Informatica® Cloud Data Integration
Summer 2018

Amazon Redshift V2 Connector Guide

Informatica Cloud Data Integration Amazon Redshift V2 Connector Guide
Summer 2018
December 2018

Publication Date: 2019-02-04

Table of Contents

CHAPTER 1

Introduction to Amazon Redshift V2 Connector

This chapter includes the following topics:

Amazon Redshift V2 Connector Overview

You can use Amazon Redshift V2 Connector to securely read data from and write data to Amazon Redshift. Amazon Redshift V2 sources and targets represent records in Amazon Redshift. When you read data from or write data to Amazon Redshift, you can specify the Secure Agent.

You can create an Amazon Redshift V2 connection and use the connection in mass ingestion tasks, mappings, and mapping tasks. Create a mass ingestion task to transfer files from any source that mass ingestion task supports to an Amazon Redshift target. Create a mapping task to process data based on the data flow logic defined in a mapping or integration template.

When you run an Amazon Redshift V2 mass ingestion task, mapping, or mapping task, the Secure Agent writes data to Amazon Redshift based on the workflow and Amazon Redshift V2 connection configuration. The Secure Agent connects and writes data to Amazon Simple Storage Service (Amazon S3) through a TCP/IP network. Amazon S3 is a storage service in which you can copy data from a source and simultaneously move data to Amazon Redshift clusters. The Secure Agent issues a copy command that copies data from Amazon S3 to the Amazon Redshift target table.

You can move data from any data source to Amazon Redshift. Data Integration uses the Amazon driver to communicate with Amazon Redshift.

Amazon Redshift V2 Connector supports Hosted Agent.

Note: Informatica recommends that you use Amazon Redshift Connector if you want to create a synchronization task or mapping to read data from and write data to Amazon Redshift. For more information about using Amazon Redshift Connector, see the Amazon Redshift Connector documentation.

Example

You work for an organization that stores purchase order details, such as customer ID, item codes, and item quantity in an on-premise MySQL database. You need to analyze purchase order details and move data from the on-premise MySQL database to an affordable cloud-based environment. Create a mapping to read all the purchase records from the MySQL database and write them to an Amazon Redshift target for data analysis.

Amazon Redshift V2 Supported Task Types

The following table lists the task types that Amazon Redshift V2 Connector supports:

Task Type	Source	Target
Mapping	Yes	Yes
Mass Ingestion	No	Yes

Introduction to Amazon Redshift

Amazon Redshift is a cloud-based petabyte-scale data warehouse service that organizations can use to analyze and store data.

Amazon Redshift uses columnar data storage, parallel processing, and data compression to store data and to achieve fast query execution. Amazon Redshift uses a cluster-based architecture that consists of a leader node and compute nodes. The leader node manages the compute nodes and communicates with the external client programs. The leader node interacts with the client applications and communicates with compute nodes. A compute node stores data and runs queries for the leader node. Any client that uses a PostgreSQL driver can communicate with Amazon Redshift.

Amazon Redshift Spectrum Overview

Amazon Redshift Spectrum enables you to run complex Amazon Redshift SQL queries on a large amount data of different formats stored in Amazon S3. With Amazon Redshift Spectrum, you can directly run queries to read Amazon S3 data files without the need to load or transform the data.

You can run queries for the large amount of Amazon S3 data files without the need to scale the specified Amazon Redshift cluster.

Amazon Redshift Spectrum resides on Amazon Redshift servers independent of the Amazon Redshift cluster. When you run queries using Amazon Redshift Spectrum, the queries run faster and uses less Amazon Redshift cluster processing capacity as Amazon Redshift Spectrum pushes all the compute-intensive tasks to the Amazon Redshift Spectrum layer.

External Schema and External Table

To use Amazon Redshift Spectrum, you must create an external table within an external schema that references a database in an external data catalog. You can create the external table for Avro, ORC, Parquet, RCFile, SequenceFlile, and Textfile file formats.

The metadata of the external database and external table are stored in the external data catalog. You must provide authorization to Amazon Redshift to access the data catalog and the data files in Amazon S3.

You can create an external database in Amazon Redshift. You can read data from a single external table, multiple external table, or from a standard Amazon Redshift table that is joined to the external table.

Multiple Amazon Redshift clusters can contain multiple external tables. You can run a query for the same data on Amazon S3 from any Amazon Redshift cluster in the same region. When you update the data in Amazon S3, the data is immediately available in all the Amazon Redshift clusters.

When you create an external table, you must specify the Amazon S3 location from where you want to read the data. You can create the external tables by defining the structure of the Amazon S3 data files and registering the external tables in the external data catalog. Then, you can run queries or join the external tables.

When you add an external table as source and create a mapping, the external table name is displayed in the `spectrum_schemaname` format in the **Select Source Object** dialog box.

Note: You can only read data from the Amazon Redshift Spectrum external table. You cannot insert or update data in the Amazon Redshift Spectrum external table.

When you create an external table using Athena or Glu data catalogs, ensure that you create the external tables using the data types that Amazon Redshift V2 Connector supports.

The following lists the data types that Amazon Redshift V2 Connector supports when you create an external table:

- Smallint (INT2)
- Integer (INT, INT4)
- Bigint (INT8)
- Decimal (NUMERIC)
- Real (FLOAT4)
- Double Precision (FLOAT8)
- Boolean (BOOL)
- Char (CHARACTER)
- Varchar (CHARACTER VARYING)
- Date

  Note: Applicable when you create an external table for the ORC, Parquet, and Textfile file formats.
- Timestamp

For more information on how to create an external table, see the AWS documentation.

Administration of Amazon Redshift V2 Connector

As a user, you can use Amazon Redshift V2 Connector after the organization administrator ensures that users have access to the Secure Agent directory that contains the success and error files. This directory path

must be the same on each Secure Agent machine in the runtime environment. The organization administrator must also perform the following tasks:

- Get the Amazon Redshift JDBC URL.
- Manage Authentication. Use either of the following two methods:
  - Create an Access Key ID and Secret Access Key.
    Provide the values for access key ID and secret access key when you configure the Amazon Redshift V2 connection. For more information about creating an access key ID and secret access key, see the AWS documentation.
  - Configure AWS Identity and Access Management (IAM) Authentication to enhance security.
    If you use IAM authentication, do not provide access key ID and secret access key explicitly in the Amazon Redshift V2 connection. Instead, you must create an Redshift Role Amazon Resource Name (ARN), add the minimal Amazon S3 bucket policy to the Redshift Role ARN, and add the Redshift Role ARN to the Redshift cluster.

    Provide the Redshift Role ARN in the AWS_IAM_ROLE option in the UNLOAD and COPY commands when you create a task.

    If you specify both, access key ID and secret access key in the connection properties and AWS_IAM_ROLE in the UNLOAD and COPY commands, AWS_IAM_ROLE takes the precedence.

    You must add IAM EC2 role and IAM Redshift role to the customer master key when you use IAM authentication and server-side encryption using customer master key.

    Hosted Agent does not support IAM authentication. For more information about how to configure IAM authentication for Amazon Redshift V2 Connector, see "IAM Authentication " on page 10
- Configure Amazon Redshift for SSL if you want to support an SSL connection.
- Create a master symmetric key if you want to enable client-side encryption.
- Create an AWS Key Management Service (AWS KMS)-managed customer master key if you want to enable server-side encryption.
- Create minimal Amazon S3 bucket policy for Amazon Redshift V2 Connector.
- To access the data catalog and the data files in Amazon S3 by using Amazon Redshift Spectrum, ensure that the Amazon Redshift cluster has the required authorization.
- Configure a CDC source if you want to create a mapping to capture changed data from the CDC source, and then run the associated mapping tasks to write the changed data to an Amazon Redshift target. To create a mapping with a CDC source, ensure that you have the PowerExchangeClient and CDC licenses.

Configure Amazon Redshift for SSL

You can configure the Secure Agent to support an SSL connection to Amazon Redshift.

1. Download the Amazon Redshift certificate from the following location: https://s3.amazonaws.com/redshift-downloads/redshift-ssl-ca-cert.pem.
2. Run the following command to add the certificate file to the key store: `${JAVA_HOME}/bin/keytool -keystore {JAVA_HOME}/lib/security/cacerts -import -alias <string_value> -file <certificate_filepath>`.
3. In Administrator, select **Runtime Environments**.
4. Select the Secure Agent for which you want to increase memory from the list of available Secure Agents.
5. In the upper-right corner, click **Edit**.
6. In the **System Configuration Details** section, change the **Type** to **DTM**.

7. Click the **Edit Agent Configuration** icon next to **JVMOption1** and add the following command: -Djavax.net.ssl.trustStore=<keystore_name>.
8. Click the **Edit Agent Configuration** icon next to **JVMOption2** and add the following command:-Djavax.net.ssl.trustStorePassword=<password>.
9. Add the following parameter to the JDBC URL you specified in your Amazon Redshift V2 connection properties: ssl=true. For example, jdbc:redshift://mycluster.xyz789.us-west-2.redshift.amazonaws.com:5439/dev?ssl=true.
10. Click **OK** to save your changes.

Create Minimal Amazon S3 Bucket Policy

The minimal Amazon S3 bucket policy restricts user operations and user access to particular Amazon S3 buckets by assigning an AWS IAM policy to users. You can configure the AWS IAM policy through the AWS console.

You can use the following minimum required permissions to successfully read data from and write data to Amazon Redshift resources:

- PutObject
- GetObject
- DeleteObject
- ListBucket

Sample Policy:

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

Note: The **Test Connection** does not validate the IAM policy assigned to users. The Amazon S3 bucket name is available in the advanced properties for source and target.

You must make sure that the Amazon S3 bucket and Amazon Redshift cluster reside in the same region to run the mapping tasks successfully.

You can only read data from or write data to the regions supported by AWS SDK used by the Connector. The supported regions are:

- Asia Pacific(Mumbai)
- Asia Pacific(Seoul)
- Asia Pacific(Singapore)
- Asia Pacific(Sydney)

- Asia Pacific(Tokyo)
- AWS GovCloud (US)
- Canada(Central)
- China(Bejing)
- China(Ningxia)
- EU(Ireland)
- EU(Frankfurt)
- EU(Paris)
- South America(Sao Paulo)
- US East(N. Virginia)
- US East(Ohio)
- US West(N. California)
- US West(Oregon)

IAM Authentication

Optional. You can configure IAM authentication when the Secure Agent is installed on an Amazon Elastic Compute Cloud (EC2) system. Use IAM authentication for secure and controlled access to Amazon Redshift resources when you run mappings and mapping tasks.

Use IAM authentication when you want to run the mappings and mapping tasks on Secure agent installed on an EC2 system. Perform the following steps to configure IAM authentication:

- Step 1: Create Minimal Amazon S3 Bucket Policy.
- Step 2: Create the Amazon EC2 role. The Amazon EC2 role is used when you create an EC2 system in the Redshift cluster. For more information about creating the Amazon EC2 role, see the AWS documentation.
- Step 3: Create an EC2 instance. Assign the Amazon EC2 role that you created in step #2 to the EC2 instance.
- Step 4: Create the Amazon Redshift Role ARN for secure access to Amazon Redshift resources. You can use the Amazon Redshift Role ARN in the UNLOAD and COPY commands. For more information about creating the Amazon Redshift Role ARN, see the AWS documentation.
- Step 5: Add the Amazon Redshift Role ARN to the Amazon Redshift cluster to successfully perform the read and write operations. For more information about adding the Amazon Redshift Role ARN to the Amazon Redshift cluster, see the AWS documentation.
- Step 6: Install Secure Agent on the EC2 system.

Amazon Redshift Spectrum Prerequisite Task

To read data from an Amazon Redshift Spectrum external table, you must provide the required authorization to Amazon Redshift cluster to access the data catalog and the data files in Amazon S3.

1. Create an AWS Identity and Access Management (IAM) role to authorize the Amazon Redshift cluster access to the external data catalog and data files in Amazon S3.
2. Associate the IAM Role with the specified Amazon Redshift cluster.
3. Create an external schema.
4. Provide Amazon Redshift Role ARN for the IAM Role in the external schema.

5. Create an external table within the external schema and specify the Amazon S3 location from where you want to read the data.
   For more information about creating external tables, see the AWS documentation.

Note: The Amazon Redshift cluster and the Amazon S3 bucket that contains the data files must belong to the same region. The Amazon Redshift cluster must be of version 1.0.1294 or later.

CHAPTER 2

Amazon Redshift V2 Connections

This chapter includes the following topics:

Amazon Redshift V2 Connections Overview

Amazon Redshift V2 connection enables you to read data from or write data to Amazon Redshift. You can use Amazon Redshift V2 connections to specify sources or targets in mappings and mapping tasks. You can use Amazon Redshift V2 connections to specify targets in mass ingestion tasks.

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

Create an Amazon Redshift V2 connection on the **Connections** page and associate it with a mapping, mapping task, or mass ingestion task. Define the source and target properties to read or write data to Amazon Redshift.

Note: If you enable both HTTP and SOCKS proxies, SOCKS proxy is used by default. If you want to use HTTP proxy instead of SOCKS proxy, set the value of the **DisableSocksProxy** property to true in the System property.

Amazon Redshift V2 connection properties

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

The following table describes the Amazon Redshift V2 connection properties:

Connection property	Description
Runtime Environment	The name of the runtime environment where you want to run the tasks.
Username	User name of the Amazon Redshift account.
Password	Password for the Amazon Redshift account.

Connection property	Description
AWS Access Key ID	Amazon S3 bucket access key ID.
AWS Secret Access Key	Amazon S3 bucket secret access key ID.
Master Symmetric Key	Optional. Provide a 256-bit AES encryption key in the Base64 format when you enable client-side encryption. You can generate a key using a third-party tool. If you specify a value, ensure that you specify the encryption type as client side encryption in the advanced target properties.
Customer Master Key ID	Optional. Specify the customer master key ID or alias name generated by AWS Key Management Service (AWS KMS). You must generate the customer master key ID for the same region where Amazon S3 bucket reside. You can either specify the customer generated customer master key ID or the default customer master key ID.
JDBC URL	Amazon Redshift V2 connection URL.

CHAPTER 3

Amazon Redshift V2 Sources and Targets

This chapter includes the following topics:

Amazon Redshift V2 Sources

You can use an Amazon Redshift V2 object as a source in a mapping. You can encrypt data, specify the location of the staging directory, and securely unload the results to Amazon Redshift.

When you configure the advanced source properties, you configure properties specific to Amazon Redshift V2. You can encrypt data, retain the staging files on Amazon S3, and securely unload the results of a query to files on Amazon Redshift.

Amazon Redshift Staging Directory for Amazon Redshift V2 Sources

The Secure Agent creates a staging file in the directory that you specify in the source properties. The Secure Agent reads the data from Amazon Redshift V2 source and writes the data to the staging directory before it writes data to Amazon S3.

The Secure Agent deletes the staged files from the staging directory after it writes the data to Amazon S3. Specify a staging directory in the mapping properties with an appropriate amount of disk space for the volume of data that you want to process. Specify a directory on the machine that hosts the Secure Agent.

Encryption Type

To protect data, you can encrypt the data when you read the data from a source.

You can select the type of the encryption in the **Encryption Type** field under the Amazon Redshift V2 advanced source properties on the **Schedule** page. The Unload command creates staging files on Amazon S3 for server-side encryption with the AWS-managed encryption keys and AWS Key Management Service key.

Use the customer master key ID generated by AWS Key Management Service in the Unload command for server-side encryption. You can select the following types of encryption:

None

The data is not encrypted.

SSE-S3

If you select the **SSE-S3** encryption type, the Unload command creates the staging files in the Amazon S3 bucket and Amazon S3 encrypts the file using AWS-managed encryption keys for server-side encryption.

SSE-KMS

If you select the **SSE-KMS** encryption type, the Unload command creates the staging files in the Amazon S3 bucket and Amazon S3 encrypts the file using AWS KMS-managed customer master key for server-side encryption.

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 when you select the **SSE-KMS** encryption type.

CSE-SMK

If you select the **CSE-SMK** encryption type, Amazon Redshift uploads the data to the Amazon S3 server by using the master symmetric key and then loads the data by using the copy command with the Encrypted option and a private encryption key for additional security.

You must provide a master symmetric key ID in the connection property to enable **CSE-SMK** encryption type.

To enable client-side encryption, perform the following tasks:

1. Provide the master symmetric key when you create an Amazon Redshift V2 connection.
   Ensure that you provide a 256-bit AES encryption key in Base64 format.
2. Update the `local_policy.jar` and the `US_export_policy.jar` files in the following directory: `<JAVA_HOME>\lib\security`.
   You can download the JAR files supported by your JAVA environment from the Oracle website.
3. Select the **S3 Client Side Encryption** option in the advanced source properties.

Note: Amazon Redshift V2 Connector does not support the server-side encryption with the master symmetric key and client-side encryption with the customer master key.

Unload Command

You can use the Unload command to extract data from Amazon Redshift and create staging files on Amazon S3. The Unload command uses a secure connection to load data into one or more files on Amazon S3.

You can specify the Unload command options directly in the **Unload Options** field. Enter the options in uppercase and use a semicolon to separate the options. For example:

```
DELIMITER = \036;ESCAPE = OFF;PARALLEL = ON;AWS_IAM_ROLE=arn;aws;iam;;<account ID>;role/
<role-name>
```

It is recommended to use octal representation of non-printable characters as DELIMITER.

If you run the Unload command as a pre-SQL or post-SQL command, specify the `ALLOWOVERWRITE` option to overwrite the existing objects.

By default, the UNLOAD property field is empty.

Unload Command Options

The Unload command options extract data from Amazon Redshift and load data to staging files on Amazon S3 in a particular format. You can delimit the data with a particular character or load data to multiple files in parallel.

To add options to the Unload command, use the **Unload Options** option.

You can set the following options:

DELIMITER

A single ASCII character to separate fields in the input file. You can use characters such as pipe (|), tilde (~), or a tab (\t). The delimiter you specify should not be a part of the data. If the delimiter is a part of data, use ESCAPE to read the delimiter character as a regular character. Default value is \036, the octal representation of the non-printable character, record separator.

ESCAPE

You can add an escape character for CHAR and VARCHAR columns in delimited unload files before the delimiter character is specified for the unloaded data. By default, the escape option is **ON**. To disable the escape option, specify **OFF** as the value of the escape option. For example, `ESCAPE = OFF`.

PARALLEL

The Unload command writes data in parallel to multiple files, according to the number of slices in the cluster. Default is on. If you turn the Parallel option off, the Unload command writes data serially. The maximum size of a data file is 6.5 GB.

AWS_IAM_ROLE

Specify the Amazon Redshift Role Resource Name (ARN) to run the mapping on Secure Agent installed on an Amazon EC2 system in the following format: `AWS_IAM_ROLE=arn:aws:iam::<account ID>:role/<role-name>`

For example: `arn:aws:iam::123123456789:role/redshift_read`

Source Partitioning

When you read data from Amazon Redshift, you can configure partitioning to optimize the mapping performance at run time. The partition type controls how the agent distributes data among partitions at partition points.

You can define the partition type as key range partitioning. Configure key range partitioning to partition Amazon Redshift data based on the value of a fields or set of fields. With key range partitioning, the Secure Agent distributes rows of source data based the fields that you define as partition keys. The Secure Agent compares the field value to the range values for each partition and sends rows to the appropriate partition.

Use key range partitioning for columns that have an even distribution of data values. Otherwise, the partitions might have unequal size. For example, a column might have 10 rows between key values 1 and 1000 and the column might have 999 rows between key values 1001 and 2000.

With key range partitioning, a query for one partition might return rows sooner than another partition. Or, one partition can return rows while the other partitions are not returning rows. This situation occurs when the rows in the table are in a similar order as the key range. One query might be reading and returning rows while the other queries are reading and filtering the same rows.

Note: You can configure a partition key only of the Integer and String data types.

When you configure more than two partitions in a mapping, the Secure Agent ignore the values that you specify in the start range for the first partition and end range for the last partition. The Secure Agent uses the start range value for the first partition as less than 10 and the end range value for the last partition as greater than the value you specify for the last partition.

For example, if you configure three partitions in a mapping and specify the start range value for the first partition as 5 and the end range value for the last partition as 90, the mapping runs successfully. However, the Secure Agent ignores the values that you specify and uses the start range value for the first partition as less than 10 and the end range value for the last partition as greater than 90.

Amazon Redshift V2 Targets

You can use an Amazon Redshift V2 object as a target in a mapping, mapping task, or mass ingestion task. You can also create an Amazon Redshift V2 target based on the input source.

When you configure the advanced target properties, you configure properties specific to Amazon Redshift V2. You can encrypt data, update statistical metadata of the database tables to improve the efficiency of queries, load data into Amazon Redshift from flat files in an Amazon S3 bucket, and use vacuum tables to recover disk space and sort rows in tables.
Note: If the distribution key column in a target table contains null values and you configure a task with an upsert operation for the same target table, the task might create duplicate rows. To avoid creating duplicate rows, you must perform one of the following tasks:

- Replace the null value with a non-null value when you load data.
- Do not configure the column as a distribution key if you expect null values in the distribution key column.
- Remove the distribution key column from the target table temporarily when you load data. You can use the Pre-SQL and Post-SQL properties to remove and then add the distribution key column in the target table.

Amazon Redshift Staging Directory for Amazon Redshift V2 Targets

The Secure Agent creates a staging file in the directory that you specify in the target properties. The Secure Agent writes the data to the staging directory before it writes data to Amazon Redshift.

The Secure Agent deletes the staged files from the staging directory after it writes the data to Amazon S3. Specify a staging directory in the mapping properties with an appropriate amount of disk space for the volume of data that you want to process. Specify a directory on the machine that hosts the Secure Agent.

The Secure Agent creates subdirectories in the staging directory. Subdirectories use the following naming convention: `<staging directory>/infaRedShiftStaging<MMddHHmmssSSS+xyz>`

Analyze Target Table

To optimize query performance, you can configure a task to analyze the target table. Target table analysis updates statistical metadata of the database tables.

You can use the Analyze Target Table option to extract sample rows from the table, analyze the samples, and save the column statistics. Amazon Redshift then updates the query planner with the statistical metadata. The query planner uses the statistical metadata to build and choose optimal plans to improve the efficiency of queries.

You can run the Analyze Target Table option after you load data to an existing table by using the Copy command. If you load data to a new table, the Copy command performs an analysis by default.

Data Encryption in Amazon Redshift V2 Targets

To protect data, you can enable server-side encryption or client-side encryption to encrypt the data that you insert in Amazon Redshift.

If you enable both server-side and client-side encryption for an Amazon Redshift target, then the client-side encryption is used for data load.

Server-side Encryption for Amazon Redshift V2 Targets

If you want Amazon Redshift to encrypt data while uploading and staging the `.csv` files to Amazon S3, you must enable server-side encryption.

To enable server-side encryption, select **S3 Server Side Encryption** in the advanced target properties and specify the **Customer Master key ID** in the connection properties.

You can configure the customer master key ID generated by AWS Key Management Service (AWS KMS) in the connection properties for server-side encryption. You must add IAM EC2 role and IAM Redshift role to the customer master key when you use IAM authentication and server-side encryption using customer master key. If you select the server-side encryption in the advanced target properties and do not specify the customer master key ID in the connection properties, Amazon S3-managed encryption keys are used to encrypt data.

Client-side Encryption for Amazon Redshift V2 Targets

Client-side encryption is a technique to encrypt data before transmitting the data to the Amazon Redshift server.

When you enable client-side encryption for Amazon Redshift V2 targets, the Secure Agent fetches the data from the source, writes the data to the staging directory, encrypts the data, and then writes the data to an Amazon S3 bucket. The Amazon S3 bucket then writes the data to Amazon Redshift.

If you enable both server-side and client-side encryption for an Amazon Redshift V2 target, then the client-side encryption is used for data load.

To enable client-side encryption, you must provide a master symmetric key in the connection properties. The Secure Agent encrypts the data by using the master symmetric key. The master symmetric key is a 256-bit AES encryption key in the Base64 format. Amazon Redshift V2 Connector uploads the data to the Amazon S3 server by using the master symmetric key and then loads the data to Amazon Redshift by using the copy command with the Encrypted option and a private encryption key for additional security. To enable client-side encryption, perform the following tasks:

1. Provide the master symmetric key when you create an Amazon Redshift V2 connection. Ensure that you provide a 256-bit AES encryption key in Base64 format.
2. Download the `local_policy.jar` and the `US_export_policy.jar` files for your JAVA environment from the Oracle website. Replace the existing `local_policy.jar` and the `US_export_policy.jar` files in the following directory: `<JAVA_HOME>\lib\security`.
3. Select **S3 Client Side Encryption** in the advanced target properties.

Retain Staging Files

You can retain staging files on Amazon S3 after the Secure Agent writes data to the target. You can retain files to create a data lake of your organizational data on Amazon S3. The files you retain can also serve as a backup of your data.

When you create a target connection, you can configure a file prefix or directory prefix to save the staging files. After you provide the prefixes, the Secure Agent creates files within the directories at Amazon S3 location specified in the target connection. Configure one of the following options for the **Prefix for Retaining Staging Files on S3** property:

- Provide a directory prefix and a file prefix. For example, `backup_dir/backup_file`. The Secure Agent creates the following directories and files:

  - `backup_dir_<year>_<month>_<date>_<timestamp_inLong>`

  - `backup_file.batch_<batch_number>.csv.<file_number>.<encryption_if_applicable>`

- Provide a file prefix. For example, backup_file. The Secure Agent creates the following directories and files:

  - `<year>_<month>_<date>_<timestamp_inLong><3 digit of random number>00<ProcessID><PartitionId>`

  - `backup_file.batch_<batch_number>.csv.<file_number>.<encryption_if_applicable>`

- Do not provide a prefix. The Secure Agent does not save the staging files.

Copy Command

You can use the Copy command to append data in a table. The Copy command uses a secure connection to load data from flat files in an Amazon S3 bucket to Amazon Redshift.

You can specify the Copy command options directly in the **Copy Options** field. Enter the options in uppercase and use a semicolon to separate the options. For example:

```
DELIMITER = \036;ACCEPTINVCHARS = #;QUOTE = \037;COMPUPDATE =
ON;AWS_IAM_ROLE=arn;aws;iam;;<account ID>;role/<role-name>
```

It is recommended to use octal representation of non-printable characters as DELIMITER and QUOTE.

Copy Command Options

The Copy command options read data from Amazon S3 and write data to Amazon Redshift in a particular format. You can apply compression to data in the tables or delimit the data with a particular character.

To add options to the Copy command, use the **CopyOptions Property File** option. You can set the following options:

DELIMITER

A single ASCII character to separate fields in the input file. You can use characters such as pipe (|), tilde (~), or a tab (\t). The delimiter must not be a part of the data. Default is \036, the octal representation of the non-printable character, record separator.

ACCEPTINVCHARS

Loads data into VARCHAR columns even if the data contains UTF-8 characters that are not valid. When you specify ACCEPTINCHARS, the Secure Agent replaces UTF-8 character that is not valid with an equal length string consisting of the character specified in ACCEPTINVCHARS. If you have specified '|' in ACCEPTINVCHARS, the Secure Agent replaces the three-byte UTF-8 character with '|||'.

If you do not specify ACCEPTINVCHARS, the COPY command returns an error when it encounters an UTF-8 character that is not valid. You can use the ACCEPTINVCHARS option on VARCHAR columns. Default is question mark (?).

QUOTE

Specifies the quote character to use with comma separated values. Default is \037, the octal representation of the non-printable character, unit separator.

COMPUPDATE

Overrides current compression encoding and applies compression to an empty table. Use the COMPUPDATE option in an insert operation when the rows in a table are more than 100,000. The behavior of COMPUPDATE depends on how it is configured:

- If you do not specify COMPUPDATE, the COPY command applies compression if the target table is empty and all columns in the table have either RAW or no encoding.
- If you specify COMPUPDATE ON, the COPY command replaces the existing encodings if the target table is empty and the columns in the table have encodings other than RAW.
- If you specify COMPUPDATE OFF, the COPY command does not apply compression.

Default is OFF.

TRUNCATECOLUMN

Truncates the data of the VARCHAR and CHAR data types column before writing the data to the target. If the size of the data that you want to write to the target is larger than size of the target column, the Secure Agent truncates the data before writing data to the target coulmn.

By default, the TRUNCATECOLUMNS option is OFF. To enable the TRUNCATECOLUMNS option, specify ON as the value of the TRUNCATECOLUMNS option. For example, TRUNCATECOLUMNS=ON.

AWS_IAM_ROLE

Specify the Amazon Redshift Role Resource Name (ARN) to run the task on Secure Agent installed on an Amazon EC2 system in the following format: `AWS_IAM_ROLE=arn:aws:iam::<account ID>:role/<role-name>`

For example: `arn:aws:iam::123123456789:role/redshift_write`

Vacuum Tables

You can use vacuum tables to recover disk space and sorts rows in a specified table or all tables in the database.

After you run bulk operations, such as delete or load, or after you run incremental updates, you must clean the database tables to recover disk space and to improve query performance on Amazon Redshift. Amazon Redshift does not reclaim and reuse free space when you delete and update rows.

Vacuum databases or tables often to maintain consistent query performance. You can recover disk space for the entire database or for individual tables in a database. You must run vacuum when you expect minimal activity on the database or during designated database administration schedules. Long durations of vacuum might impact database operations. Run vacuum often because large unsorted regions result in longer vacuum times.

You can enable the vacuum tables option when you configure the advanced target properties. You can select the following recovery options:

None

Does not sort rows or recover disk space.

Full

Sorts the specified table or all tables in the database and recovers disk space occupied by rows marked for deletion by previous update and delete operations.

Sort Only

Sorts the specified table or all tables in the database without recovering space freed by deleted rows.

Delete Only

Recovers disk space occupied by rows marked for deletion by previous update and delete operations, and compresses the table to free up used space.

Reindex

Analyzes the distribution of the values in the interleaved sort key columns to configure the entire Vacuum table operations for a better performance.

Octal Values as DELIMITER and QUOTE

In addition to printable ASCII characters, you can use octal values for printable and non-printable ASCII characters as DELIMITER and QUOTE.

To use a printable character as DELIMITER or QUOTE, you can either specify the ASCII character or the respective octal value. However, to use a non-printable character as DELIMITER or QUOTE, you must specify the respective octal value.

Example for a printable character:

`DELIMITER=#` or `DELIMITER=\043`

Example for a non-printable character, file separator:

`QUOTE=\034`

Octal values 000-037 and 177 represent non-printable characters and 040-176 represent printable characters. The following table lists the recommended octal values, for QUOTE and DELIMITER in the Copy command and as DELIMITER in the Unload command, supported by Amazon Redshift:

Command Option	Recommended Octal Values
COPY QUOTE	001-010, 016-037, 041-054, 057, 073-100,133, 135-140, 173-177
COPY DELIMITER	001-011, 013, 014, 016, 017, 020-046, 050-054, 057, 073-133, 135-177
UNLOAD DELIMITER	001-011, 013, 014, 016, 017, 020-041, 043-045, 050-054, 056-133, 135-177

Success and Error Files

The Secure Agent generates success and error files after you run a mapping. Success and error files are `.csv` files that contain row-level details.

The Secure Agent generates a success file after you run a mapping. The success file contains an entry for each record that successfully writes into Amazon Redshift. Each entry contains the values that are written for all the fields of the record. Use this file to understand the data that the Secure Agent writes to the Amazon S3 bucket and then to the Amazon Redshift target.

The error file contains an entry for each data error. Each entry in the file contains the values for all fields of the record and the error message. Use the error file to understand why the Secure Agent does not write data to the Amazon Redshift target.

The Secure Agent does not overwrite success or error files. Access the error rows files and success rows files directly from the directories where they are generated. You can manually delete the files that you no longer need.

Consider the following guidelines when you configure the mapping properties for success files:

- You must provide the file path where you want the Secure Agent to generate the success rows file.
- The success rows file uses the following naming convention: `<timestamp>success`

Consider the following guidelines when you configure the mapping properties for error files:

- You must provide the file path where you want the Secure Agent to generate the error rows file.
- The success rows file uses the following naming convention: `<timestamp>error`

  Note: The insert and upsert tasks error rows file follows the same naming convention.
- When you define a error file directory, you can use the variable `$PMBadFileDir`. When you use the `$PMBadFileDir` variable, the application writes the file to the following Secure Agent directory: `<Secure Agent installation directory>/apps/Data_Integration_Server/data/error`.

CHAPTER 4

Mass Ingestion Task with Amazon Redshift V2 Connector

This chapter includes the following topics:

Mass ingestion Task Overview

Use mass ingestion tasks to transfer a large number of files of any file type between on-premises and cloud repositories, and to track and monitor file transfers.

Create an Amazon Redshift V2 connection and use the connection to perform a mass ingestion task. When you create a mass ingestion task, select the target connection and specify which files you want to move from the source to the Amazon Redshift target.

Example

You work for an organization that stores purchase order details data, such as customer ID, item codes, and item quantity in an on-premise flat file system. You need to move the files that contains the purchase order details data from an on-premise flat file system to a cloud-based environment for data analysis.

You can create a mass ingestion task to move all the files that contains the purchase order details data from a flat file system to an Amazon Redshift target at once, instead of moving single row of data separately.

Before You Begin

Before you create mass ingestion tasks, verify that the following conditions exist:

- The organization has the following licenses:
  - Mass Ingestion
  - Mass Ingestion Runtime
- The Mass Ingestion application is running on the Secure Agent.
- Source and target connections exist, based on the sources from where you want to transfer files and the targets to where you want to transfer files.

Amazon Redshift V2 Targets in Mass ingestion Task

In a mass ingestion task, you can configure the Amazon Redshift V2 target properties to transfer files from any source that mass ingestion task supports to an Amazon Redshift target.

The following table describes the Amazon Redshift V2 target properties that you can configure in a mass ingestion task:

Target Property	Description
Connection Type	Type of the target connection. Select **Amazon Redshift V2** as the connection type.
Connection	Select the connection from a list of configured connections.

Amazon Redshift V2 Connector provides the following options that you must select to perform the copy command method:

- **Define Redshift Copy Command Properties**. Select this option to define the Amazon Redshift copy command properties.
- **Enter Custom Redshift Copy Command**. Select this option to provide a custom Amazon Redshift copy command that the mass ingestion task uses.

The following table describes the advanced target properties that you can configure in a mass ingestion task if you select the **Define Redshift Copy Command Properties** option:

Property	Description
Target Table Name	Name of the table in Amazon Redshift to which the files are loaded.
Schema	The Amazon Redshift schema name. Default is the schema that is used while creating connection.
Truncate Target Table	Truncate the target table before loading.

Property	Description
Analyze Target Table	The analyze command collects statistics about the contents of tables in the database to help determine the most efficient execution plans for queries.
Vacuum Target Table	You can select to vacuum the target table to recover disk space and sorts rows in a specified table in the database. You can select the following recovery options: - Full. Sorts the specified table and recovers disk space occupied by rows marked for deletion by previous update and delete operations. - Sort. Sorts the specified table without recovering space freed by deleted rows. - Delete. Recovers disk space occupied by rows marked for deletion by previous update and delete operations, and compresses the table to free up used space.
Copy Options	Select the format with which to copy data. The following options are available: - DELIMITER. A single ASCII character to separate fields in the input file. You can use characters such as pipe (|), tilde (~), or a tab (\t). The delimiter you specify cannot be a part of the data. - QUOTE. Specifies the quote character used to identify `nvarchar` characters and skip them. - COMPUPDATE. Overrides current compression encoding and applies compression to an empty table. - AWS_IAM_ROLE. Specify the Amazon Redshift Role Resource Name to run on an Amazon EC2 system. - IGNOREHEADER. Select to ignore headers. For example, if you specify `IGNOREHEADER 0`, the task processes data from row 0. - DATEFORMAT. Specify the format for date fields. - TIMEFORMAT. Specify the format for time fields.

The following table describes the advanced target properties that you can configure in a mass ingestion task if you select the **Enter Custom Redshift Copy Command** option:

Property	Description
Copy Command	Amazon Redshift COPY command appends the data to any existing rows in the table. If the Amazon S3 staging directory and the Amazon Redshift target belongs to different regions, you must specify the region in the COPY command. For example, `copy public.messages from '{{FROM-S3PATH}}' credentials 'aws_access_key_id={{ACCESS-KEY-ID}};aws_secret_access_key={{SECRET-ACCESS-KEY-ID}}' MAXERROR 0 REGION '' QUOTE '"' DELIMITER ',' NULL '' CSV;` Where `public` is the schema and `messages` is the table name. For more information about the COPY command, see the AWS documentation.

The following table describes the Amazon Redshift advanced target properties that you can configure in a mass ingestion task after you select one of the copy command methods:

Property	Description
Pre SQL	SQL command to run before the mass ingestion task runs the COPY command.
Post SQL	SQL command to run after the mass ingestion task runs the COPY command.

Property	Description
S3 Staging Directory	Specify the Amazon S3 staging directory. You must specify the Amazon S3 staging directory in `<bucket_name/folder_name>` format. The staging directory is deleted after the mass ingestion task runs.
Upload to Redshift with no Intermediate Staging	Upload files from Amazon S3 to Amazon Redshift directly from the Amazon S3 source directory with no additional, intermediate staging. If you select this option, ensure that the Amazon S3 bucket and the Amazon S3 staging directory belongs to the same region. If you do not select this option, ensure that the Amazon S3 staging directory and Amazon Redshift target belongs to the same region.
File Compression	Determines whether or not files are compressed before they are transferred to the target directory. The following options are available: - None. Files are not compressed. - GZIP. Files are compressed using GZIP compression.
File Encryption Type	Type of Amazon S3 file encryption to use during file transfer. The following options are available: - None. Files are not encrypted during transfer. - S3 server-side encryption. Amazon S3 encrypts the file using AWS-managed encryption keys. - S3 client-side encryption. Ensure that unrestricted policies are implemented for the AgentJVM, and that the master symmetric key for the connection is set. **Note:** Client-side encryption does not apply to tasks where Amazon S3 is the source.
S3 Accelerated Transfer	Select whether to use Amazon S3 Transfer Acceleration on the S3 bucket. To use Transfer Acceleration, accelerated transfer must be enabled for the bucket. The following options are available: - Disabled. Do not use Amazon S3 Transfer Acceleration. - Accelerated. Use Amazon S3 Transfer Acceleration. - Dualstack Accelerated. Use Amazon S3 Transfer Acceleration on a dual-stack endpoint.
Minimum Upload Part Size	Minimum upload part size when uploading a large file as a set of multiple independent parts, in megabytes. Use this option to tune the file load to Amazon S3.
Multipart Upload Threshold	Multipart download minimum threshold to determine when to upload objects in multipleparts in parallel.

Custom Amazon Redshift Copy Command

If you select to use an Amazon Redshift target connection, you can create a custom copy command that the mass ingestion task triggers to load files to Amazon Redshift.

You must specify credentials and variables in the command in the following format:

```
<ID> = <variable>
```

You can use the following credential IDs and variables for the custom copy command:

ID	Variable and Description
aws_access_key_id	Variable: {{ACCESS-KEY-ID}} Description: The AWS access key ID.
aws_secret_access_key	Variable: {{SECRET-ACCESS-KEY-ID}} Description: The AWS secret access key.
master_symmetric_key	Variable: {{MASTER-KEY}} Description: The master symmetric key.
from	Variable: {{FROM-S3PATH}} Description: The Amazon S3 folder.

You can also specify the format with which to copy data. The following options are available:

- DELIMITER. A single ASCII character to separate fields in the input file. You can use characters such as pipe (|), tilde (~), or a tab (\t). The delimiter you specify cannot be a part of the data.
- QUOTE. Specifies the quote character to use with comma separated values.
- COMPUPDATE. Overrides current compression encoding and applies compression to an empty table.
- AWS_IAM_ROLE. Specify the Amazon Redshift Role Resource Name to run on an Amazon EC2 system.
- IGNOREHEADER. Select to ignore headers. For example, if you specify `IGNOREHEADER 0`, the task processes data from row 0.
- DATEFORMAT. Specify the format for date fields.
- TIMEFORMAT. Specify the format for time fields.

For more information, see the Amazon Redshift copy command documentation at http://docs.aws.amazon.com/redshift/latest/dg/r_COPY.html.

The following code provides an example of a custom copy command:

```
copy "public"."test_str_tgt" ("col1" , "col2") from '{{FROM-S3PATH}}'
credentials 'aws_access_key_id={{ACCESS-KEY-ID}};aws_secret_access_key={{SECRET-ACCESS-
KEY-ID}}'
MAXERROR 0
QUOTE '"'
DELIMITER ','
DATEFORMAT AS 'YYYY-MM-DD HH24:MI:SS'
ROUNDEC
TIMEFORMAT AS 'YYYY-MM-DD HH24:MI:SS'
NULL ''
CSV
MANIFEST
```

Creating a Mass ingestion Task

You can create a mass ingestion task to transfer files from any source that mass ingestion task supports to an Amazon Redshift target.

1. In Data Integration, click **New** > **Tasks**.

2. Select **Mass Ingestion** and then click **Create**.

   The **Definition** tab appears.

3. In the **Definition** tab, configure the following properties:

Property	Description
Task Name	Name of the mass ingestion task. The names of mass ingestion tasks must be unique within the organization. Task names can contain alphanumeric characters, spaces, and underscores. Names must begin with an alphabetic character or underscore. Task names are not case sensitive.
Location	Project folder in which the task resides.
Description	Optional description of the task. Maximum length is 1024 characters.
Runtime Environment	Runtime environment that contains the Secure Agent used to run the task. The Mass Ingestion application must run on the Secure Agent.

4. Click **Next**.

   The **Source** tab appears.

5. On the **Source Details** page, select connection from a list of configured connections in the **Connection Type** field.

   You can select one of the following sources that mass ingestion task supports:

   - Local folder
   - Advanced FTP
   - Advanced FTPS
   - Advanced SFTP
   - Amazon S3

6. Click **View** to view the connection details.
7. Click **Test** to test the connection in the **View Connection** dialog.
8. Click **Next**.

   The **Target** tab appears.

9. On the **Target Details** section, select the **Connection Type** as **Amazon Redshift V2** and configure the Amazon Redshift V2 target properties.
10. Click **View** to view the connection details.
11. Click **Test** to test the connection in the **View Connection** dialog.
12. Click **Next**.

    The **Schedule** tab appears where you can select whether to run the task on a schedule or without a schedule.

13. Click **Run this task on schedule** to run a task on a schedule and select the schedule you want to use.

    If you want to remove a task from a schedule, click **Do not run this task on a schedule**.

14. Click **Finish** to save and close the task wizard.

You can edit, run, or view the mass ingestion task on the **Explore** page after you create the mass ingestion task.

Viewing Mass Ingestion Task Details

You can view details about a mass ingestion task, including the source and target connections and the associated schedule.

1. On the **Explore** page, navigate to the task.
2. In the row that contains the task, click **Actions** and select **View**.

   The **Task Details** page appears with task, source, target, and schedule details.
3. You can edit or run the task that you selected to view. On the **Task Details** page, click **Edit** to modify the task or click **Run** to run the task.

Running a Mass Ingestion Task

You can run a mass ingestion task in the following ways:

1. To run a mass ingestion task manually, on the **Explore** page, navigate to the task. In the row that contains the task, click **Actions** and select **Run**.

   Alternatively, you can run the task manually from the **Task Details** page. To access the **Task Details** page, click **Actions** and select **View**. In the **Task Details** page, select **Run**.
2. To run a mass ingestion task on a schedule, edit the task in the mass ingestion task wizard to associate the task with a schedule.

CHAPTER 5

Mappings and Mapping Tasks with Amazon Redshift V2 Connector

This chapter includes the following topics:

Amazon Redshift V2 Objects in Mappings

When you create a mapping, you can configure a Source or Target transformation to represent an Amazon Redshift V2 object.

Note: When you select an Amazon Redshift V2 object that contains a boolean data type and preview the data, the Secure Agent truncates the value of the boolean data type and displays only the first letter of the boolean value.

If you use a simple filter, specify the filter condition in `YYYY-MM-DD HH24:MI:SS.MS` format. If you use an advanced filter, specify the filter condition in `date_time_fix.f_timestamp < to_date('2012-05-24 09:13:57','YYYY-MM-DD HH24:MI:SS.MS')` format.

Amazon Redshift V2 Sources in Mappings

In a mapping, you can configure a Source transformation to represent an Amazon Redshift V2 source.

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

Property	Description
Connection	Name of the source connection.
Source type	Type of the source object. Select Single Object, Multiple Objects, Query, or Parameter.
Object	Name of the source object. You can select single or multiple source objects.

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

Property	Description
S3 Bucket Name	Amazon S3 bucket name for the Amazon Redshift source data. You can also specify the bucket name with the folder path. Use an S3 bucket in the same region as your Amazon Redshift cluster.
Enable Compression	Decompresses staging files before reading the files to Amazon Redshift. The task performance improves when the Secure Agent compresses the staging files. Default is selected.
Staging Directory Location	Amazon Redshift staging directory. When you run a task in Secure Agent runtime environment, specify a directory path that is available on the corresponding Secure Agent machine in the runtime environment. Specify the directory path in the following manner: `<staging directory>` For example, C:\Temp
Unload Options	Unload command options. Add options to the Unload command to write data from an Amazon Redshift object to an S3 bucket. Provide the Amazon Redshift Role Amazon Resource Name (ARN). You can add the following options: - DELIMITER - ESCAPE - PARALLEL - AWS_IAM_ROLE For example: `DELIMITER = \036;ESCAPE = OFF;PARALLEL = ON;AWS_IAM_ROLE=arn;aws;iam;;<account ID>;role/<role-name>` Specify a directory on the machine that hosts the Secure Agent. **Note:** If you do not add the options to the Unload command manually, the Secure Agent considers the default values.

Property	Description
Encryption Type	Select the source encryption type. You can select from the following encryption types: - None - SSE-S3 - SSE-KMS - CSE-SMK Default is None. For more information, see "Encryption Type" on page 14
Download S3 Files in Multiple Parts	Downloads large Amazon S3 objects in multiple parts. When the file size of an Amazon S3 object is greater than 8 MB, you can choose to download the object in multiple parts in parallel. Default is 5 MB.
Multipart Download Threshold Size	Specifies the part size of an object. Default is 5 MB.
Pre-SQL	The pre-SQL commands to run a query before you read data from Amazon Redshift. You can also use the UNLOAD or COPY command. The command you specify here is processed as a plain text.
Post-SQL	The post-SQL commands to run a query after you write data to Amazon Redshift. You can also use the UNLOAD or COPY command. The command you specify here is processed as a plain text.
Select Distinct	Selects unique values. The agent includes a SELECT DISTINCT statement if you choose this option. Amazon Redshift ignores trailing spaces. Therefore, the agent might extract fewer rows than expected. **Note:** If you select the source type as query or use the **SQL Query** property and select the **Select Distinct** option, the Secure Agent ignores the **Select Distinct** option.
Tracing Level	Use the verbose tracing level to get the amount of detail that appears in the log for the Source transformation.

Configuring Key Range Partition

Configure key range partition to partition Amazon Redshift data based on field values.

1. In **Source Properties**, click the **Partitions** tab.
2. Select the required **Partition Key** from the list.
3. Click **Add New key Range** to add partitions.
4. Specify the **Start range** and **End range**.

Amazon Redshift V2 Targets in Mappings

To write data to Amazon Redshift, configure an Amazon Redshift V2 object as the target in a mapping.

When you enable the source partition, the Secure Agent uses the pass-through partitioning to write data to Amazon Redshift to optimize the mapping performance at run time. Specify the name and description of the Amazon Redshift V2 target. Configure the target and advanced properties for the target object

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

Property	Description
Connection	Name of the target connection.
Target Type	Type of the target object. Select Single Object or Parameter.
Object	Name of the target object. Target object for a single target.
Operation	Target operation. Select Insert, Update, Upsert, or Delete.
Create Target	Creates a target. Enter a name for the target object. You can provide the schema name and create a target table within the schema. By default, the field is empty. **Note:** The Secure Agent converts the table names that you specify in the **Create Target** field into lower case.

The following table describes the Amazon Redshift V2 advanced target properties:

Property	Description
S3 Bucket Name	Amazon S3 bucket name for the Amazon Redshift target data. You can also specify the bucket name with the folder path. Use an S3 bucket in the same region as your Amazon Redshift cluster.
Enable Compression	Compresses staged files before writing the files to Amazon Redshift. Mapping performance improves when the Secure Agent compresses the staged files. Default is selected.
Staging Directory Location	Amazon Redshift staging directory. When you run a task in Secure Agent runtime environment, specify a directory path that is available on the corresponding Secure Agent machine in the runtime environment. Specify the directory path in the following manner: `<staging directory>` For example, C:\Temp
Batch Size	Minimum number of rows in a batch. Enter a number greater than 0. Default is 2000000.
Max Errors per Upload Batch for INSERT	Number of errors within a batch that causes a batch to fail. Enter a positive integer. If the number of errors is equal to or greater than the property value, the Secure Agent writes the entire batch to the error file. Default is 1.
Truncate Target Table Before Data Load	Truncates an Amazon Redshift target before writing data to the target.
Require Null Value For Char and Varchar	A string value that you want to replace as NULL when data is uploaded to Amazon Redshift. Default is an empty string.
WaitTime In Seconds For S3 File Consistency	Number of seconds to wait for the Secure Agent to make the staged files consistent with the list of files available on Amazon S3. Default is 0.

Property	Description
Copy Options	Name of the property file. Add additional options to the copy command for writing data from an Amazon S3 source to an Amazon Redshift target when the default delimiter comma (,) or double-quote (") is used in the data. Provide the Amazon Redshift Role Amazon Resource Name (ARN). You can add the following options: - DELIMITER - ACCEPTINVCHARS - QUOTE - COMPUPDATE - AWS_IAM_ROLE For example: `DELIMITER = \036;ACCEPTINVCHARS = #;QUOTE = \037 COMPUPDATE = ON;AWS_IAM_ROLE=arn:aws:iam::<account ID>:role/<role-name>` Specify a directory on the machine that hosts the Secure Agent. **Note:** If you do not add the options to the Copy command manually, the Secure Agent considers the default values.
S3 Server Side Encryption	Indicates that Amazon S3 encrypts data during upload and decrypts data at the time of access. You must provide a **Customer Master key ID** in the connection property to enable this property. Default is not selected.
S3 Client Side Encryption	Indicates that the Secure Agent encrypts data by using a private encryption key. If you enable both server side and client side encryption, the Secure Agent ignores the server side encryption. You must provide a **Master Symmetric Key ID** in the connection property to enable this property.
Analyze Target Table	Improve the efficiency of the write operations. The query planner on Amazon Redshift updates the statistical metadata to build and choose optimal plans to improve the efficiency of queries.
Vacuum Target Table	Recovers disk space and sorts row in a specified table or all tables in the database. You can select the following recovery options: - None - Full - Sort Only - Delete Only - Reindex Default is None. For more information about the vacuum tables, see "Vacuum Tables" on page 20.
Prefix to retain staging files on S3	Retains staging files on Amazon S3. Provide both a directory prefix and a file prefix separated by a slash (/) or only a file prefix to retain staging files on Amazon S3. For example, `backup_dir/backup_file` or `backup_file`.
Success File Directory	Directory for the Amazon Redshift success file. Specify a directory on the machine that hosts the Secure Agent.
Error File Directory	Directory for the Amazon Redshift error file. Specify a directory on the machine that hosts the Secure Agent.

Property	Description
Treat Source Rows As	Overrides the Amazon Redshift target. Default is **INSERT**. Select one of the following override options: **NONE** By default, none is enabled. The Secure Agent considers the task operation that you select in the **Operation** target property. **INSERT** Performs insert operation. If enabled, the Secure Agent inserts all rows flagged for insert. If disabled, the Secure Agent rejects the rows flagged for insert. **DELETE** Performs delete operation. If enabled, the Secure Agent deletes all rows flagged for delete. If disabled, the Secure Agent rejects all rows flagged for delete. **UPDATE and UPSERT** Performs update and upsert operations. To perform an update operation, you must map the primary key column and at least one column other than primary key column. You can select the following data object operation attributes: - Update as Update: The Secure Agent updates all rows as updates. - Update else Insert: The Secure Agent updates existing rows and inserts other rows as if marked for insert. **Note:** Amazon Redshift V2 Connector does not support the Upsert operation in the Upgrade Strategy transformation. To use an Update Strategy transformation to write data to an Amazon Redshift target, you must select **Treat Source Rows As** as **None**. By default, the Secure Agent performs the task operation based on the value that you specify in the **Operation** target property. However, if you specify an option in the **Treat Source Rows As** property, the Secure Agent ignores the value of that you specify in the **Operation** target property or in the Update Strategy transformation.
TransferManager Thread Pool Size	Specifies the number of the threads to write data in parallel. Default is 10.
Number of files per batch	Calculates the number of the staging files per batch. If you do not provide the number of files, Amazon Redshift V2 Connector calculates the number of the staging files.
Target table name	Enter the target table name to override the default target table name.
Pre-SQL	The pre-SQL commands to run a query before you read data from Amazon Redshift. You can also use the UNLOAD or COPY command. The command you specify here is processed as a plain text.
Post-SQL	The post-SQL commands to run a query after you write data to Amazon Redshift. You can also use the UNLOAD or COPY command. The command you specify here is processed as a plain text.
Minimum Upload Part Size	Specifies the part size of an object. Default is 5 MB.
Forward Rejected Rows	This property is not applicable for Amazon S3 V2 Connector.

Amazon Redshift V2 Objects in Mapping Tasks

When you configure a mapping task, you can configure advanced properties for Amazon Redshift V2 targets.

Amazon Redshift V2 Sources in Mapping Tasks

For Amazon Redshift V2 source connections used in template-based mapping tasks, you can configure advanced properties in the Sources page.

You can configure the following advanced properties:

Property	Description
S3 Bucket Name	Amazon S3 bucket name for the Amazon Redshift source data. You can also specify the bucket name with the folder path. Use an S3 bucket in the same region as your Amazon Redshift cluster.
Enable Compression	Decompresses staging files before reading the files to Amazon Redshift. The task performance improves when the Secure Agent compresses the staging files. Default is selected.
Staging Directory Location	Amazon Redshift staging directory. When you run a task in Secure Agent runtime environment, specify a directory path that is available on the corresponding Secure Agent machine in the runtime environment. Specify the directory path in the following manner: `<staging directory>` For example, C:\Temp
Unload Options	Unload command options. Add options to the Unload command to write data from an Amazon Redshift object to an S3 bucket. Provide the Amazon Redshift Role Amazon Resource Name (ARN). You can add the following options: - DELIMITER - ESCAPE - PARALLEL - AWS_IAM_ROLE For example: `DELIMITER = \036;ESCAPE = OFF;PARALLEL = ON;AWS_IAM_ROLE=arn;aws;iam;;<account ID>;role/<role-name>` Specify a directory on the machine that hosts the Secure Agent. **Note:** If you do not add the options to the Unload command manually, the Secure Agent considers the default values.
Encryption Type	Select the source encryption type. You can select from the following encryption types: - None - SSE-S3 - SSE-KMS - CSE-SMK Default is None. For more information, see "Encryption Type" on page 14
Download S3 Files in Multiple Parts	Downloads large Amazon S3 objects in multiple parts. When the file size of an Amazon S3 object is greater than 8 MB, you can choose to download the object in multiple parts in parallel. Default is 5 MB.

Property	Description
Multipart Download Threshold Size	Specifies the part size of an object. Default is 5 MB.
Pre-SQL	The pre-SQL commands to run a query before you read data from Amazon Redshift. You can also use the UNLOAD or COPY command. The command you specify here is processed as a plain text.
Post-SQL	The post-SQL commands to run a query after you write data to Amazon Redshift. You can also use the UNLOAD or COPY command. The command you specify here is processed as a plain text.
SQL Query	Define an SQL query for source objects in a mapping. Enclose column names in double quotes. The SQL query is case sensitive. Specify an SQL statement supported by the Amazon Redshift database.
Select Distinct	Selects unique values. The agent includes a SELECT DISTINCT statement if you choose this option. Amazon Redshift ignores trailing spaces. Therefore, the agent might extract fewer rows than expected. **Note:** If you select the source type as query or use the **SQL Query** property and select the **Select Distinct** option, the Secure Agent ignores the **Select Distinct** option.
Tracing Level	Use the verbose tracing level to get the amount of detail that appears in the log for the Source transformation.

Amazon Redshift V2 Targets in Mapping Tasks

For Amazon Redshift V2 target connections used in mapping tasks, you can configure advanced target properties in the **Targets** page of the Mapping Task wizard.

You can configure the following advanced target properties:

Property	Description
S3 Bucket Name	Amazon S3 bucket name for the Amazon Redshift target data. You can also specify the bucket name with the folder path. Use an S3 bucket in the same region as your Amazon Redshift cluster.
Enable Compression	Compresses staged files before writing the files to Amazon Redshift. Mapping performance improves when the Secure Agent compresses the staged files. Default is selected.
Staging Directory Location	Amazon Redshift staging directory. When you run a task in Secure Agent runtime environment, specify a directory path that is available on the corresponding Secure Agent machine in the runtime environment. Specify the directory path in the following manner: `<staging directory>` For example, C:\Temp
Batch Size	Minimum number of rows in a batch. Enter a number greater than 0. Default is 2000000.

Property	Description
Max Errors per Upload Batch for INSERT	Number of errors within a batch that causes a batch to fail. Enter a positive integer. If the number of errors is equal to or greater than the property value, the Secure Agent writes the entire batch to the error file. Default is 1.
Truncate Target Table Before Data Load	Truncates an Amazon Redshift target before writing data to the target.
Require Null Value For Char and Varchar	A string value that you want to replace as NULL when data is uploaded to Amazon Redshift. Default is an empty string.
WaitTime In Seconds For S3 File Consistency	Number of seconds to wait for the Secure Agent to make the staged files consistent with the list of files available on Amazon S3. Default is 0.
Copy Options	Name of the property file. Add additional options to the copy command for writing data from an Amazon S3 source to an Amazon Redshift target when the default delimiter comma (,) or double-quote (") is used in the data. Provide the Amazon Redshift Role Amazon Resource Name (ARN). You can add the following options: - DELIMITER - ACCEPTINVCHARS - QUOTE - COMPUPDATE - AWS_IAM_ROLE For example: `DELIMITER = \036;ACCEPTINVCHARS = #;QUOTE = \037 COMPUPDATE = ON;AWS_IAM_ROLE=arn:aws:iam::<account ID>:role/<role-name>` Specify a directory on the machine that hosts the Secure Agent. **Note:** If you do not add the options to the Copy command manually, the Secure Agent considers the default values.
S3 Server Side Encryption	Indicates that Amazon S3 encrypts data during upload and decrypts data at the time of access. You must provide a **Customer Master key ID** in the connection property to enable this property. Default is not selected.
S3 Client Side Encryption	Indicates that the Secure Agent encrypts data by using a private encryption key. If you enable both server side and client side encryption, the Secure Agent ignores the server side encryption. You must provide a **Master Symmetric Key ID** in the connection property to enable this property.
Analyze Target Table	Improve the efficiency of the write operations. The query planner on Amazon Redshift updates the statistical metadata to build and choose optimal plans to improve the efficiency of queries.

Property	Description
Vacuum Target Table	Recovers disk space and sorts row in a specified table or all tables in the database. You can select the following recovery options: - None - Full - Sort Only - Delete Only - Reindex Default is None. For more information about the vacuum tables, see "Vacuum Tables" on page 20.
Prefix to retain staging files on S3	Retains staging files on Amazon S3. Provide both a directory prefix and a file prefix separated by a slash (/) or only a file prefix to retain staging files on Amazon S3. For example, `backup_dir/backup_file` or `backup_file`.
Success File Directory	Directory for the Amazon Redshift success file. Specify a directory on the machine that hosts the Secure Agent.
Error File Directory	Directory for the Amazon Redshift error file. Specify a directory on the machine that hosts the Secure Agent.
Treat Source Rows As	Overrides the Amazon Redshift target. Default is **INSERT**. Select one of the following override options: **NONE** By default, none is enabled. The Secure Agent considers the task operation that you select in the **Operation** target property. **INSERT** Performs insert operation. If enabled, the Secure Agent inserts all rows flagged for insert. If disabled, the Secure Agent rejects the rows flagged for insert. **DELETE** Performs delete operation. If enabled, the Secure Agent deletes all rows flagged for delete. If disabled, the Secure Agent rejects all rows flagged for delete. **UPDATE and UPSERT** Performs update and upsert operations. To perform an update operation, you must map the primary key column and at least one column other than primary key column. You can select the following data object operation attributes: - Update as Update: The Secure Agent updates all rows as updates. - Update else Insert: The Secure Agent updates existing rows and inserts other rows as if marked for insert. **Note:** Amazon Redshift V2 Connector does not support the Upsert operation in the Upgrade Strategy transformation. To use an Update Strategy transformation to write data to an Amazon Redshift target, you must select **Treat Source Rows As** as **None**. By default, the Secure Agent performs the task operation based on the value that you specify in the **Operation** target property. However, if you specify an option in the **Treat Source Rows As** property, the Secure Agent ignores the value of that you specify in the **Operation** target property or in the Update Strategy transformation.

Property	Description
TransferManager Thread Pool Size	Specifies the number of the threads to write data in parallel. Default is 10.
Number of files per batch	Calculates the number of the staging files per batch. If you do not provide the number of files, Amazon Redshift V2 Connector calculates the number of the staging files.
Target table name	Enter the target table name to override the default target table name.
Pre-SQL	The pre-SQL commands to run a query before you read data from Amazon Redshift. You can also use the UNLOAD or COPY command. The command you specify here is processed as a plain text.
Post-SQL	The post-SQL commands to run a query after you write data to Amazon Redshift. You can also use the UNLOAD or COPY command. The command you specify here is processed as a plain text.
Minimum Upload Part Size	Specifies the part size of an object. Default is 5 MB.
Forward Rejected Rows	This property is not applicable for Amazon S3 V2 Connector.

Amazon Redshift Lookups in Mapping Tasks

When you configure field mappings in a mapping task, you can create a cached and uncached lookup to an Amazon Redshift object. Use the JDBC URL specified in the connection properties to create a cached and uncached lookup.

For more information about the cached and uncached lookup, see "Lookup Transformation" in the *PowerCenter Transformation Guide*.

CHAPTER 6

Data Type Reference

This chapter includes the following topics:

Data Type Reference Overview

Data Integration uses the following data types in mappings, mapping tasks, and mass ingestion tasks with Amazon Redshift:

Amazon Redshift native data types

Amazon Redshift data types appear in 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.

Amazon Redshift and Transformation Data Types

The following table lists the Amazon Redshift data types that Data Integration supports and the corresponding transformation data types:

Amazon Redshift Data Type	Transformation Data Type	Description
Bigint	Bigint	Signed eight-byte integer.
Boolean	Small Integer	Logical Boolean (true/false).
Char	String	Fixed-length character string.

Amazon Redshift Data Type	Transformation Data Type	Description
Date	Timestamp	Calendar date (year, month, day).
Decimal	Decimal	Exact numeric of selectable precision.
Double Precision	Double	Double precision floating-point number.
Integer	Integer	Signed four-byte integer.
Real	Double	Single precision floating-point number.
Smallint	Small Integer	Signed two-byte integer.
Timestamp	Timestamp	Date and time (without time zone).
Varchar	String	Variable-length character string with a user-defined limit.

CHAPTER 7

Troubleshooting

This chapter includes the following topics:

Troubleshooting Overview

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

Troubleshooting for Amazon Redshift V2 Connector

How to configure AWS IAM authentication for Amazon Redshift V2 Connector?

For information about configuring AWS IAM authentication, see https://kb.informatica.com/h2l/HowTo%20Library/1/0972-ConfiguringAWSIAMforAmazonRedshiftandAmazonRedshiftV2Connectors-H2L.pdf

INDEX