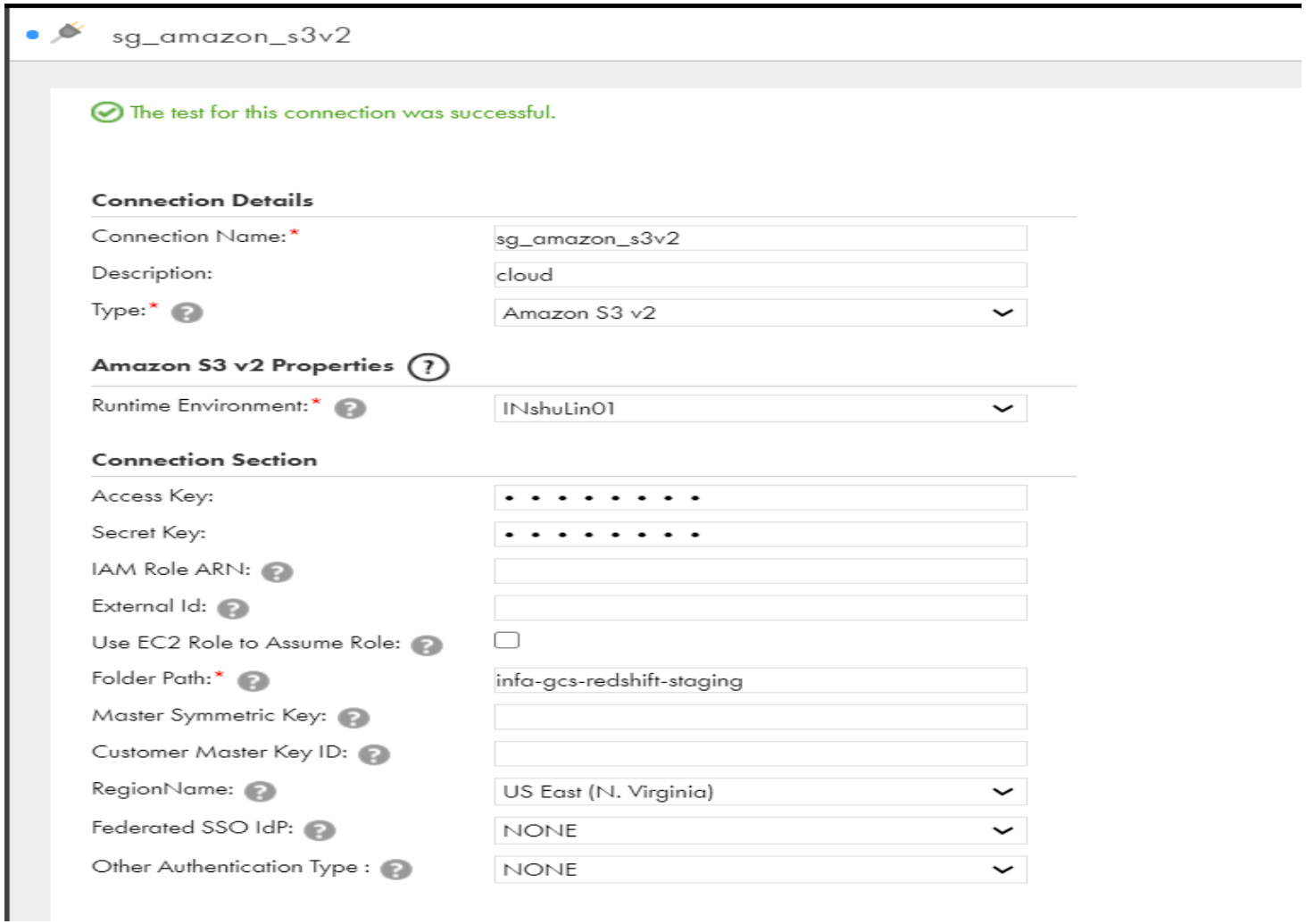
- Create a minimal Amazon S3 bucket policy for Amazon S3 V2 Connector.
- To run a mapping that reads from or writes to a complex file, ensure that either Cloudera 5.8, Cloudera 6.1, Hortonworks 2.5, or Hortonworks 2.6 license is enabled.
- Ensure that Cloudera 5.8 or Cloudera 6.1 license is enabled to preview the data successfully when you create an elastic mapping that reads from or writes to an Avro, JSON, ORC, or Parquet file.

Minimal S3 Bucket Policy

- The minimal Amazon S3 bucket policy restricts user operations and user access to particular Amazon S3 buckets by assigning an AWS Identity and Access Management (IAM) policy to users.
- You can configure the IAM policy through the AWS console. Use AWS Identity and Access Management (IAM) authentication to securely control access to Amazon S3 resources.
- You can use the following minimum required actions for users to successfully read data from and write data to Amazon S3 bucket:
- Sample Policy:

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

Amazon S3 V2 Connection Details



Informatica

Amazon S3 V2 connection properties

- Connection Name : Name of the connection.
- Description : Optional. Description of the connection.
- Type : The Amazon S3 V2 connection type.
- Runtime Environment : Name of the runtime environment where you want to run the tasks.
- Access Key : Access key to access the Amazon S3 bucket.
- Secret Key : Secret access key to access the Amazon S3 bucket.
- IAM Role ARN : The ARN of the IAM role assumed by the user to use the dynamically generated temporary security credentials.
- External Id : Optional. Specify the external ID for a more secure access to the Amazon S3 bucket when the Amazon S3 bucket is in a different AWS account.
- Use EC2 Role to Assume Role : Optional. Select the check box to enable the EC2 role to assume another IAM role specified in the IAM Role ARN option.

Amazon S3 V2 connection properties

- Folder Path : Bucket name or complete folder path to the Amazon S3 objects.
- Master Symmetric Key : Optional. Provide a 256-bit AES encryption key in the Base64 format when you enable client-side encryption.
- Customer Master Key ID : Optional. Specify the customer master key ID or alias name generated by AWS Key Management Service (AWS KMS) or the Amazon Resource Name (ARN) of your custom key for cross-account access.
- Region Name : The AWS region of the bucket that you want to access.
- Federated SSO IdP : SAML 2.0-enabled identity provider for the federated user single sign-on to use with the AWS account.
- Other Authentication Type : Select NONE or Credential Profile File Authentication.

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.

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.

IAM Authentication

Perform the following steps to configure IAM authentication on EC2:

1.Create minimal Amazon S3 bucket policy.

2.Create the Amazon EC2 role. The Amazon EC2 role is used when you create an EC2 system. For more information about creating the Amazon EC2 role, see the AWS documentation.

3.Link the minimal Amazon S3 bucket policy with the Amazon EC2 role.

4.Create an EC2 instance. Assign the Amazon EC2 role that you created in step #2 to the EC2 instance.

5.Install the Secure Agent on the EC2 system.

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

Designing Integration using Mappings and Mapping Tasks in IICS Cloud Data Integration

Source Transformation

You can use Source Transformation in the mapping to connect to Amazon S3 to read data from Amazon S3 by configuring the Amazon S3 V2 source and advanced properties for the source object.

You can select the following file format types:

-None

-Delimited

-Avro

-ORC

-Parquet

Note: You need to open the Formatting Options dialog box to define the format of the file.

If you select None is as the format type, the Secure Agent reads data from Amazon S3 files in binary format.

Target Transformation

You can use a Target transformation to connect to Amazon S3 V2 object as the target to write data to Amazon S3 by configure the Amazon S3 V2 target and advanced properties for the target object.

You have an option to create Target at runtime by specifying the name and path for the target object.

For example,

- If you specify the path as folder1/folder2 and target object name as Records, then the Secure Agent will create the target object in the following location: <bucket_name>/folder1/folder2/Records.

Target Transformation

- - If you do not specify the path and specify the target object name as Records, the Secure Agent will create the target object within the bucket that you specify in the Folder Path connection property in the following location: <bucket_name>/Records.
- - If you do not specify the path, the Secure Agent will create target object name within the bucket that you specify in the Folder Path connection property in the following format: <bucket_name>/<target_object_name>.

Encryption

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

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

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

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

IICS Amazon S3 V2 connector DEMO

- I will be showing you a demo on how to read data from and write data to S3 bucket using Amazon S3 V2 connector in IICS.

Configuring Proxy Settings

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

Contact your network administrator for the correct proxy settings.

Consider the following rules when you configure a proxy server:

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

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

- Proxy settings do not apply to elastic mappings.

- You cannot use an unauthenticated proxy server when you read from complex files.

Amazon S3 V2 Debugging options

- To capture the Amazon S3 V2 connector Request/Response in the session log of the task, enable log level debug property at the agent level:

Troubleshooting Options

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

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

1.Select **Administrator** > **Runtime Environments**.
2.On the **Runtime Environments** page, select the Secure Agent for which you want to increase memory from the list of available Secure Agents.
3.In the upper-right corner, click **Edit**.
4.In the **System Configuration Details** section, select the **Type** as **DTM** for the Data Integration Service.
5.Edit the **JVMOption1** as **-Xmx8046m**.
6.Restart the Secure Agent manually.

Troubleshooting Options

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

1.Select **Administrator** > **Runtime Environments**.

2.On the **Runtime Environments** page, select the Secure Agent.

3.In the upper-right corner, click **Edit**.

4.In the **Custom Configuration Details** section, select the **Type** as **Tomcat JRE** for the Data Integration Service.

5.Enter the **Name** as **FFParserRetainNullString** and the **Value** as **true**.

6.Click **Save**

References

- Amazon S3 V2 connector guide

https://network.informatica.com/onlinehelp/IICS/prod/CDI/en/index.htm#page/cloud-data-integration-amazon-s3-v2-connector/Introduction_to_Amazon_S3_V2_Connector.html

- Informatica Cloud Data Integration Amazon S3 V2 Connector Frequently Asked Questions

https://kb.informatica.com/h2l/HowTo%20Library/1/1207-InformaticaCDIAmazonS3V2ConnectorFAQs-H2L.pdf

Q & A


Thank You