Informatica®

Informatica® Cloud Data Integration

Amazon Redshift Connectors

Informatica Cloud Data Integration Amazon Redshift Connectors
July 2024

© Copyright Informatica LLC 2019, 2024

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

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

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

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

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

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

NOTICES

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

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

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

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

Publication Date: 2024-07-18

Table of Contents

Preface 8

Informatica Resources 8

- Informatica Documentation 8
- Informatica Intelligent Cloud Services web site 8
- Informatica Intelligent Cloud Services Communities 8
- Informatica Intelligent Cloud Services Marketplace 9
- Data Integration connector documentation 9
- Informatica Knowledge Base 9
- Informatica Intelligent Cloud Services Trust Center 9
- Informatica Global Customer Support 9

Part I: Introduction to Amazon Redshift connectors 10

Chapter 1: Introduction to Amazon Redshift connectors 11

Secure Agent and Amazon Redshift integration 12

Chapter 2: Connector comparison 13

Mapping functionality 13

Source functionality 14

Target functionality 15

Chapter 3: Task Examples 16

Synchronization task use case 16

Mapping and mapping task use case 16

Mapping task with Oracle CDC sources use case 16

Use case for a mapping in advanced mode 17

Part II: Data Integration with Amazon Redshift V2 Connector 18

Chapter 4: Introduction to Amazon Redshift V2 Connector 19

Amazon Redshift V2 Connector assets 19

Introduction to Amazon Redshift 20

Amazon Redshift Spectrum overview 20

Chapter 5: Connections for Amazon Redshift V2 21

Prepare for authentication 21

Create a minimal Amazon IAM policy 23

Configure IAM authentication 23

Configure an assume role for Amazon Redshift 24

- Generate temporary security credential policies for Amazon Redshift 24

Generate temporary security credentials using AssumeRole for EC2. . . . 26
Configure an assume role for Amazon S3 staging. . . . 27
Generate temporary security credentials using AssumeRole for Amazon S3 staging. . . . 28
Generate temporary security credentials using AssumeRole for EC2. . . . 30
Enable encryption. . . . 31
Connect to Amazon Redshift. . . . 32
Before you begin. . . . 32
Connection details. . . . 32
Authentication types. . . . 32
Proxy server settings. . . . 42
Configure SSL. . . . 42
Configure SSL with the serverless runtime environment. . . . 43
Configure client-side encryption with the serverless runtime environment. . . . 44
Configure SSE-KMS encryption for mappings in advanced mode. . . . 45
Amazon Redshift Serverless connectivity. . . . 45
Requirements to use Amazon Redshift Spectrum. . . . 45
Private communication with Amazon Redshift. . . . 46
Private communication with Amazon S3. . . . 46
VPC peering between the serverless runtime environment and Amazon Redshift. . . . 47

Chapter 6: Amazon Redshift sources and targets. . . . 48

Amazon Redshift V2 sources. . . . 48
Read from Amazon Redshift with or without staging. . . . 48
Data encryption in Amazon Redshift V2 sources. . . . 49
Unload command. . . . 49
Source partitioning. . . . 51
Amazon Redshift V2 targets. . . . 52
Amazon Redshift staging directory for Amazon Redshift V2 targets. . . . 52
Data encryption in Amazon Redshift V2 targets. . . . 52
Copy command. . . . 53
Analyze target table. . . . 55
Retain staging files. . . . 55
Vacuum tables. . . . 55
Recovery and restart processing. . . . 56
Preserve record order on Write. . . . 57
Octal values as DELIMITER and QUOTE. . . . 57
Success and error files. . . . 58

Chapter 7: Mappings and mapping tasks with Amazon Redshift. . . . 59

Before you begin. . . . 59
Create an external schema and table for Amazon Redshift Spectrum. . . . 60
Amazon Redshift V2 objects in mappings. . . . 61
Amazon Redshift V2 sources in mappings. . . . 62

Rules and guidelines for reading data. . . . 65
Rules and guidelines for configuring SQL query. . . . 65
Adding multiple source objects. . . . 66
Amazon Redshift V2 targets in mappings. . . . 69
Amazon Redshift V2 lookups in mappings. . . . 75
Mapping task with Oracle CDC sources example. . . . 82
Mapping in advanced mode example. . . . 83
SQL transformation. . . . 84
Dynamic schema handling for mappings. . . . 85
Configuring key range partition. . . . 86
Bulk processing for write operations. . . . 87
Optimize the staging performance for a mapping. . . . 87

Chapter 8: Migrating a mapping. . . . 89

Use the same object path for the migrated mapping. . . . 89
Use a different object path for the migrated mapping. . . . 89
Migration options. . . . 90
Rules and guidelines for migrating a mapping. . . . 91

Chapter 9: Upgrade the connection type. . . . 92

Connection switching example. . . . 93
Advanced properties retained after the switch. . . . 95
Rules and guidelines. . . . 96

Chapter 10: SQL ELT with Amazon Redshift V2 Connector. . . . 97

SQL ELT configuration options. . . . 97
Create a mapping in SQL ELT mode . . . 97
Enable SQL ELT optimization for a mapping task . . . 99
SQL ELT query preview. . . . 99
Mappings in SQL ELT mode for Amazon Redshift. . . . 99
Sources in mappings in SQL ELT mode. . . . 100
Targets in mappings in SQL ELT mode. . . . 100
Transformations in mappings in SQL ELT mode. . . . 101
Functions in mappings in SQL ELT mode. . . . 101
Operators in mappings in SQL ELT mode. . . . 102
Rules and guidelines in mappings in SQL ELT mode. . . . 102
SQL ELT optimization for mapping tasks . . . 102
SQL ELT optimization types. . . . 103
SQL ELT optimization scenarios. . . . 103
SQL ELT compatibility. . . . 104
Configuring SQL ELT optimization. . . . 107
Transformations for Amazon Redshift V2 mappings. . . . 107
Features. . . . 110

SQL ELT optimization for multiple targets. . . . 113
Determine the order of processing for multiple targets. . . . 113
Clean stop a SQL ELT optimization job. . . . 114
Rules and guidelines for SQL ELT optimization. . . . 115

Chapter 11: Data type reference. . . . 123
Amazon Redshift and transformation data types. . . . 123
Rules and guidelines for data types. . . . 125

Chapter 12: Troubleshooting. . . . 128
Troubleshooting for Amazon Redshift V2 Connector. . . . 128

Part III: Data Integration with Amazon Redshift Connector. . . . 130

Chapter 13: Introduction to Amazon Redshift Connector. . . . 131
Amazon Redshift Connector assets. . . . 131
Introduction to Amazon Redshift. . . . 132
Amazon Redshift Connector example. . . . 132
Administration of Amazon Redshift Connector. . . . 132
Configure Amazon Redshift Connector for SSL. . . . 133
Create a minimal Amazon IAM policy. . . . 133
IAM authentication. . . . 134

Chapter 14: Amazon Redshift connections. . . . 136
Amazon Redshift connection properties. . . . 136
Configuring proxy settings. . . . 137
Configuring proxy settings on Windows. . . . 137
Configuring proxy settings on Linux. . . . 138

Chapter 15: Amazon Redshift sources and targets. . . . 139
Amazon Redshift sources. . . . 139
Amazon Redshift staging directory for Amazon Redshift sources. . . . 139
Server-side encryption for Amazon Redshift sources. . . . 140
Client-side encryption for Amazon Redshift sources. . . . 141
Unload command. . . . 141
Partitioning. . . . 142
Amazon Redshift targets. . . . 143
Amazon Redshift staging directory for Amazon Redshift targets. . . . 143
Analyze target table. . . . 143
Data encryption in Amazon Redshift targets. . . . 144
Retain staging files. . . . 144
Copy command. . . . 145
Field mappings. . . . 146

Vacuum tables. . . . 146
Working with large tables. . . . 147
Octal values as DELIMITER and QUOTE. . . . 147
Success and error files. . . . 148

Chapter 16: Synchronization tasks with Amazon Redshift. . . . 150

Amazon Redshift sources in synchronization tasks. . . . 150
Amazon Redshift targets in synchronization tasks. . . . 152
Amazon Redshift lookups in synchronization tasks. . . . 155
Rules and guidelines for synchronization tasks. . . . 155
Synchronization task example. . . . 155

Chapter 17: Mappings and mapping tasks with Amazon Redshift. . . . 157

Amazon Redshift sources in mappings. . . . 157
Configuring key range partitioning. . . . 159
Amazon Redshift targets in mappings. . . . 160
Amazon Redshift lookups in mappings. . . . 163
Amazon Redshift objects in mapping tasks. . . . 163
Amazon Redshift sources in mapping tasks. . . . 163
Amazon Redshift targets in mapping tasks. . . . 164

Chapter 18: Amazon Redshift SQL ELT optimization. . . . 168

Supported functions and operators for Amazon Redshift mappings. . . . 168
Configuring Amazon Redshift ODBC connection. . . . 171
Configuring Amazon Redshift ODBC connection on Windows. . . . 171
Configuring Amazon Redshift ODBC connection on Linux. . . . 173
Creating an ODBC connection. . . . 174
Cross-schema SQL ELT optimization. . . . 175
Configuring cross-schema SQL ELT optimization for an Amazon Redshift mapping task. . . . 176
Rules and guidelines for functions in SQL ELT optimization. . . . 176

Chapter 19: Data type reference. . . . 178

Amazon Redshift and transformation data types. . . . 178

Chapter 20: Troubleshooting. . . . 180

Troubleshooting for Amazon Redshift Connector. . . . 180
Troubleshooting Amazon Redshift connection. . . . 180

Index. . . . 181

Preface

Use *Amazon Redshift Connectors* to learn about *Amazon Redshift* and *Amazon Redshift V2 Connectors*. Use Part I to know the overview and the functionality comparison between Amazon Redshift V2 and Amazon Redshift connectors. Use Part II and Part III to learn about the functionality available for Amazon Redshift Connectors.

Informatica Resources

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

Informatica Documentation

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

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

Informatica Intelligent Cloud Services web site

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

Informatica Intelligent Cloud Services Communities

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

Access the Informatica Intelligent Cloud Services Community at:

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

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

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

Informatica Intelligent Cloud Services Marketplace

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

https://marketplace.informatica.com/

Data Integration connector documentation

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

Informatica Knowledge Base

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

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

Informatica Intelligent Cloud Services Trust Center

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

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

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

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

Informatica Global Customer Support

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

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

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

Part I: Introduction to Amazon Redshift connectors

This part contains the following chapters:

- Introduction to Amazon Redshift connectors, 11
- Connector comparison, 13
- Task Examples, 16

CHAPTER 1

Introduction to Amazon Redshift connectors

You can use Amazon Redshift connectors to read data from and write data to Amazon Redshift. Use the connectors to create sources and targets that represent records in Amazon Redshift.

When you use Amazon Redshift connectors to create and run a Data Integration task, the Secure Agent reads from and writes data to Amazon Redshift based on the taskflow and Amazon Redshift connection configuration. The Secure Agent connects reads data from and writes data to Amazon Simple Storage Service (Amazon S3) through a TCP/IP network. The Secure Agent uses the Amazon driver to communicate with Amazon Redshift.

You can move data from any data source to Amazon Redshift.

Use the following connectors to create connections and integrate data to and from Amazon Redshift:

Amazon Redshift V2 Connector

This is the recommended connector to connect to Amazon Redshift. Use Amazon Redshift V2 Connector to create a mapping or a mapping task.

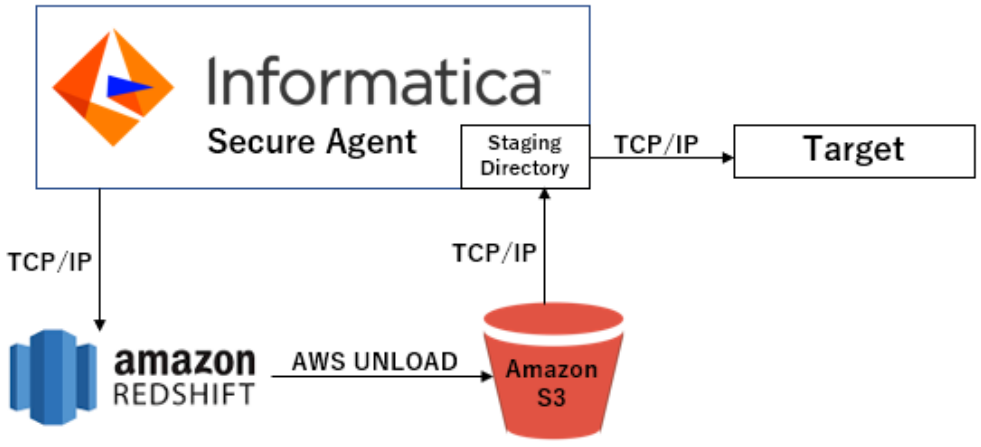
Amazon Redshift Connector

This is an older version of Amazon Redshift Connector. Informatica recommends that you use Amazon Redshift Connector to read or write data only when you want to use a synchronization task.

Secure Agent and Amazon Redshift integration

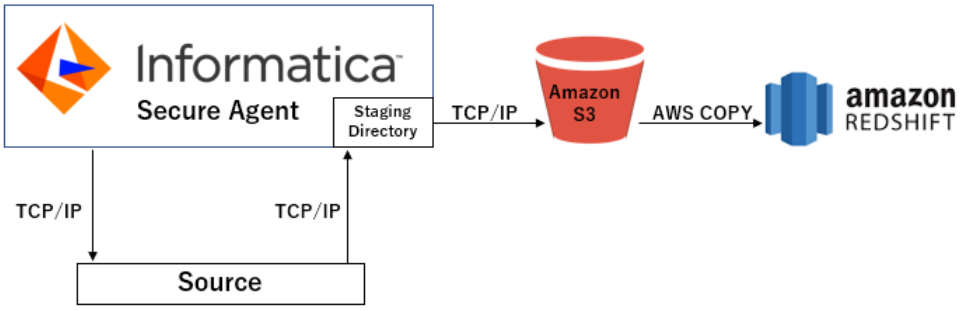
The Secure Agent uses the Amazon Redshift connection to connect to Amazon Redshift.

The following image shows how the Secure Agent connects to Amazon Redshift to read data:



The Secure Agent connects to Amazon Redshift and issues UNLOAD command to read data from Amazon Redshift to the Amazon Simple Storage Service (Amazon S3) bucket specified in the connection properties. The Secure Agent then stores data in a staging directory on the Secure Agent system over a TCP/IP network.

The following image shows how the Secure Agent connects to Amazon Redshift to write data:



The Secure Agent reads data from a staging directory on the Secure Agent system and writes data to Amazon Simple Storage Service (Amazon S3) through a TCP/IP network. The Secure Agent then issues COPY command to write data from Amazon S3 to the Amazon Redshift target table.

CHAPTER 2

Connector comparison

Based on your requirements to integrate or ingest data, you can use either Amazon Redshift Connector or Amazon Redshift V2 Connector to create data integration tasks.

The functionality to create integration tasks and configure read and write operations differ in both connectors. Informatica recommends to use Amazon Redshift V2 connector as the new features and enhancements are provided for Amazon Redshift V2 Connector.

Mapping functionality

The following table compares the mapping functionality supported by Amazon Redshift connectors:

Mapping functionality	Amazon Redshift Connector	Amazon Redshift V2 Connector
Hosted agent	Yes	Yes
Proxy server	Yes	Yes
Synchronization task	Yes	No
Mapping task	Yes	Yes (Preferred)
Mapping in advanced mode	No	Yes
Read from Oracle CDC Sources	No	Yes

Mapping functionality	Amazon Redshift Connector	Amazon Redshift V2 Connector
SQL ELT optimization	Use an ODBC connection with **ODBC Subtype**=`Redshift` to enable source or full SQL ELT optimization between Amazon Redshift source and target.	- Use an Amazon S3 V3 source connection and an Amazon Redshift V2 target connection to enable full SQL ELT optimization between Amazon S3 source and Amazon Redshift target. - Use an ODBC connection with **ODBC Subtype**=`Redshift` to SQL ELT optimization between Amazon Redshift source and target. **Note:** SQL ELT optimization does not apply to mappings in advanced mode.
Lookup transformation	Cache, uncached, and connected	Cache, uncached, connected, and unconnected **Note:** Uncached lookups do not apply to mappings in advanced mode.

The following table lists the Amazon Redshift functionality supported by the Amazon Redshift connectors:

Amazon Redshift Functionality	Amazon Redshift Connector	Amazon Redshift V2 Connector
VPC endpoints	Yes	Yes
Redshift Spectrum	No	Yes

Source functionality

When you import an object from Amazon Redshift to read data, you can configure the advance source properties to determine the read operation behavior. For example, you can read data in an encrypted format or you can configure partitioning for optimal performance.

The following table lists the source functionality you can use when you read data from an Amazon Redshift source for Amazon Redshift and Amazon Redshift V2 connectors:

Feature	Amazon Redshift Connector	Amazon Redshift V2 Connector
Staging directory[1]	Yes	Yes
Server-side encryption	Yes	Yes
Server-side encryption with KMS	Yes	Yes
Client-side encryption[1]	Yes	Yes
Unload command	Yes	Yes
Partitioning[1]	Yes	Yes
Working with large tables	Yes	Yes
Octal Values as DELIMITER and QUOTE	Yes	Yes

Feature	Amazon Redshift Connector	Amazon Redshift V2 Connector
Success and error files	Yes	Yes
Import objects from different schema	No	Yes
[1]Doesn't apply to mappings in advanced mode.		

Target functionality

When you import an object from Amazon Redshift to write data, you can configure the advance target properties to determine the write operation behavior. For example, you can write data in an encrypted format or you can retain Amazon S3 staging files after the write operation is complete.

The following table lists the target functionality you can use when you write data to an Amazon Redshift target:

Target functionality	Amazon Redshift Connector	Amazon Redshift V2 Connector
Staging directory[1]	Yes	Yes
Server-side encryption	Yes	Yes
Client-side encryption[1]	Yes	Yes
Analyze target table	Yes	Yes
Retain staging files	Yes	Yes
Copy Command	Yes	Yes
Vacuum Tables	Yes	Yes
Recovery and restart processing[1]	No	Yes
Preserve record order on write[1]	No	Yes
Working with Large Tables	Yes	Yes
Octal Values as DELIMITER and QUOTE	Yes	Yes
Success and Error Files[1]	Yes	Yes
Import objects from different schema	No	Yes
[1]Doesn't apply to mappings in advanced mode.		

CHAPTER 3

Task Examples

This section lists all the task examples that Amazon Redshift and Amazon Redshift V2 Connectors supports.

Synchronization task use case

You work for an e-commerce organization that stores sales order details in a MySQL database. Your organization needs to move the data from the MySQL database to an Amazon Redshift target.

Use Amazon Redshift Connector to create a synchronization task to write to an Amazon Redshift target.

Mapping and mapping task use case

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 to know the items ordered in a particular state and move data from the on-premise MySQL database to state-wise target tables in an affordable cloud-based environment.

Use Amazon Redshift V2 Connector to create a parameterized mapping to state-wise read purchase records from the MySQL database and write them to multiple Amazon Redshift targets to prepare an upcoming marketing campaign for all states.

Mapping task with Oracle CDC sources use case

our organization needs to replicate real-time changed data from a mission-critical Oracle production system to minimize intrusive, non-critical work, such as offline reporting or analytical operations system.

Use Amazon Redshift V2 Connector to capture changed data from the Oracle CDC source and write the changed data to an Amazon Redshift target table. Add the Oracle CDC sources in mappings, and then run the associated mapping tasks to write the changed data to the target.

Use case for a mapping in advanced mode

You work for an organization that stores large amount of purchase order details, such as customer ID, item codes, and item quantity in Amazon S3. You need to port the data from Amazon S3 to another cloud-based environment to quickly analyze the purchase order details and to increase future revenues.

Use Amazon Redshift V2 Connector to create a mapping in advanced mode to achieve faster performance when you read all the purchase records from Amazon S3 and write the records to an Amazon Redshift target.

Part II: Data Integration with Amazon Redshift V2 Connector

This part contains the following chapters:

- Introduction to Amazon Redshift V2 Connector, 19
- Connections for Amazon Redshift V2 , 21
- Amazon Redshift sources and targets, 48
- Mappings and mapping tasks with Amazon Redshift, 59
- Migrating a mapping, 89
- Upgrade the connection type, 92
- SQL ELT with Amazon Redshift V2 Connector, 97
- Data type reference, 123
- Troubleshooting, 128

CHAPTER 4

Introduction to Amazon Redshift V2 Connector

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.

You can also connect to Amazon Redshift Serverless cluster to read from or write data. You can move data from any data source to Amazon Redshift. Data Integration uses the Amazon driver to communicate with Amazon Redshift.

You can create an Amazon Redshift V2 connection and use the connection in mappings and mapping tasks. You can switch the mapping to advanced mode to include transformations and functions that enable advanced functionality. The advanced cluster can be a self-service cluster, a local cluster, or hosted on Amazon Web Services.

Create a mapping task to process data based on the data flow logic defined in a mapping or integration template. You can also create a mapping task to capture changed data from an Oracle CDC source and write the changed data to an Amazon Redshift target table.

When you run an Amazon Redshift V2 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.

Amazon Redshift V2 Connector assets

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

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

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

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

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.

CHAPTER 5

Connections for Amazon Redshift V2

You can use an Amazon Redshift V2 connection to read from or write to Amazon Redshift. You can use Amazon Redshift V2 connections to specify sources or targets in mappings and mapping tasks.

Prepare for authentication

You can configure **Default** and **Redshift IAM Authentication via AssumeRole** authentication types in an Amazon Redshift V2 connection to connect to Amazon Redshift. Additionally, you need to complete the S3 staging prerequisites to access S3 resources. You can also configure encryption, if required, to connect to Amazon Redshift.

See the following sections for a summary of the authentication, staging, and encryption prerequisites.

Authentication prerequisites

Before you begin, you need to have a registered user account with Amazon Redshift.

Get the minimum required details from your Amazon Redshift account from the AWS Console for the authentication type that you want to configure, as listed in the following table:

Default authentication	Redshift IAM Authentication via Assume Role
- JDBC URL - User name - Password	- JDBC URL - User name - Database name - Cluster identifier - Redshift IAM role ARN*
*To use the Redshift IAM role ARN, configure the Redshift IAM role ARN with the required trust policies to generate temporary security credentials to access Amazon Redshift. For instructions, see "Configure an assume role for Amazon Redshift" on page 24.	

Staging prerequisites

To enable staging on Amazon S3 and to gain access to S3 resources when you read or write data, you need to configure the staging properties in the Amazon Redshift V2 connection.

Note: If you do not require staging for the read operation, you can ignore the staging properties in the connection. You need to select the Direct mode in the Source transformation properties. Data Integration directly accesses and retrieves the data from Amazon Redshift without staging the data.

The following table summarizes the staging options that you can configure in the connection for both default and Redshift IAM Authentication via AssumeRole authentication and the tasks that you need to perform to get the required details for S3 staging:

S3 staging options	Tasks
Generate temporary credentials for the IAM user who assumes the S3 IAM role to access S3 staging.	**AWS configurations** Enable IAM users to assume an S3 IAM role and generate temporary credentials. For instructions, see the following references: - "Generate temporary security credentials using AssumeRole for Amazon S3 staging" on page 28. - Using an assume role for Amazon S3 resources How-To Library article. **Redshift V2 connection configurations** - Enter the value of the **S3 IAM Role ARN**. - Enter the **S3 Access Key ID** and **S3 Secret Access Key** values.
Generate temporary security credentials for an EC2 instance that assumes an S3 IAM role to access S3 staging.	**AWS configurations** Define an EC2 instance to assume an S3 IAM role and generate the temporary credentials for S3 staging. For instructions, see "Generate temporary security credentials using AssumeRole for EC2" on page 30. **Redshift V2 connection configurations** Configure the following minimum required properties: - Enable **Use EC2 Role to Assume Role**. - Enter the value of the **S3 IAM Role ARN**.
Generate the S3 access and secret access keys for the IAM user with access to the S3 bucket.	**AWS configurations** To generate the credentials, perform the following tasks: 1. "Create a minimal Amazon IAM policy" on page 23. 2. Create an IAM user, assign the policy to that user, and then generate the S3 access key ID and S3 secret access key in the AWS console. For more information about how to create an IAM user and generate keys, see the AWS documentation. **Redshift V2 connection configurations** Enter the **S3 Access Key ID** and **S3 Secret Access Key** values.
Configure IAM authentication	**AWS configurations** If you have an EC2 instance, and do not want to specify the keys or use the IAM role ARN, then assign the minimum policy to the EC2 with access to the S3 bucket. For instructions, see "Configure IAM authentication" on page 23. **Redshift V2 connection configurations** In this case, you do not need to enable or specify any of the staging properties in the connection. However, before you run the mapping, specify the Redshift role assigned to the Redshift cluster in the UnloadOptions and CopyOptions Property File fields in the mapping.

Encryption prerequisites

To configure client-side and server-side encryption for the Default authentication and Redshift IAM authentication via AssumeRole during staging, see "Enable encryption" on page 31.

Create a minimal Amazon IAM policy

To stage the data in Amazon S3, you need to create an IAM policy with the minimum required permissions to access the S3 resources.

You can either attach the policy to the IAM user and generate the S3 access key ID and S3 secret access keys to access S3 resources. Or, if you have an EC2 instance, you can assign the minimum policy to the EC2 instance to access the S3 bucket for staging.

You need the following minimum required permissions in the policy:

- PutObject
- GetObject
- DeleteObject
- ListBucket
- ListBucketMultipartUploads. Applicable only for mappings in advanced mode.

You can use the following sample Amazon IAM policy:

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

For mappings in advanced mode, you can use different AWS accounts within the same AWS region in the source and target connection. Make sure that the Amazon IAM policy confirms access to the AWS accounts specified in mappings.

Note: The **Test Connection** does not validate the IAM policy assigned to users. Hence, ensure that the policy assigned to the user is valid.

Configure IAM authentication

Configure AWS Identity and Access Management (IAM) authentication and create a minimal Amazon IAM policy for both the EC2 role and Redshift role.

For instructions, see the following How-to-Library article: Configuring AWS IAM Authentication

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

Configure an assume role for Amazon Redshift

To use the Redshift IAM role ARN, configure the Redshift IAM role ARN with the required trust policies to generate temporary security credentials to access Amazon Redshift.

You can use one of the following options to generate the temporary security credentials:

AWS configurations	Connection details
Option 1. Configure an AssumeRole to enable an IAM user.	To use the AssumeRole for the IAM user, specify the following IAM user details: - Redshift Access Key ID - Redshift Secret Access Key - Redshift IAM Role ARN
Option 2. Define an EC2 instance to assume a Redshift IAM role.	To use the AssumeRole for Amazon EC2: - Specify the **Redshift IAM Role ARN** value. - Enable the **Use EC2 Role to Assume Role** check box.

For more information about configuring an AssumeRole, see the following How-to-Library article:
Configure AssumeRole authentication for Amazon Redshift V2 Connector

Generate the temporary security credentials based on your requirement.

Generate temporary security credential policies for Amazon Redshift

To use the temporary security credentials to connect to Amazon Redshift, both the IAM user and IAM role require policies.

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

IAM user

An IAM user must have the `sts:AssumeRole` policy to use the temporary security credentials in the same or different AWS account. The IAM user credentials are used to key-in the Redshift access key and Redshift secret key in the connection properties.

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

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

Note: To run mappings in advanced mode, ensure to assign this policy to the Worker node role.

Redshift IAM role trust policy

The Redshift IAM role policy pertains to the role that is specified in the Redshift IAM Role ARN. An IAM role must have a trust policy attached with it to allow the IAM user to access Redshift using the temporary security credentials.

The following policy is a sample trust policy:

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

For example, you can specify the role or user in the following format:

```
{
"Version": "2012-10-17",
"Statement": [
    {
"Effect": "Allow",
"Principal": {
"AWS": "arn:aws:iam::<AWS-account>:role/<name-of-the-role>"
    },
       "Action": "sts:AssumeRole"
    }
 ]
}


{
"Version": "2012-10-17",
"Statement": [
    {
"Effect": "Allow",
"Principal": {
"AWS": " arn:aws:iam::<AWS-account>:user/<name-of-the-user>"
    },
       "Action": "sts:AssumeRole"
    }
 ]
}
```

Redshift IAM role trust policy for mappings in advanced mode

An IAM role must have a trust policy attached with it to allow the worker node to assume the Redshift role and access Amazon Redshift through the AssumeRole.

The following policy is a sample trust policy:

```
{
"Effect": "Allow",
"Principal": { "AWS": "arn:aws:iam::<ACCOUNT-ID>:role/<WORKER-NODE-ROLE-ARN>" },
"Action": "sts:AssumeRole"
}
```

Minimum permission policies of the Redshift IAM role

The following policy shows the permissions required to the Redshift IAM Role, which will be assumed by an IAM user to connect to the Redshift database using an existing Amazon Redshift user:

```
{
"Version": "2012-10-17",
"Statement": [
{
"Effect": "Allow",
"Action": [
          "redshift:GetClusterCredentials",

          "redshift:DescribeClusters"
```

```
            ],

"Resource": [
          "arn:aws:redshift:<REGION>:<ACCOUNT-ID>:dbuser:<Cluster_Identifier>/
<USER_NAME>",

          "arn:aws:redshift:<REGION>:<ACCOUNT-ID>:dbname:<Cluster_Identifier>/
<DATABASE_NAME>"

            ]

}
}
]
}
}
```

The following policy shows the permissions needed to be attached to the Redshift IAM Role, which will be assumed by an IAM user to connect to the Redshift database with a newly created user by the `Auto create DBUser` check box:

```
{

"Version": "2012-10-17",

"Statement": [

{

"Effect": "Allow",

"Action": [
          "redshift:GetClusterCredentials",

          "redshift:DescribeClusters",

          "redshift:CreateClusterUser",

          "redshift:JoinGroup"
            ],

"Resource": [
          "arn:aws:redshift:<REGION>:<ACCOUNT-ID>:dbuser:<Cluster_Identifier>/
<USER_NAME>",

          "arn:aws:redshift:<REGION>:<ACCOUNT-ID>:dbname:<Cluster_Identifier>/
<DATABASE_NAME>",

          "arn:aws:redshift:<REGION>:<ACCOUNT-ID>:dbgroup:<Cluster_Identifier>/
<GROUP_NAME>"
            ]
}

}

]

}
```

Generate temporary security credentials using AssumeRole for EC2

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

The Amazon EC2 role can assume another IAM role from the same or different AWS account without requiring a Redshift access key and Redshift secret key.

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

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

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

EC2 service role trust policy

The following is a sample trust policy that is defined in a trust relationship of the EC2 role attached to the EC2 instance:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "ec2.amazonaws.com"
            },
            "Action": "sts:AssumeRole"
        }
    ]
}
```

The following is a sample trust policy of the Redshift IAM role when you enable EC2 assume role:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "redshift.amazonaws.com"
            },
            "Action": "sts:AssumeRole"
        },
        {
            "Effect": "Allow",
            "Principal": {
                "AWS": "arn:aws:iam::<Account-ID:role>/ec2_role_attached_to_ec2_instance"
            },
            "Action": "sts:AssumeRole"
        }
    ]
}
```

The permission policy that is required to be attached to the EC2 instance is same as the policy defined for the IAM user.

Configure an assume role for Amazon S3 staging

To configure AssumeRole authentication for S3 staging, you need to attach the minimum permission policies and trust policies for the IAM user and IAM role in the AWS console.

An IAM user can use the AssumeRole to temporarily gain access to the Amazon S3 resources. For more information about using an assume role for Amazon S3 resources, you can also refer to the How-to-Library article: Using an assume role for Amazon S3 resources

You can generate temporary security credentials using AssumeRole for Amazon S3 staging to access the Amazon S3 staging bucket. If you want EC2 instances to assume an IAM role to gain access to the S3 staging bucket securely, use the temporary security credentials generated using AssumeRole for EC2 instances.

Note: Do not use the root user credentials of the AWS account to generate the temporary security credentials. You need to use the credentials of an IAM user to generate the temporary security credentials.

Generate the temporary security credentials based on your requirement.

Generate temporary security credentials using AssumeRole for Amazon S3 staging

You can use the temporary security credentials using AssumeRole to access the Amazon S3 staging bucket from the same or different AWS accounts.

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

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

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

External ID

You can specify the external ID for a more secure access to the Amazon S3 bucket when the Amazon S3 bucket is in a different AWS account than the IAM user or EC2 instance.

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

The external ID must be a string. The following sample shows an external ID condition in the assumed IAM role's trust policy:

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

Temporary security credentials policy

To use the temporary security credentials to access the Amazon S3 staging bucket, both the IAM user and IAM role require policies.

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

IAM user

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

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

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

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

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

IAM role

An IAM role must have a `sts:AssumeRole` policy and a trust policy attached with the IAM role to allow the IAM user to access the Amazon S3 bucket using the temporary security credentials. The policy specifies the Amazon S3 bucket that the IAM user can access and the actions that the IAM user can perform. The trust policy specifies the IAM user from the AWS account that can access the Amazon S3 bucket.

The following policy is a sample trust policy:

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

Temporary security credentials for KMS

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

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

- GenerateDataKey
- DescribeKey

- Encrypt
- Decrypt
- ReEncrypt

You can use the following sample policy:

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

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

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

Generate temporary security credentials using AssumeRole for EC2

You can use temporary security credentials using AssumeRole for an Amazon EC2 role to access the Amazon S3 staging bucket from the same or different AWS accounts.

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

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

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

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

Enable encryption

You can enable client-side and server-side encryption in the Amazon Redshift V2 connection for staging data in Amazon S3.
You can also configure encryption in the Amazon Redshift V2 Source and Target transformations in a mapping. For more information on how to configure encryption in the mapping, see the topics "Data encryption in Amazon Redshift V2 sources" and "Data encryption in Amazon Redshift V2 targets."

Complete the prerequisites based on the type of encryption that you want to configure in the Amazon Redshift V2 connection.

Client-side encryption

Client-side encryption requires a 256-bit AES encryption key in the Base64 format. You can generate a key using a third-party tool.

Specify the key value in the **Master Symmetric Key** field when you create an Amazon Redshift V2 connection.

Server-side encryption

To enable server-side encryption, create an AWS Key Management Service (AWS KMS)-managed customer master key.

Generate the customer master key ID for the same region where your Amazon S3 staging bucket resides. For more information about generating a customer master key, see the AWS documentation.

To enable encryption with the customer master key, you need to create a minimal KMS policy. You can specify the customer master key ID when you create an Amazon Redshift V2 connection.

Note: You cannot configure server-side encryption with the master symmetric key and client-side encryption with the customer master key.

Create a minimal policy for using AWS KMS

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

You can perform the following operations to enable encryption with KMS:

- GenerateDataKey
- DescribeKey
- Encrypt
- Decrypt
- ReEncrypt

Sample policy:

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

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

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

Connect to Amazon Redshift

Let's configure the Amazon Redshift V2 connection properties to connect to Amazon Redshift.

Before you begin

Check out "Prepare for authentication" on page 21 to learn about the authentication requirements before you configure a connection.

Connection details

The following table describes the basic connection properties:

Property	Description
Connection Name	Name of the connection. Each connection name must be unique within the organization. Connection names can contain alphanumeric characters, spaces, and the following special characters: _. + -, Maximum length is 255 characters.
Description	Description of the connection. Maximum length is 4000 characters.
Type	Amazon Redshift V2
Runtime Environment	Name of the runtime environment where you want to run tasks. Select a Secure Agent, Hosted Agent, or serverless runtime environment. **Note:** Hosted Agent doesn't apply for mappings that run on an advanced cluster. You also cannot use the Hosted Agent for IAM authentication and EC2 AssumeRole authentication.

Authentication types

You can configure default and Redshift IAM AssumeRole authentication types to access Amazon Redshift. Select the required authentication method and then configure the authentication-specific parameters.

Default authentication

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

Properties	Description
JDBC URL	The JDBC URL to connect to the Amazon Redshift cluster. You can get the JDBC URL from your Amazon AWS Redshift cluster configuration page. Enter the JDBC URL in the following format: `jdbc:redshift://<cluster_endpoint>:<port_number>/<database_name>`, where the endpoint includes the Redshift cluster name and region. For example, `jdbc:redshift://infa-rs-cluster.abc.us-west-2.redshift.amazonaws.com:5439/rsdb` In the example, - infa-rs-qa-cluster is the name of the Redshift cluster. - us-west-2.redshift.amazonaws.com is the Redshift cluster endpoint, which is the US West (Oregon) region. - 5439 is the port number for the Redshift cluster. - rsdb is the specific database instance in the Redshift cluster to which you want to connect.
Username	User name of your database instance in the Amazon Redshift cluster.
Password	Password of the Amazon Redshift database user.
Use EC2 Role to Assume Role	Enables the EC2 instance that assumes an S3 IAM role to access the S3 resources to stage data using the temporary security credentials. The EC2 role must have a policy attached with permissions to assume an S3 IAM role. The S3 IAM role and the EC2 instance can be in the same or different AWS account. Select the check box to enable the EC2 role to assume an S3 IAM role specified in the S3 IAM Role ARN option to access the S3 resources for staging data. For instructions, see "Generate temporary security credentials using AssumeRole for EC2" on page 30.
S3 IAM Role ARN	The Amazon Resource Number (ARN) of the IAM role assumed by the IAM user or EC2 to use the dynamically generated temporary security credentials to stage data in Amazon S3. This property applies when you want to generate temporary security credentials to access the S3 staging buckets by using either the EC2 instance or the IAM user who assumes the S3 IAM role. Specify the S3 IAM role name to use the temporary security credentials to access the Amazon S3 staging bucket. For more information about how to get the ARN of the S3 IAM role, see the AWS documentation.

Advanced settings

The following table describes the advanced connection properties for default authentication:

Properties	Description
S3 Access Key ID	Access key of the IAM user to access the Amazon S3 staging bucket. Enter the access key ID when you use the following methods for S3 staging: - When the IAM user has access to S3 staging. - When the IAM user who assumes the S3 IAM role uses the temporary security credentials to access S3. You do not need to enter the S3 access key ID if you use IAM authentication or the assume role for EC2 to access S3.
S3 Secret Access Key	Secret access key to access the Amazon S3 staging bucket. The secret key is associated with the access key and uniquely identifies the account. Enter the secret access key value when you use following methods for S3 staging: - When the IAM user has access to S3 staging. - When the IAM user who assumes the S3 IAM role uses the temporary security credentials to access S3. You do not need to enter the S3 secret access key if you use IAM authentication or the assume role for EC2 to access S3.
S3 VPC Endpoint Type	The type of Amazon Virtual Private Cloud endpoint for Amazon S3. You can use a VPC endpoint to enable private communication with Amazon S3. Select one of the following options: - Default. Select if you do not want to use a VPC endpoint. - Interface Endpoint. Select to establish private communication with Amazon S3 through an interface endpoint which uses a private IP address from the IP address range of your subnet. It serves as an entry point for traffic destined to an AWS service.
Endpoint DNS Name for Amazon S3	The DNS name for the Amazon S3 interface endpoint. Replace the asterisk symbol with the bucket keyword in the DNS name. Enter the DNS name in the following format: `bucket.<DNS name of the interface endpoint>` For example, `bucket.vpce-s3.us-west-2.vpce.amazonaws.com`
STS VPC Endpoint Type	The type of Amazon Virtual Private Cloud endpoint for AWS Security Token Service. You can use a VPC endpoint to enable private communication with Amazon Security Token Service. Select one of the following options: - Default. Select if you do not want to use a VPC endpoint. - Interface Endpoint. Select to establish private communication with Amazon Security Token Service through an interface endpoint which uses a private IP address from the IP address range of your subnet.
Endpoint DNS Name for AWS STS	The DNS name for the AWS STS interface endpoint. For example, `vpce-01f22cc14558c241f-s8039x4c.sts.us-west-2.vpce.amazonaws.com`

Properties	Description
KMS VPC Endpoint Type	The type of Amazon Virtual Private Cloud endpoint for AWS Key Management Service. You can use a VPC endpoint to enable private communication with Amazon Key Management Service. Select one of the following options: - Default. Select if you do not want to use a VPC endpoint. - Interface Endpoint. Select to establish private communication with Amazon Key Management Service through an interface endpoint which uses a private IP address from the IP address range of your subnet.
Endpoint DNS Name for AWS KMS	The DNS name for the AWS KMS interface endpoint. For example, `vpce-0e722f5c721e19232-g2pkm2r7.kms.us-west-2.vpce.amazonaws.com`
External ID	The external ID associated with the IAM role. You can specify the external ID if you want to provide a more secure access to the Amazon S3 bucket. The Amazon S3 staging bucket and the IAM role can be in the same or different AWS accounts. If required, you also have the option to specify the external ID in the AssumeRole request to the AWS Security Token Service (STS) using an external ID condition in the assumed IAM role's trust policy. For more information about using an external ID, see External ID when granting access to your AWS resources.

Properties	Description
Cluster Region	The AWS cluster region in which the Redshift cluster resides. Select the cluster region from the list if you choose to provide a custom JDBC URL with a different cluster region from that specified in the **JDBC URL** field property. To continue to use the cluster region name specified in the **JDBC URL** field property, select **None** as the cluster region in this property. You can only read data from or write data to the cluster regions supported by the AWS SDK. Select one of the following cluster regions: None Asia Pacific(Mumbai) Asia Pacific(Seoul) Asia Pacific(Singapore) Asia Pacific(Sydney) Asia Pacific(Tokyo) Asia Pacific(Hong Kong) AWS GovCloud (US) AWS GovCloud (US-East) Canada(Central) China(Bejing) China(Ningxia) EU(Ireland) EU(Frankfurt) EU(Paris) EU(Stockholm) South America(Sao Paulo) Middle East(Bahrain) US East(N. Virginia) US East(Ohio) US West(N. California) US West(Oregon) Default is **None**.
Connection Environment SQL	The SQL statement to set up the database environment that applies for the entire session. Separate multiple values with a semicolon (;). Specify only the configurations for the database environment in the SQL statement. Do not specify any DDL or DML commands in the SQL statement.
Master Symmetric Key[1]	A 256-bit AES encryption key in the Base64 format that enables client-side encryption to encrypt your data before you send them for staging in Amazon S3. For more information, see "Enable encryption" on page 31.
Customer Master Key ID	The customer master key ID generated by AWS Key Management Service (AWS KMS) or the ARN of your custom key for cross-account access when you stage data in Amazon S3. The customer master key serves to encrypt your data at the destination before they are saved in Amazon S3. You can either enter the customer-generated customer master key ID or the default customer master key ID. You can use a cross account KMS key in a connection in a mapping in advanced mode. The cluster and the staging bucket needs to be in the same region.

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

Redshift IAM Authentication via AssumeRole

The Redshift AssumeRole authentication enables the user to assume an IAM role or define an EC2 role configured with required trust policies to generate temporary security credentials to access Amazon Redshift.

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

Properties	Description
JDBC URL	The JDBC URL to connect to the Amazon Redshift cluster. You can get the JDBC URL from your Amazon AWS Redshift cluster configuration page. Enter the JDBC URL in the following format: `jdbc:redshift://<cluster_endpoint>:<port_number>/<database_name>`, where the endpoint includes the Redshift cluster name and region. For example, `jdbc:redshift://infa-rs-cluster.abc.us-west-2.redshift.amazonaws.com:5439/rsdb` In the example, - infa-rs-qa-cluster is the name of the Redshift cluster. - us-west-2.redshift.amazonaws.com is the Redshift cluster endpoint, which is the US West (Oregon) region. - 5439 is the port number for the Redshift cluster. - rsdb is the specific database instance in the Redshift cluster to which you want to connect.
Username	User name of your database instance in the Amazon Redshift cluster.
Cluster Identifier	The unique identifier of the cluster that hosts Amazon Redshift. Specify the Amazon Redshift cluster name.
Database Name	Name of the Amazon Redshift database where the tables that you want to access are stored.
Redshift IAM Role ARN	The Amazon Resource Number (ARN) of the IAM role assumed by EC2 to use the dynamically generated temporary security credentials to access Amazon Redshift. Enter the Redshift IAM role ARN to access the Amazon Redshift cluster.

Properties	Description
Use EC2 Role to Assume Role	Enables the EC2 role to assume an IAM role, either to connect to Redshift or to stage data using the temporary security credentials: **Connect to Redshift with IAM authentication using the EC2 role** Select the check box to enable the EC2 role that assumes a Redshift IAM role specified in the **Redshift IAM Role ARN** field to access Amazon Redshift. The EC2 role must have a policy attached with permissions to assume a Redshift IAM role from the same or different account. **Access S3 resources to stage data** Select the check box to enable the EC2 role to assume an S3 IAM role specified in the **S3 IAM Role ARN** field and dynamically generate the temporary security credentials to access the S3 staging buckets. The EC2 role must have a policy attached with permissions to assume an S3 IAM role from the same or different AWS account.
S3 IAM Role ARN	The Amazon Resource Number (ARN) of the S3 IAM role assumed by the IAM user or EC2 to use the dynamically generated temporary security credentials to stage data in Amazon S3. This property applies when you want to generate the temporary security credentials to access the S3 staging buckets by using either the EC2 instance or the IAM user who assumes the S3 IAM role. Specify the S3 IAM role name to use the temporary security credentials to access the Amazon S3 staging bucket. For more information about how to get the ARN of the IAM role, see the AWS documentation.

Advanced settings

The following table describes the advanced connection properties for Redshift IAM AssumeRole authentication:

Properties	Description
Redshift Access Key ID	The access key of the IAM user that has permissions to assume the Redshift IAM AssumeRole ARN. This property doesn't apply to Amazon Redshift AssumeRole authentication with EC2 role.
Redshift Secret Access Key	The secret access key of the IAM user that has permissions to assume the Redshift IAM Assume Role ARN. This property doesn't apply to Amazon Redshift AssumeRole authentication with EC2 role.
Database Group	The name of the database group to which you want to add the database user when you select the **Auto Create DBUser** option in this connection property. The user that you add to this database group inherits the specified group privileges. If you do not specify a database group name, the user is added to the public group and inherits its associated privileges. You can also enter multiple database groups, separated by a comma, to add the user to each of the specified database groups.
Expiration Time	The time duration that the password for the Amazon Redshift database user expires. Specify a value between 900 seconds and 3600 seconds. Default is 900.

Properties	Description
Auto Create DBUser	Select to create a new Amazon Redshift database user at run time. The agent adds the user you specified in the **Username** field to the database group. The added user assumes the privileges assigned to the database group. Default is disabled.
S3 Access Key ID	Access key of the IAM user to access the Amazon S3 staging bucket. Enter the access key ID when you use the following methods for S3 staging: - When the IAM user has access to S3 staging. - When the IAM user who assumes the S3 IAM role uses the temporary security credentials to access S3. You do not need to enter the S3 access key ID if you use IAM authentication or the assume role for EC2 to access S3.
S3 Secret Access Key	Secret access key to access the Amazon S3 staging bucket. The secret key is associated with the access key and uniquely identifies the account. Enter the secret access key value when you use following methods for S3 staging: - When the IAM user has access to S3 staging. - When the IAM user who assumes the S3 IAM role uses the temporary security credentials to access S3. You do not need to enter the S3 secret access key if you use IAM authentication or the assume role for EC2 to access S3.
S3 VPC Endpoint Type	The type of Amazon Virtual Private Cloud endpoint for Amazon S3. You can use a VPC endpoint to enable private communication with Amazon S3. Select one of the following options: - Default. Select if you do not want to use a VPC endpoint. - Interface Endpoint. Select to establish private communication with Amazon S3 through an interface endpoint which uses a private IP address from the IP address range of your subnet. It serves as an entry point for traffic destined to an AWS service.
Endpoint DNS Name for Amazon S3	The DNS name for the Amazon S3 interface endpoint. Replace the asterisk symbol with the bucket keyword in the DNS name. Enter the DNS name in the following format: `bucket.<DNS name of the interface endpoint>` For example, `bucket.vpce-s3.us-west-2.vpce.amazonaws.com`
STS VPC Endpoint Type	The type of Amazon Virtual Private Cloud endpoint for AWS Security Token Service. You can use a VPC endpoint to enable private communication with Amazon Security Token Service. Select one of the following options: - Default. Select if you do not want to use a VPC endpoint. - Interface Endpoint. Select to establish private communication with Amazon Security Token Service through an interface endpoint which uses a private IP address from the IP address range of your subnet.
Endpoint DNS Name for AWS STS	The DNS name for the AWS STS interface endpoint. For example, `vpce-01f22cc14558c241f-s8039x4c.sts.us-west-2.vpce.amazonaws.com`

Properties	Description
KMS VPC Endpoint Type	The type of Amazon Virtual Private Cloud endpoint for AWS Key Management Service. You can use a VPC endpoint to enable private communication with Amazon Key Management Service. Select one of the following options: - Default. Select if you do not want to use a VPC endpoint. - Interface Endpoint. Select to establish private communication with Amazon Key Management Service through an interface endpoint which uses a private IP address from the IP address range of your subnet.
Endpoint DNS Name for AWS KMS	The DNS name for the AWS KMS interface endpoint. For example, `vpce-0e722f5c721e19232-g2pkm2r7.kms.us-west-2.vpce.amazonaws.com`
External ID	The external ID associated with the IAM role. You can specify the external ID if you want to provide a more secure access to the Amazon S3 bucket when the Amazon S3 staging bucket is in same or different AWS accounts. If required, you also have the option to specify the external ID in the AssumeRole request to the AWS Security Token Service (STS) using an external ID condition in the assumed IAM role's trust policy. For more information about using an external ID, see External ID when granting access to your AWS resources.

Properties	Description
Cluster Region	The AWS geographical region in which the Redshift cluster resides. Select the cluster region from the list if you choose to provide a custom JDBC URL with a different cluster region from that specified in the **JDBC URL** field property. To continue to use the cluster region name specified in the **JDBC URL** field property, select **None** as the cluster region in this property. You can only read data from or write data to the cluster regions supported by the AWS SDK. Select one of the following cluster regions: None Asia Pacific(Mumbai) Asia Pacific(Seoul) Asia Pacific(Singapore) Asia Pacific(Sydney) Asia Pacific(Tokyo) Asia Pacific(Hong Kong) AWS GovCloud (US) AWS GovCloud (US-East) Canada(Central) China(Bejing) China(Ningxia) EU(Ireland) EU(Frankfurt) EU(Paris) EU(Stockholm) South America(Sao Paulo) Middle East(Bahrain) US East(N. Virginia) US East(Ohio) US West(N. California) US West(Oregon) Default is **None**.
Connection Environment SQL	The SQL statement to set up the database environment that applies for the entire session. Separate multiple values with a semicolon (;). Specify only the configurations for the database environment in the SQL statement. Do not specify any DDL or DML commands in the SQL statement.
Master Symmetric Key[1]	A 256-bit AES encryption key in the Base64 format that enables client-side encryption to encrypt your data before you send them for staging in Amazon S3. For more information, see "Enable encryption" on page 31.

Properties	Description
Customer Master Key ID	The customer master key ID generated by AWS Key Management Service (AWS KMS) or the ARN of your custom key for cross-account access when you stage data in Amazon S3. The customer master key serves to encrypt your data at the destination before they are saved in Amazon S3. You can either enter the customer-generated customer master key ID or the default customer master key ID. You can use a cross account KMS key in a connection in a mapping in advanced mode. The cluster and the staging bucket needs to be in the same region. For more information about how to configure server-side encryption, see "Enable encryption" on page 31.
[1]Doesn't apply to mappings in advanced mode.	

Proxy server settings

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

You can configure the Secure Agent to use the proxy server on Windows and Linux. You can use only an unauthenticated proxy server. The proxy settings applies to connections used in mappings and in mappings in advanced mode.

To configure the proxy settings for the Secure Agent, use one of the following methods:

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

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

To configure proxy settings for the serverless runtime environment, see "Using a proxy server" in *Runtime Environments* in the Administrator help.

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.

Configure SSL

To use SSL to connect to Amazon Redshift, you need to configure the Secure Agent for SSL and enable SSL through the JDBC URL in the Amazon Redshift V2 connection properties.

1. Download the Amazon Redshift certificate from the following location: https://s3.amazonaws.com/redshift-downloads/redshift-ssl-ca-cert.pem.
2. At the command prompt, 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 from the list of 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 that you specify in the 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.

Configure SSL with the serverless runtime environment

You can use the serverless runtime environment in an Amazon Redshift V2 connection to connect to an SSL-enabled Amazon Redshift database.

Before you configure a secure Amazon Redshift V2 connection using the serverless runtime environment, perform the following tasks:

- Add the SSL certificate in the Amazon S3 bucket or Azure container.
- Configure the .yml serverless configuration file.
- Configure the serverless environment.
- Configure the connection properties to use SSL.

Add the SSL certificate in the Amazon S3 bucket or Azure container

Perform the following steps to configure an SSL connection in a serverless runtime environment:

1. Create the following structure for the serverless agent configuration in AWS or Azure: `<Supplementary file location>/serverless_agent_config`
2. Add the certificate name and source path in the Amazon S3 bucket or Azure container in the following location in your AWS or Azure account: `<Supplementary file location>/serverless_agent_config/SSL`

Configure the .yml serverless configuration file

Perform the following steps to configure the .yml serverless configuration file in the serverless runtime environment and add the certificate name and path entries so that Amazon Redshift V2 Connector can use SSL:

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

```
version: 1
agent:
  agentAutoApply:
    general:
      sslStore:
        - fileCopy:
            sourcePath: SSL/<cert_name>
        - importCerts:
            certName: <cert_name>
            alias: <alias name of the certificate>
```

where the source path is the directory path of the certificate files in AWS or Azure.

2. Ensure that the syntax and indentations are valid, and then save the file as `serverlessUserAgentConfig.yml` in the following AWS or Azure location: `<Supplementary file location>/serverless_agent_config`
When the .yml file runs, the SSL certificates are copied from the AWS or Azure location to the serverless agent directory.

Configure the serverless environment

Configure the JVMOption1 and JVMOption2 properties for SSL in the serverless runtime environment:

1. Navigate to your serverless runtime environment properties, and click **Edit**.
2. On the **Runtime Configuration Properties** tab, click **JVMoption1** and add the following property:
`-Djavax.net.ssl.trustStore=/home/cldagnt/SystemAgent/jdk/jre/lib/security/cacerts`
3. Click **JVMoption2** and add the following property:
`-Djavax.net.ssl.trustStorePassword=changeit`
4. Click **Save**.
5. Redeploy the runtime environment.

Configure the connection properties to use SSL

After you set the runtime properties in the serverless runtime environment, specify `ssl=true` in the **JDBC URL** connection property.

For example, `jdbc:redshift://mycluster.xyz789.us-west-2.redshift.amazonaws.com:5439/dev?ssl=true`

Configure client-side encryption with the serverless runtime environment

You can use the serverless runtime environment to configure client-side encryption when you connect to Amazon Redshift.

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

Configure the .yml serverless configuration file

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

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

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

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

`<Supplementary file location>/serverless_agent_config`

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

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

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

Configure SSE-KMS encryption for mappings in advanced mode

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

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

Amazon Redshift Serverless connectivity

Amazon Redshift Serverless is a serverless offering of Amazon Web Services (AWS) that allows the same scalability and capability of Amazon Redshift without the need to set up and manage the provisioned Redshift cluster.
Amazon Redshift V2 Connector provides out-of-the-box support to connect to the Amazon Redshift Serverless endpoint.

For more information about how to access the Amazon Redshift serverless endpoint, see the How-to-Library article: Using Amazon Redshift Serverless with Cloud Data Integration

Requirements to use Amazon Redshift Spectrum

When you use a connection in a mapping to read data from an Amazon Redshift Spectrum external table, provide the required authorization to the Amazon Redshift cluster to access the data catalog and the data files in Amazon S3.
Important: 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.

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.
6. 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.

Private communication with Amazon Redshift

If you do not want to expose your traffic to the public internet, you can enable private communication with Amazon Redshift by configuring a gateway endpoint on the AWS console.

To establish a private connection with Amazon Redshift, ensure that the Secure Agent is a part of the subnet in the AWS Virtual Private Cloud (VPC). You can create a gateway endpoint and stage the Amazon S3 data to Amazon Redshift.

To configure private communication to connect to Amazon Redshift, you need to perform the following tasks:

- Create a cluster subnet group.
- Create a Redshift-managed VPC endpoint.
- Configure the gateway endpoint.

You can then specify the gateway endpoint in the Amazon Redshift V2 connection properties.

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

Private communication with Amazon S3

You can configure an Amazon Redshift V2 connection to establish private communication with Amazon S3 for staging.

You need to configure an interface endpoint on the AWS console to enable private communication to stage data in Amazon S3. The AWS S3 VPC endpoint enables an S3 request to be routed to the Amazon S3 service, without connecting to the internet.

Consider the following guidelines to establish private communication with Amazon S3:

- When an Amazon Redshift Cluster is spawned, an Elastic Network Interface (ENI) is generated for the cluster within the subnet. Ensure that the route table of the S3 gateway endpoint corresponds to the same subnet where your Redshift cluster ENI is created.
- To connect to an Amazon S3 bucket using a VPC endpoint, the Amazon Redshift cluster and the Amazon S3 bucket that it connects to must be in the same AWS region.

VPC peering between the serverless runtime environment and Amazon Redshift

When you use the serverless runtime environment and if the serverless runtime environment and the Amazon Redshift cluster reside in different VPCs, you need to configure VPC peering.

For more information about configuring VPC peering, see the How-to-Library article:
Configure VPC peering between Amazon Redshift clusters

CHAPTER 6

Amazon Redshift sources and targets

You can assign the source and target properties for an Amazon Redshift connection.

Amazon Redshift V2 sources

You can use an Amazon Redshift V2 object as a source in a mapping. You can use tables, views, or materialized views as Amazon Redshift V2 sources.

When you configure the advanced source properties, 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.

Read from Amazon Redshift with or without staging

When you configure a read operation, you have the option to read data from the Amazon Redshift source with or without staging data on the S3 bucket.

You can set the read mode to direct or staging in the Source and Lookup transformations. You cannot configure direct mode in a mapping in advanced mode.

Staging read mode

If you choose to set the read mode to staging, the agent creates a staging file in the directory that you specify in the mapping properties. You need to specify a staging directory on the Secure Agent machine that has sufficient disk space corresponding to the volume of data that you want to process.

When you run the mapping, the UNLOAD command loads the data from the Amazon Redshift table to the Amazon S3 bucket, and then the agent loads the data from S3 to the staging directory on the agent machine. After the agent completes the read operation, the staging files are deleted.

Direct read mode

When you set the read mode to direct, Data Integration directly accesses and retrieves the data from Amazon Redshift without staging the data.

When you configure the direct read mode, you can also specify the fetch size to determine the number of rows of data to retrieve from the Redshift source at a given time. The advanced source attributes applicable for S3 staging in the mapping properties and the staging optimization property in the agent properties are not required for the direct mode.

For the list of advanced source attributes that apply for direct and staging read modes, see "Amazon Redshift V2 sources in mappings" on page 62.

Data encryption in Amazon Redshift V2 sources

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

Sslect the type of the encryption in the **Encryption Type** field under the Amazon Redshift V2 advanced source properties on the **Runtime Options** 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 or Amazon Resource Name (ARN) for server-side encryption.

The AWS KMS-managed customer master key or ARN that you specify 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.

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 or specify the path to a text file that contains all the command options.

Enter the options in uppercase and use a semicolon to separate the options.

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

Note: You cannot specify the path to a text file that contains the Unload command options in a mapping in advanced mode.

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 must 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.

The default value is \036 which represents the octal representation of the non-printable known as the record separator.

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

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`.

Note: If you enable the optimization property that you set for staging data, you must use `ESCAPE=OFF` in Unload command options. If you do not specify `ESCAPE=OFF`, the mapping runs without optimizing the staging performance.

REGION

You can use the REGION attribute when the Amazon S3 staging bucket is not in the same region as the cluster region. If Amazon Redshift resides in the US East (N. Virginia) region, you can use an Amazon S3 bucket residing in the Asia Pacific (Mumbai) region to create staging files. For example, `REGION = ap-south-1`.

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.

NULL

You can use NULL Unload command option to replace the null values in an Amazon Redshift source table with the string that you specify using the NULL Unload command option.

Enter the value of the NULL Unload command option in the following format: `NULL=text`. Do not add spaces when you enter the string value.

Note: The NULL Unload command option does not apply to mappings in advanced mode.

For more information about the NULL Unload command, see the AWS documentation.

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`

ADDQUOTES

ADDQUOTES is implemented with the UNLOAD command by default. The Unload command adds quotation marks to each data field. With added quotation marks, the UNLOAD command can read data values that contain the delimiter. If double quote (") is a part of data, use ESCAPE to read the double quote as a regular character.

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 Secure Agent distributes data among partitions at partition points. You cannot configure source partitioning when you run a mapping in advanced mode.

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.

Note: When you configure key range partitioning for the source and create a target, the mapping task fails.

You cannot configure filters when you use source partitioning.

Amazon Redshift V2 targets

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

When you configure the advanced target properties, 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 writing the data to Amazon Redshift.

The Secure Agent deletes the staged files from the staging directory after writing 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>`

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

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 and select **S3 Client Side Encryption** in the advanced target 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.

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 or specify the path to a text file that contains all the command options.

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.

Note: You cannot specify the path to a text file that contains the Copy command options in a mapping in advanced mode.

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.

The default value is \036 which represents the octal representation of the non-printable known as the 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.

REGION

You can use the REGION attribute when the Amazon S3 staging bucket is not in the same region as the cluster region. If Amazon Redshift resides in the US East (N. Virginia) region, you can use an Amazon S3 bucket residing in the Asia Pacific (Mumbai) region to create staging files. For example, `REGION = ap-south-1`.

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 column.

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`

IGNOREBLANKLINES

If the input rows contain NULL values, set the `IGNOREBLANKLINES` property to `ON` in the **Copy Options** to ignores blank lines while inserting the data to the target. Specify the property in the following format: `IGNOREBLANKLINES=ON`The property helps ignores blank lines that only contain a line feed in a data file and does not try to load them.

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.

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 the Amazon S3 location specified in the target. 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.

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.

Recovery and restart processing

When you run a mapping task to capture changed data from a CDC source and write the changed data to an Amazon Redshift target table, Amazon Redshift V2 Connector supports recovery and restart processing. You cannot use recovery and restart processing when you run a mapping in advanced mode.

If a mapping task fails or is stopped before completing the task, the Secure Agent uses the recovery information stored in the `infa_recovery_table` table on the target system to resume the extraction of changed data from the point of interruption. This functionality prevents changed data loss and inconsistencies between the source and target.

To enable recovery and restart processing, set the **Recovery Strategy** advanced session property to **Resume from last checkpoint** on the **Runtime Options** page when you create or edit a mapping task. With this setting, the mapping task can resume processing changed data from the point of interruption.

In special situations, you can specify a restart point for a mapping task. Typically, the first time you start a mapping task, you specify a restart point that corresponds to the target materialization time so that no change records are skipped. The default restart point is the end of log (EOL), which is the current point of CDC processing in the log. You can specify a restart point that corresponds to the extraction processing starting from the earliest available record in the log or from a specific date and time. When you use a time-based restart point, extraction processing starts in the log that contains the first unit-of-work (UOW) that has an end time later than the restart time.

When you specify a restart point, consider the following points:

- The restart point applies to all sources in the mapping that is associated with the mapping task.
- If you set a restart point that is too early, it might correspond to a expired log file. In this case, the value of the restart point is considered as the earliest available record in the available log files.
- If you set a restart point that is later than the latest record in the log files, an error message is issued.

Note: Restart information is associated with a mapping task, a specific source and target combination. If you change the source object in a mapping, you must either create a new mapping task for the mapping or increment the restart revision number for the existing mapping task. To increment the restart revision number, navigate to the **CDC Runtime** page for the mapping task, open the **Select Restart Point** dialog box, and click **OK**. If you do not take one of these actions, the mapping task will fail the next time you run it.

Preserve record order on Write

You can retain the order number of the changed record when you capture the changed record from a CDC source to a target table. This property enables you to avoid inconsistencies between the CDC source and target. You cannot use the preserve record order on write property when you run a mapping in advanced mode.

When you modify a single record in a row several times in a CDC source, enable the **Preserve record order on write** option in the advanced target property to retain the order number of the changed record when you write the changed record to the target table.

For example, you have a record in the following CDC source table in which you have performed multiple of operations:

Emp ID	Emp Name	Emp Description	RowType	RowID
1	John	L1	Insert	1
1	John	L2	Update	2
1	John	L3	Update	3

Here, assume that the `RowID` shows the order of the changed record in the CDC source table.

The Secure Agent writes the following changed record along with the order number in the target table:

Emp ID	Emp Name	Emp Description	RowType	RowID
1	John	L3	Update	3

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. You cannot use success and error files when you run a mapping in advanced mode.

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 7

Mappings and mapping tasks with Amazon Redshift

Create a mapping task to process data based on the data flow logic defined in a mapping or integration template. You can also create a mapping task to capture changed data from the Oracle CDC source and write the changed data to an Amazon Redshift target table. You can switch the mapping to advanced mode to include transformations and functions that enable advanced functionality.

Caution: The pre-SQL and post-SQL commands configured in the Source, Target, and Lookup transformations are not validated at run time. Ensure that the command syntax is free from potentially unsafe content and the database user you configured in the connection object has appropriate privileges. It is recommended to use TLS- enabled database connections for secure network communications.

Before you begin

Perform the following tasks before you can create a mapping:

General prerequisites

Ensure that you have access to the Secure Agent directory that contains the success and error files. The directory path must be the same on each Secure Agent machine in the runtime environment.

IAM authentication

If you use IAM authentication, you must create an Redshift Role Amazon Resource Name (ARN), add the minimal Amazon IAM 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, the 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.

Temporary security credentials

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

- Before you run a task, ensure that you have enough time to use the temporary security credentials for running the task. You cannot extend the time duration of the temporary security credentials for an ongoing task. For example, when you read from and write to Amazon Redshift and if the temporary security credentials expire, you cannot extend the time duration of the temporary security credentials that causes the task to fail.

- After the temporary security credentials expire, AWS does not authorize the IAM users or IAM roles to access the resources using the credentials. You must request for new temporary security credentials before the previous temporary security credentials expire in a mapping.
- For mappings in advanced mode, the temporary security credentials do not expire even after the configured time in the **Temporary Credential Duration** advanced source and target property elapses.
- When you create an Amazon Redshift V2 connection with the IAM Role ARN and use the SSE-KMS encryption, you must specify AWS_IAM_ROLE as the unload option in the Amazon Redshift V2 advanced source properties.
- If both the source and target in a mapping point to the same Amazon S3 bucket, use the same Amazon S3 connection in the Source and Target transformations. If you use two different Amazon S3 connections, configure the same values in the connection properties for both the connections.
- If the source and target in a mapping point to different Amazon S3 buckets, you can use two different Amazon S3 connections.
  You can configure different values in the connection properties for both the connections. However, you must select the **Use EC2 Role to Assume Role** check box in the connection property. You must also specify the same value for the **Temporary Credential Duration** field in the source and target properties.

CDC sources

To create a mapping with a CDC source, ensure that you have the PowerExchangeClient and CDC licenses. 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.

Mappings in advanced mode

If you configure a mapping to run in advanced mode, ensure that the Redshift cluster and the advanced cluster reside in the same virtual private cloud (VPC).

Create an external schema and table for Amazon Redshift Spectrum

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, SequenceFIile, and Textfile file formats.

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

When you create an external table using Athena or Glue 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:

- Bigint (INT8)
- Boolean (BOOL)
- Char (CHARACTER)
- Date

  Note: Applicable when you create an external table for the ORC, Parquet, and Textfile file formats.
- Decimal (NUMERIC)
- Double Precision (FLOAT8)
- Integer (INT, INT4)
- Real (FLOAT4)
- Smallint (INT2)
- Timestamp
- Varchar (CHARACTER VARYING)

Rules and guidelines for external tables

Consider the following rules and guidelines for external tables:

- 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.
- The Secure Agent does not remove the external table names from the list of target objects available in the Target transformation.
- You cannot use pre-SQL and post-SQL commands to perform target operations on an external table.

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

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. Additionally, you can configure Lookups for mappings and mapping tasks.

Consider the following rules and guidelines when you configure Amazon Redshift V2 objects in mappings:

- If you use a simple filter in a mapping, you must specify the filter condition in the `YYYY-MM-DD HH24:MI:SS.MS` format. If you use an advanced filter, you must specify the filter condition in the `date_time_fix.f_timestamp < to_date('2012-05-24 09:13:57','YYYY-MM-DD HH24:MI:SS.MS')` format.
- When you define queries for sources or targets and the table names or the column names contain special characters, you must enclose the table names or column names in double quotes.
- To write special characters in the source column names to the target, you can configure a Target transformation in the mapping to create a new target at runtime. Enable the `Exact Source Field Names in Target` property in the `Create New at Runtime` window. You cannot view this checkbox if you select an existing object. This feature does not apply for mappings in advanced mode.
- If the Amazon Redshift table or schema name contains a forward slash (/), the mapping fails.
- When you select an Amazon Redshift V2 object that contains a boolean data type and preview the data, the value of the boolean data type appears truncated and displays only the first letter.

- If a field contains a Date or Datetime data type, the **Data Preview** tab displays the field value with an offset of +4 hours.
- When you read or write data using a mapping in advanced mode, the source table name must not contain unicode or special characters.

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. Select a source connection, or click **New Parameter** to define a new parameter for the source connection.
Source type	Type of the source object. Select any of the following source object: - Single Object - Multiple Objects. You can use implicit joins and advanced relationships with multiple objects. - Query. When you select the source type as query, you must map all the fields selected in the query in the **Field Mapping** tab. - Parameter **Note:** You cannot override the source query object and multiple objects at runtime using parameter files in a mapping. When you select the source type as query, the boolean values are written as 0 or false to the target. The query that you specify must not end with a semicolon (;).
Object	Name of the source object. You can select single or multiple source objects.
Parameter	Select an existing parameter for the source object or click **New Parameter** to define a new parameter for the source object. The **Parameter** property appears only if you select Parameter as the source type. If you want to overwrite the parameter at runtime, select the **Overwrite Parameter** option.
Filter	Filters records based on the filter condition. You can specify a simple filter or an advanced filter.
Sort	Sorts records based on the conditions you specify. You can specify the following sort conditions: - Not parameterized. Select the fields and type of sorting to use. - Parameterized. Use a parameter to specify the sort option.

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

Property	Description
Read Mode	Specifies the read mode to read data from the Amazon Redshift source. You can select one of the following read modes: - Direct[1]. Reads data directly from the Amazon Redshift source without staging the data in Amazon S3. - Staging. Reads data from the Amazon Redshift source by staging the data in the S3 bucket. Default is Staging.
Fetch Size[1]	Determines the number of rows to read in one resultant set from Amazon Redshift. Applies only when you select the **Direct** read mode. Default is 10000. **Note:** If you specify fetch size 0 or if you don't specify a fetch size, the entire data set is read directly at the same time than in batches.
S3 Bucket Name*	Amazon S3 bucket name for staging the data. You can also specify the bucket name with the folder path. If you provide an Amazon S3 bucket name that is in a different region than the Amazon Redshift cluster, you must configure the **REGION** attribute in the Unload command options.
Enable Compression*	Compresses the staging files into the Amazon S3 staging directory. The task performance improves when the Secure Agent compresses the staging files. Default is selected.
Staging Directory Location[1]*	Location of the local 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`. Ensure that you have the write permissions on the directory.
Unload Options*	Unload command options. Add options to the Unload command to extract data from Amazon Redshift and create staging files on Amazon S3. Provide an Amazon Redshift Role Amazon Resource Name (ARN). You can add the following options: - DELIMITER - ESCAPE - PARALLEL - NULL[1] - AWS_IAM_ROLE - REGION - ADDQUOTES For example: DELIMITER = \036;ESCAPE = OFF;NULL=text;PARALLEL = ON;AWS_IAM_ROLE=arn;aws;iam;;<account ID>;role/<role-name>;REGION = ap-south-1 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 uses the default values.
Treat NULL Value as NULL*	Retains the null values when you read data from Amazon Redshift.

Property	Description
Encryption Type*	Encrypts the data in the Amazon S3 staging directory. You can select the following encryption types: - None - SSE-S3 - SSE-KMS - CSE-SMK[1] Default is None.
Download S3 Files in Multiple Parts[1]*	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[1]*	The maximum threshold size to download an Amazon S3 object in multiple parts. Default is 5 MB.
Schema Name	Overrides the default schema name. **Note:** You cannot configure a custom query when you use the schema name.
Source Table Name	Overrides the default source table name. Ensure that the metadata and column order in the override table match those in the source table imported during design time. **Note:** When you select the source type as **Multiple Objects** or **Query**, you cannot use the **Source Table Name** option.
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 Secure Agent includes a `SELECT DISTINCT` statement if you choose this option. Amazon Redshift ignores trailing spaces. Therefore, the Secure 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.
SQL Query	Overrides the default SQL query. Enclose column names in double quotes. The SQL query is case sensitive. Specify an SQL statement supported by the Amazon Redshift database. When you specify the columns in the SQL query, ensure that the column name in the query matches the source column name in the mapping.
Temporary Credential Duration	The time duration during which an IAM user can use the dynamically generated temporarily credentials to access the AWS resource. Enter the time duration in seconds. Default is 900 seconds. If you require more than 900 seconds, you can set the time duration up to a maximum of 12 hours in the AWS console and then enter the same time duration in this property.
Tracing Level	Use the verbose tracing level to get the amount of detail that appears in the log for the Source transformation.

Property	Description
[1]Does not apply to mappings in advanced mode.	
*Does not apply to direct read mode.	

Rules and guidelines for reading data

You need to follow some rules and guidelines for the read operation.

Direct read mode

Consider the following rules and guidelines when you read directly from an Amazon Redshift source:

- The order of data written to the target in the direct read mode is different from the order of data written after Amazon S3 staging.
- The precision values are rounded-off to the 6th precision for real data types and to the 14th precision for double data types.
- NULL values in char and varchar data types are written to the target without quotes.
- Trailing spaces in the char data type columns appear with trailing spaces in the target. In comparison, the trailing spaces are truncated when you set the source with staging mode.

Filter using IS NULL or IS NOT NULL operators

When you convert a simple filter condition with the IS NULL or IS NOT NULL operators to an advanced filter, these operators are converted into SQL functions and the resultant query fails. To avoid this issue when you use these operators, you need to use the Amazon Redshift compatible syntax in the advanced filter condition.

Rules and guidelines for configuring SQL query

Consider the following rules and guidelines when you configure an SQL query:

- When you select the source type as Query, you can override a SQL query over a custom query by setting the flag `precedenceToSQLOverrideOverCustomQuery=True` in the JVM option. Ensure that the number of columns in the overridden SQL query is same as the custom query.
- When you run a mapping in advanced mode and define a SQL query with alldatatypes, the target columns with the boolean data type appear as NULL.
- When you select individual columns but not all the columns in an SQL query in a mapping in advanced mode, the values are written as NULL to the Redshift target.
- When you use a parameterized query source type in a mapping to read from multiple tables, and you configure a join for one or more tables that have the same column names, the mapping fails.
  For example, see the following SQL query that involves a full outer join between two tables EMPLOYEE and DEPARTMENT that are part of the SALES.PUBLIC schema, where two columns have the same name, CITY:

  ```
  SELECT EMP_ID, NAME, CITY, DEPT_ID, DEPT_NAME, CITY FROM SALES.PUBLIC.EMPLOYEE
  FULL OUTER JOIN SALES.PUBLIC.DEPARTMENT ON EMP_ID = DEPT_ID
  ```

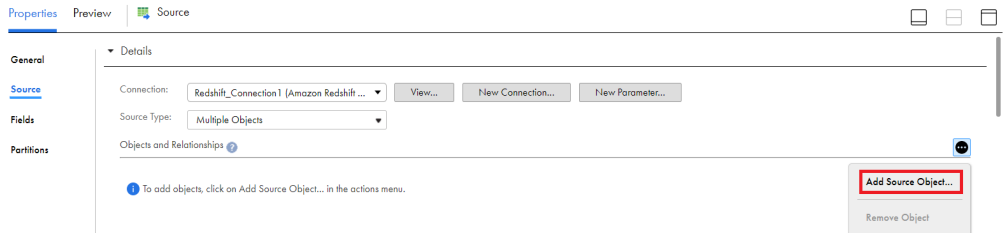
  To distinguish the conflicting column names, add aliases that the database can refer to while joining the tables:

  ```
  SELECT e.EMP_ID, e.NAME, e.CITY as ecity,d.DEPT_ID, d.DEPT_NAME, d.CITY as dcity FROM
  SALES.PUBLIC.EMPLOYEE e
  FULL OUTER JOIN SALES.PUBLIC.DEPARTMENT d ON e.EMP_ID = d.DEPT_ID
  ```

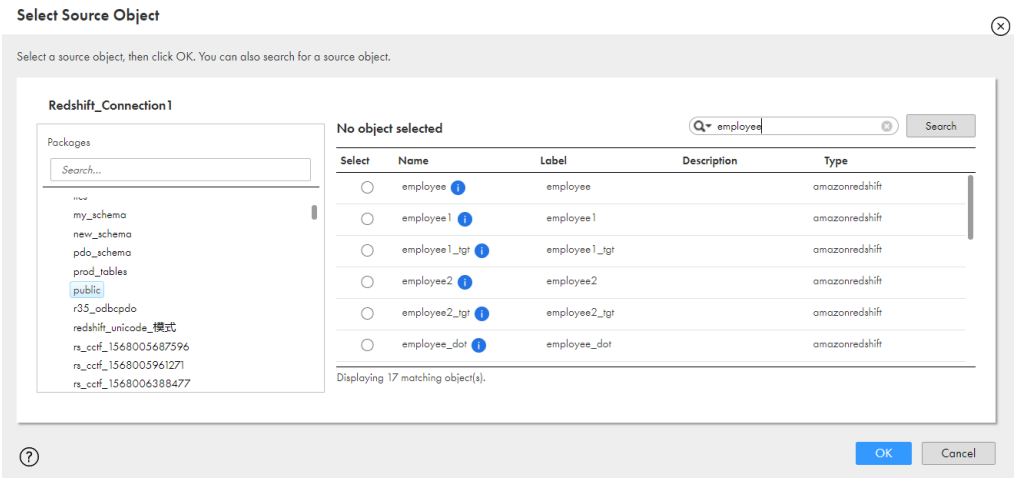
Adding multiple source objects

When you create a Source transformation, you can select Amazon Redshift V2 multiple object as the source type and then configure a join to combine the tables. You can define a relationship condition or a query to join the tables.

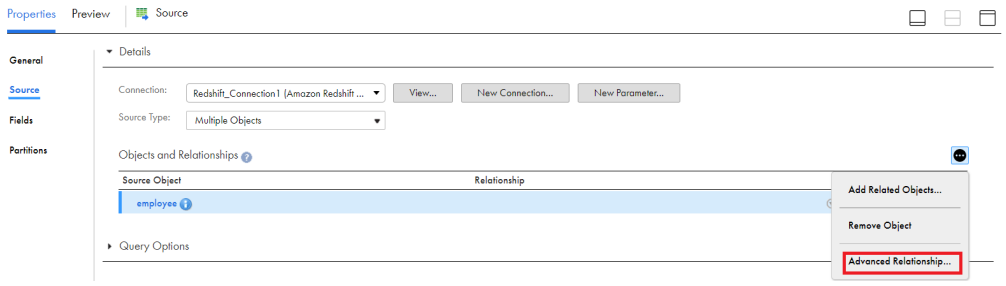
1. In the Source transformation, click the **Source Type** as **Multiple Objects**.
2. From the **Actions** menu, click **Add Source Object**.



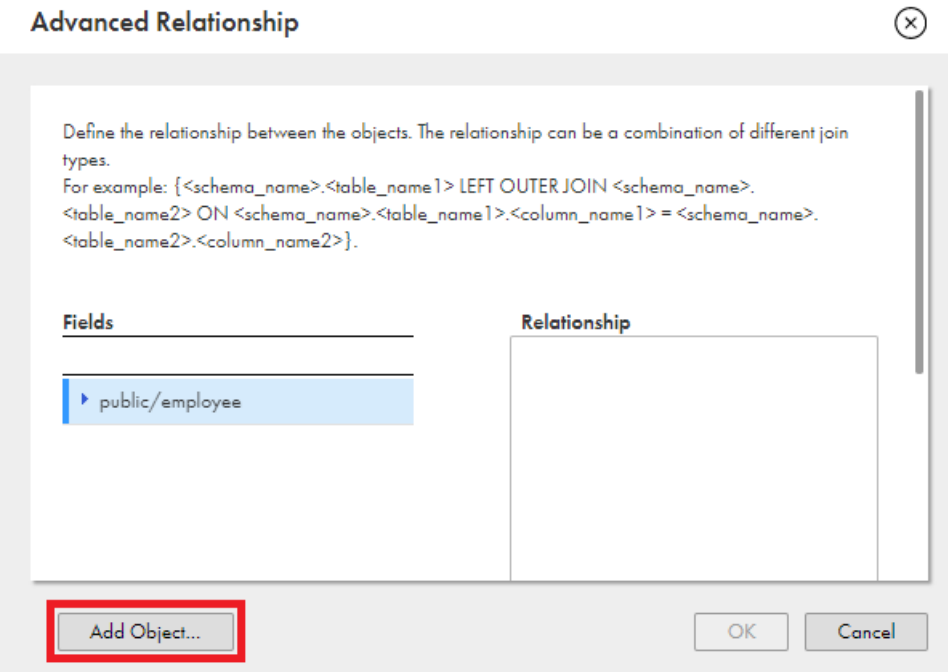
3. Select the source object that you want to add from the displayed list.



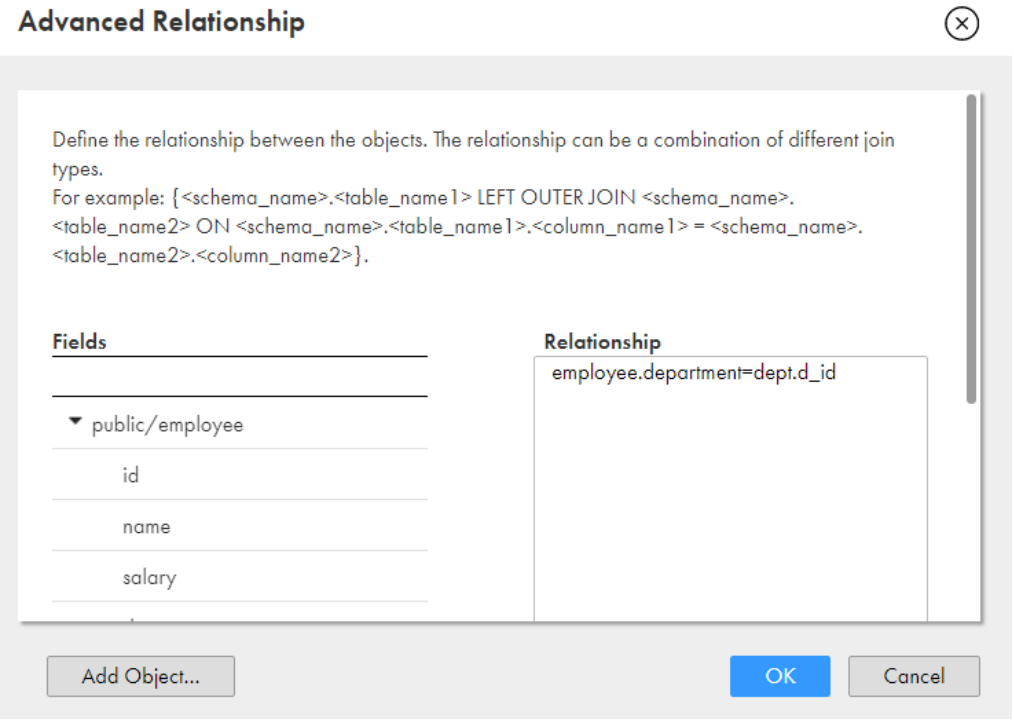
4. Click **OK**.
5. From the **Related Objects Actions** menu, select **Advanced Relationship**:



6. In the **Advanced Relationship** window, you can click **Add Object** to add more objects.



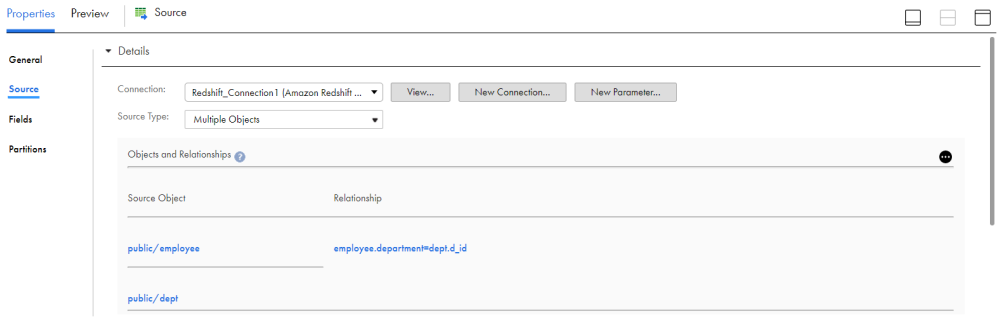
7. Click **OK**.
8. Set your own conditions or specify a query to define the relationship between the tables:



Note: When you configure a join expression, select the fields and define a join condition or a query syntax.

9. Click **OK**.

   The following image shows an example of an advanced join condition defined between the Amazon Redshift V2 tables:



Rules and guidelines for adding multiple source objects

Consider the following rules and guidelines when you add multiple source objects:

- You cannot configure partitioning when you use the advanced relationship option.
- You must specify double quotes for the table name when you use a reserved word for the table name and the column name.
- You cannot use a self join when you add multiple source objects.
- When you use special characters in column names for an advanced relationship, the query formed is not correct and the mapping task fails.
- When you click **Add Object** to add more objects in the **Advanced Relationship** window, the table might not load or take a lot of time to load. Import the object again.
- You cannot search for a schema when you add a related source object. You must scroll down and manually select the schema.
- You can use the full outer-join condition only with the `=`, `,`, and `AND` operators.
- When you use the advanced relationship option and specify the join queries with the schema name override, the mapping fails. The mapping fails with the following join queries:

  - inner join

  - left outer-join

  - right outer-join

  - full outer-join

  - cross join
- When you override the schema name and configure an advanced filter on a related source object, you must specify the schema name in the filter condition. Use the following syntax for the filter condition: `public.alldatatypes_src.f_smallint=1`
- When you override the schema name and configure an advanced filter on a related source object, the Secure Agent applies the advanced filter only on the parent object and not on the related source object.
- When you select parent and child objects that have a primary key and foreign key relationship, and the foreign key of the related object is also a primary key in the table, the mapping task fails when you create a target.
- When you select the **Multiple Objects** source type, add a source object, for example, `emp`, and define a primary key and foreign key relationship on different columns, for example, `emp.id` and `dept.d_id`, the mapping fails with the following error:

`[FATAL] Unload/Copy command failed with error: Invalid operation: column emp.d_id does not exist.`

The **Select Related Objects** list shows the join condition for the `dept` related object as `emp.d_id=dept.d_id`, even though the `emp` table does not have a `d_id` column.

- When you select the `Multiple Objects` source type, ensure that the selected table names do not contain a period (.).

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. Select a target connection, or click **New Parameter** to define a new parameter for the target connection.
Target Type	Type of the target object. Select Single Object or Parameter.
Object	Name of the target object. You can select an existing target object or create a new target object at runtime. To create a new target at runtime, select the Create New at Runtime option Creates a new target at runtime based on the table type and the path you specify. Specify the object name, table type, and path to create a new object at runtime.
Create New at Runtime	Applies to the Create New at Runtime option. Determines if you can create a target object at runtime with the same field names as the source.
Use Exact Source Field Names in Target	Applies to the Create New at Runtime option. This option does not appear for an existing target. Select to retain all the source field names in the target exactly as in the source, including any special characters. If you disable this option, special characters in the source are replaced with underscore characters in the target. Default is disabled.
Parameter	Select an existing parameter for the target object or click **New Parameter** to define a new parameter for the target object. The **Parameter** property appears only if you select Parameter as the target type. If you want to overwrite the parameter at runtime, select the **Overwrite Parameter** option.
Operation	Type of the target operation. Select one of the following operations: - Insert - Update - Upsert - Delete - Data Driven Select **Data Driven** if you want to create a mapping to capture changed data from a CDC source.

Property	Description
Data Driven Condition	Enables you to define expressions that flag rows for an insert, update, delete, or reject operation. You must specify the data driven condition for non-CDC sources. For CDC sources, you must leave the field empty as the rows in the CDC source tables are already marked with the operation types. **Note:** Appears only when you select **Data Driven** as the operation type.
Update Columns	Select columns you want to use as a logical primary key for performing update, upsert, and delete operations on the target. **Note:** This field is not required if the target table already has a primary key.
Create Target	Creates a new target. When you create a new target, enter a value of the following fields: - **Name**: Enter a name for the target object. - **Path**: Provide a schema name and create a target table within the schema. By default, the field is empty.

Note: The `Use Exact Source Field Names in Target` checkbox and `Data Driven` operation doesn't apply to mappings in advanced mode.

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

Property	Description
S3 Bucket Name	Amazon S3 bucket name for writing the files to Amazon Redshift target. You can also specify the bucket name with the folder path. If you provide an Amazon S3 bucket name that is in a different region than the Amazon Redshift cluster, you must configure the **REGION** attribute in the Copy command options.
Enable Compression	Compresses the staging files before writing the files to Amazon Redshift. The task performance improves when the Secure Agent compresses the staged files. Default is selected.
Staging Directory Location	Location of the local 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`. Ensure that you have the write permissions on the directory.
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 error rows that causes an upload insert batch to fail. Enter a positive integer. Default is 1. 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.
Truncate Target Table Before Data Load	Deletes all the existing data in the Amazon Redshift target table before loading new data.

Property	Description
Require Null Value For Char and Varchar	Replaces the string value with NULL when you write data to Amazon Redshift columns of Char and Varchar data types. Default is an empty string. **Note:** When you run a mapping to write null values to a table that contains a single column of the Int, Bigint, numeric, real, or double data type, the mapping fails. You must provide a value other than the default value in the **Require Null Value For Char And Varchar** property.
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	Copy command options. Add options to the Copy command to write data from Amazon S3 to the 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 - REGION For example: `DELIMITER = \036;ACCEPTINVCHARS = #;QUOTE = \037 COMPUPDATE = ON;AWS_IAM_ROLE=arn:aws:iam::<account ID>:role/<role-name>;REGION = ap-south-1` 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 uses the default values.
S3 Server Side Encryption	Indicates that Amazon S3 encrypts data during upload. 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 using a private key. Provide a master symmetric key ID in the connection property to enable this property. If you enable both server-side and client-side encryptions, the Secure Agent ignores the server-side encryption.
Analyze Target Table	Runs an ANALYZE command on the target table. 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 the 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.

Property	Description
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 default target operation. 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. For more information, see the Troubleshooting for Amazon Redshift V2 Connector topic. 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.
Override Target Query	Overrides the default update query that the Secure Agent generates for the update operation with the update query that you specify.
TransferManager Thread Pool Size	Number of threads to write data in parallel. Default is 10.
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.

Property	Description
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.
Preserve record order on write	Retains the order of the records when you read data from a CDC source and write data to an Amazon Redshift target. Use this property when you create a mapping to capture the changed record from a CDC source. This property enables you to avoid inconsistencies between the CDC source and target.
Minimum Upload Part Size	Minimum size of the Amazon Redshift object to upload an object. Default is 5 MB.
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.
Schema Name	Overrides the default schema name.
Target table name	Overwrites the default target table name. Ensure that the metadata and column order in the override table match those in the target table selected during design time.
Recovery Schema Name	Schema that contains recovery information stored in the `infa_recovery_table` table on the target system to resume the extraction of the changed data from the last checkpoint.
Temporary Credential Duration	The time duration during which an IAM user can use the dynamically generated temporarily credentials to access the AWS resource. Enter the time duration in seconds. Default is 900 seconds. If you require more than 900 seconds, you can set the time duration maximum up to 12 hours in the AWS console and then enter the same time duration in this property.
Forward Rejected Rows	This property is not applicable for Amazon Redshift V2 Connector.

The following properties don't apply to mappings in advanced mode:

- Staging Directory Location
- Batch Size
- WaitTime In Seconds For S3 File Consistency
- S3 Client Side Encryption
- Success File Directory
- Error File Directory
- TransferManager Thread Pool Size
- Preserve record order on write
- Minimum Upload Part Size
- Number of files per batch
- Recovery Schema Name

Rules and guidelines for creating a target

The Secure Agent converts the target table names that you specify in the **Create Target** field into lower case.

When you create a target, you can view and edit the metadata of the target object in the **Target Fields** tab. You can edit the data type, precision and define primary key of the columns in the target objects. To edit the metadata, click **Options** > **Edit Metadata** in the **Target Fields** tab.

Consider the following rules and guidelines when you create a target:

- When you create a target if the source table contains column of Text data type, the Secure Agent displays the following error message:

  ```
  Unsupported datatype - 'text' for column
  'LONGTXTAREA__C'
  ```

  You must edit the data type of the source column in the Source transformation.
- You cannot parameterize the target at runtime.
- When you create a target and specify a table name or column name in uppercase or mixed case alphabets, the Secure Agent converts the target table name or column name into lowercase alphabets and the mapping task runs successfully. However, the tomcat log shows the following error message:
  `An unexpected exception occured while fetching metadata:[The following object is not found].`
- When you create a target with Time data type, the new target will have the data type as Timestamp.
- When you run a mapping task with a vaccuum command using the pre-SQL or post-SQL queries, the mapping fails.
- When you write to a target and the target connection has **Auto Create DBuser** enabled, the new user cannot truncate a table created by an existing user.

Rules and guidelines to override the target query

You can specify a target query override to override the update query that the Secure Agent generates for the update operation.

Consider the following rules and guidelines when you use the override target query property for an Amazon Redshift target:

- Select **Update** for the **Treat source rows as** property in the advanced target properties.
- You cannot use the override target query for an upsert operation.
- Specify the override target query in the following format:

  ```
  UPDATE <Target Schema Name here>.<Target TABLE NAME here>
    SET <column1> = {{TU}}.<column1>,
        <column2> = {{TU}}.<column2>,
        <column> = {{TU}}.<column>
    FROM {{TU}}
    WHERE <Target Schema Name>.<Target TABLE NAME here> .<update column1> = {{TU}}.
  <update column1>
    AND <Target Schema Name>.<Target TABLE NAME here> .<update column2> = {{TU}}.
  <update column2>
    AND … <Target Schema Name>.<Target TABLE NAME here> .<update column> = {{TU}}.
  <update column>
  ```
- Column names for :TU must match the target table column names.
- The `WHERE` column is mandatory in the query.
- You cannot use left, right, or full outer joins in the `FROM` column of an override target query statement.
- All the column names in the query must be qualified names and the table name and schema name must be associated with the column names.

Note: For the `SET` columns, Amazon Redshift does not allow a column name with a table. SET columns must have column names only.

- You must specify the override target query with a valid SQL syntax because Amazon Redshift V2 Connector replaces :TU with a temporary table name and does not validate the update query.
- You cannot change the order of the column mappings using the override target query.
- You cannot specify multiple override target queries for an update operation.
- When you use the override target query option, do not additionally configure an override for the schema and table name from the **Schema Name** and **Target table name** fields.

Amazon Redshift V2 lookups in mappings

You can create lookups for objects in an Amazon Redshift V2 mapping. You can retrieve data from an Amazon Redshift V2 lookup object based on the specified lookup condition.

Use an Amazon Redshift V2 Lookup transformation to look up data in an Amazon Redshift object. For example, the source table includes the customer code, but you want to include the customer name in the target table to make summary data easy to read. You can use the Amazon Redshift V2 Lookup transformation to look up the customer name in another Amazon Redshift object.

You can add the following lookups to a mapping:

- Connected with cached
- Connected with uncached. Applicable only to mappings.
- Unconnected with cached
- Dynamic lookup cache. Applicable only to mappings.

Use the JDBC URL specified in the connection properties to create lookups.

The following table describes the Amazon Redshift V2 lookup object properties that you can configure in a Lookup transformation:

Property	Description
Connection	Name of the lookup connection. You can select an existing connection, create a new connection, or define parameter values for the lookup connection property. If you want to overwrite the lookup connection properties at runtime, select the **Allow parameter to be overridden at run time** option.
Source Type	Type of the source object. Select Single Object, Query, or Parameter. **Note:** You cannot configure uncached lookups when you select the source type as query.
Parameter	A parameter file where you define values that you want to update without having to edit the task. Select an existing parameter for the lookup object or click **New Parameter** to define a new parameter for the lookup object. The **Parameter** property appears only if you select parameter as the lookup type. If you want to overwrite the parameter at runtime, select the **Allow parameter to be overridden at run time** option. When the task runs, the Secure Agent uses the parameters from the file that you specify in the advanced session properties.
Lookup Object	Name of the lookup object for the mapping.

Property	Description
Multiple Matches	Behavior when the lookup condition returns multiple matches. You can return all rows, any row, the first row, the last row, or an error. You can select from the following options in the lookup object properties to determine the behavior: - Return first row - Return last row - Return any row - Return all rows - Report error
Filter	Not applicable
Sort	Not applicable

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

Property	Description
S3 Bucket Name[1]	Amazon S3 bucket name for staging the data. You can also specify the bucket name with the folder path. If you provide an Amazon S3 bucket name that is in a different region than the Amazon Redshift cluster, you must configure the **REGION** attribute in the Unload command options.
Enable Compression[1]	Compresses the staging files into the Amazon S3 staging directory. The task performance improves when the Secure Agent compresses the staging files. Default is selected.
Staging Directory Location[1]	Location of the local 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`. Ensure that you have the write permissions on the directory. Does not apply to mappings in advanced mode.
Unload Options[1]	Unload command options. Add options to the Unload command to extract data from Amazon Redshift and create staging files on Amazon S3. Provide an Amazon Redshift Role Amazon Resource Name (ARN). You can add the following options: - DELIMITER - ESCAPE - PARALLEL - NULL - AWS_IAM_ROLE - REGION - ADDQUOTES For example: DELIMITER = \036;ESCAPE = OFF;NULL=text;PARALLEL = ON;AWS_IAM_ROLE=arn;aws;iam;;<account ID>;role/<role-name>;REGION = ap-south-1 You cannot use the NULL option in a mapping in advanced mode. 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 uses the default values.

Property	Description
Treat NULL Value as NULL[1]	Retains the null values when you read data from Amazon Redshift.
Encryption Type[1]	Encrypts the data in the Amazon S3 staging directory. You can select the following encryption types: - None - SSE-S3 - SSE-KMS - CSE-SMK You can only use **SSE-S3** encryption in a mapping configured in advanced mode. Default is None.
Download S3 Files in Multiple Parts[1]	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. Does not apply to mapping in advanced mode.
Multipart Download Threshold Size[1]	The maximum threshold size to download an Amazon S3 object in multiple parts. Default is 5 MB. Does not apply to mapping in advanced mode.
Schema Name	Overrides the default schema name. **Note:** You cannot configure a custom query when you use the schema name.
Source Table Name	Overrides the default source table name. **Note:** When you select the source type as **Multiple Objects** or **Query**, you cannot use the **Source Table Name** option.
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[1]	Selects unique values. The Secure Agent includes a `SELECT DISTINCT` statement if you choose this option. Amazon Redshift ignores trailing spaces. Therefore, the Secure 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.
SQL Query[1]	Overrides the default SQL query. Enclose column names in double quotes. The SQL query is case sensitive. Specify an SQL statement supported by the Amazon Redshift database. When you specify the columns in the SQL query, ensure that the column name in the query matches the source column name in the mapping.

Property	Description
Lookup Data Filter	Limits the number of lookups that the mapping performs on the cache of the lookup source table based on the value you specify in the filter condition. This property is applicable when you select Single Object as the source type and enable lookup cache on the **Advanced** tab in the Lookup transformation properties. Maximum length is 32768 characters. For more information about this property, see *Transformations* in the Data Integration documentation.
Temporary Credential Duration[1]	The time duration during which an IAM user can use the dynamically generated temporarily credentials to access the AWS resource. Enter the time duration in seconds. Default is 900 seconds. If you require more than 900 seconds, you can set the time duration up to a maximum of 12 hours in the AWS console and then enter the same time duration in this property.
Tracing Level	Use the verbose tracing level to get the amount of detail that appears in the log for the Source transformation.
[1] Does not apply to uncached lookups.	

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

Unconnected Lookup transformation

You can configure an unconnected Lookup transformation for the Amazon Redshift source in a mapping. Use the Lookup transformation to retrieve data from Amazon Redshift based on a specified lookup condition.

An unconnected Lookup transformation is a Lookup transformation that is not connected to any source, target, or transformation in the pipeline.

An unconnected Lookup transformation receives input values from the result of a :LKP expression in another transformation. The Integration Service queries the lookup source based on the lookup ports and condition in the Lookup transformation and passes the returned value to the port that contains the :LKP expression. The :LKP expression can pass lookup results to an expression in another transformation.

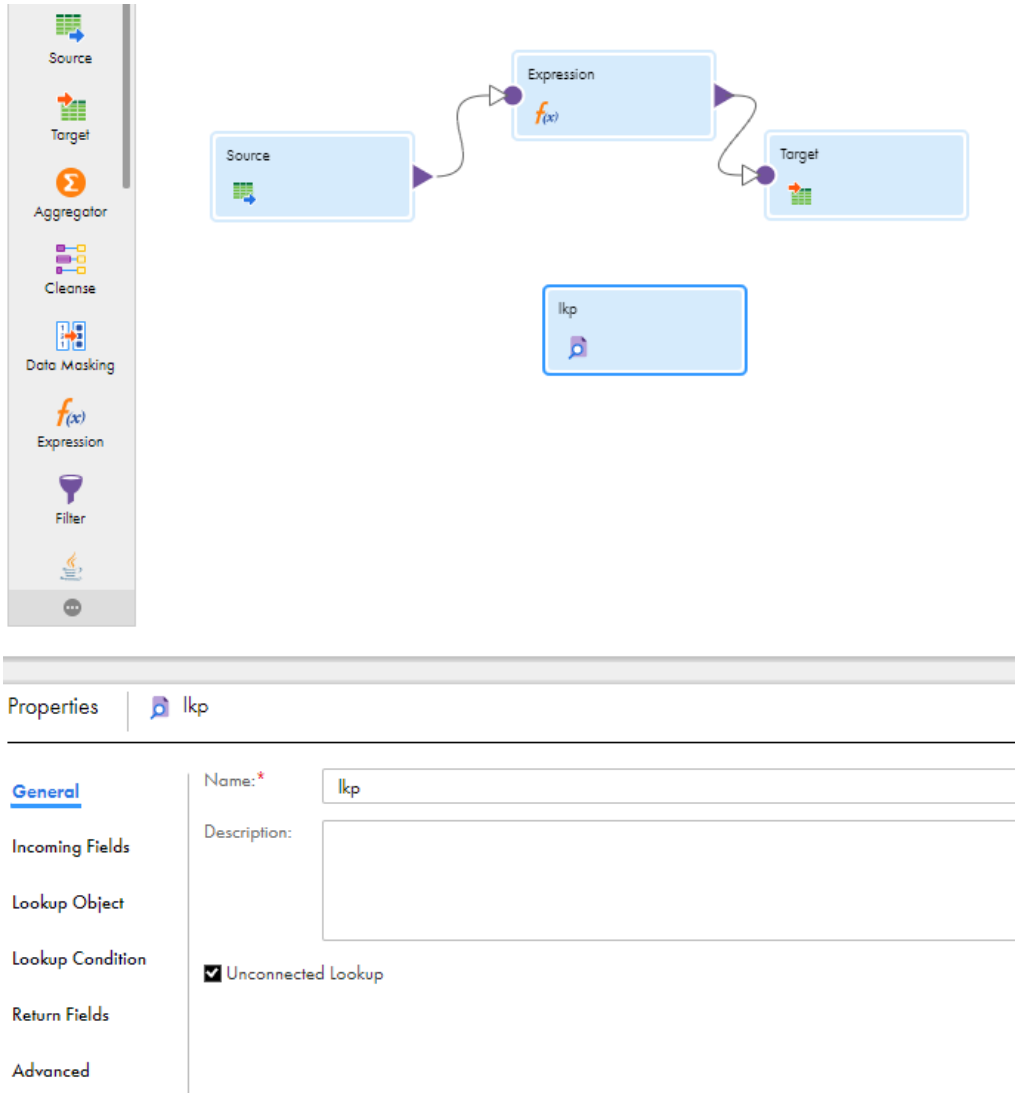
For more information about the Lookup transformation, see *Transformations*.

Configuring an unconnected Lookup transformation

To configure an unconnected Lookup transformation, select the **Unconnected Lookup** option, add incoming fields, configure the lookup condition, and designate a return value. Then configure a lookup expression in a different transformation.

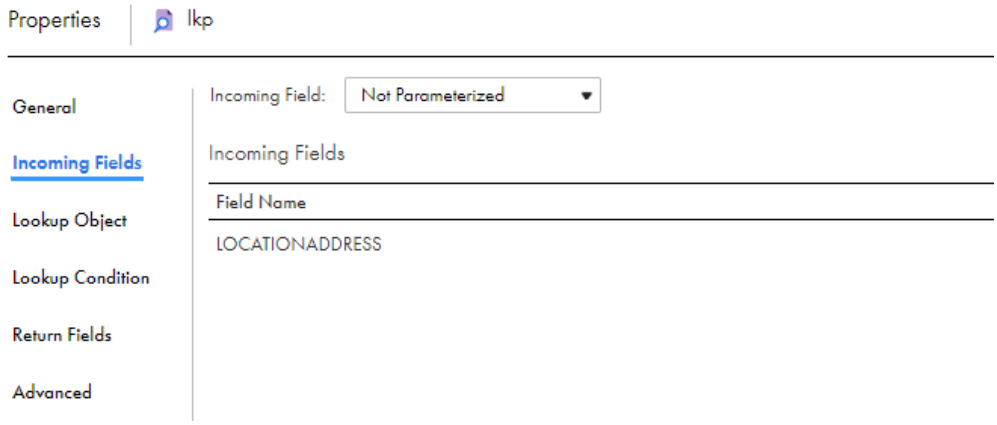
1. Add a Lookup transformation in a mapping.

2. On the **General** tab of the Lookup transformation, enable the **Unconnected Lookup** option.

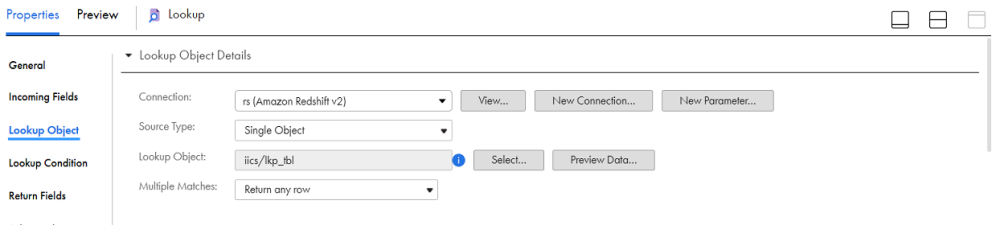


3. On the **Incoming Fields** tab of the Lookup transformation, create an incoming field for each argument in the :LKP expression.

For each lookup condition you create, add an incoming field to the Lookup transformation.



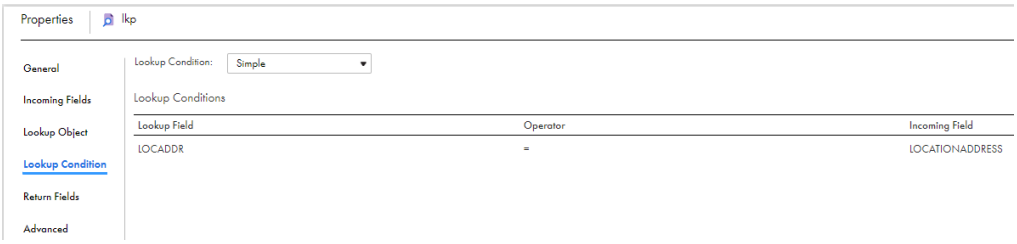
4. In the **Lookup Object** tab, import the lookup object.



The **Multiple Matches** property value **Return all rows** in an unconnected lookup is not applicable.

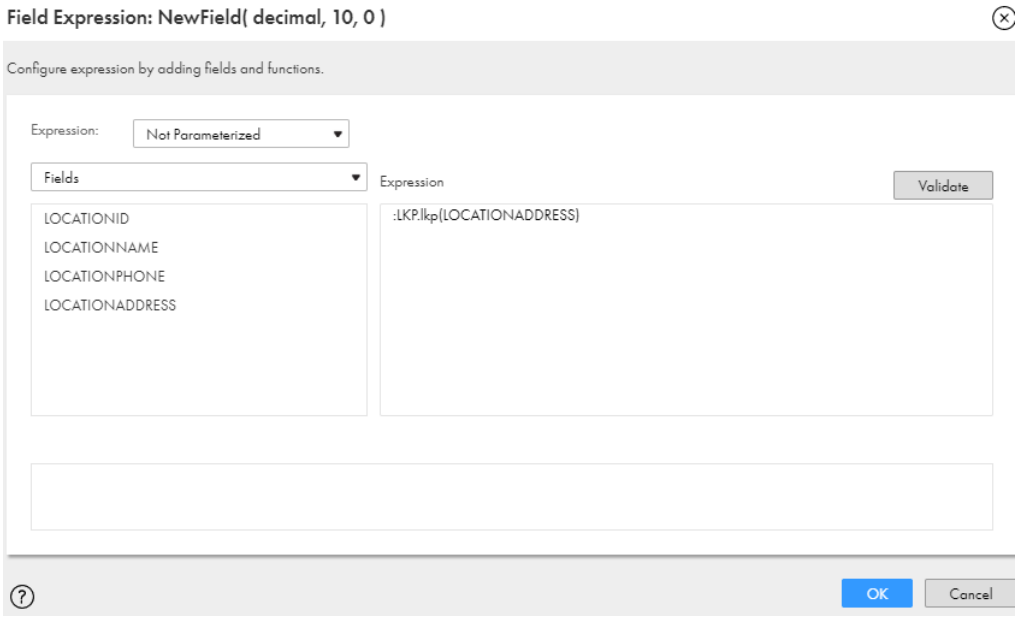
5. Designate a return value.

   You can pass multiple input values into a Lookup transformation and return one column of data. Data Integration can return one value from the lookup query. Use the return field to specify the return value.



6. Configure a lookup expression in another transformation.

Supply input values for an unconnected Lookup transformation from a :LKP expression in a transformation that uses an Expression transformation. The arguments are local input fields that match the Lookup transformation input fields used in the lookup condition.



7. Map the fields with the target.

Enabling lookup caching

When you configure a Lookup transformation in a mapping, you can cache the lookup data during the runtime session.

When you select **Lookup Caching Enabled**, Data Integration queries the lookup source once and caches the values for use during the session, which can improve performance. You can specify the directory to store the cached lookup.

Lookup Cache Persistent

Use lookup cache persistent to save the lookup cache file to reuse it the next time Data Integration processes a Lookup transformation configured to use the cache.

You can specify the file name prefix to use with persistent lookup cache files in the **Cache File Name Prefix** field.

If the lookup table changes occasionally, you can enable the **Re-cache from Lookup Source** property to rebuild the lookup cache.

Dynamic Lookup Cache

Use a dynamic lookup cache to keep the lookup cache synchronized with the target. By default, the dynamic lookup cache is disabled and represents static cache.

If the cache is static, the data in the lookup cache does not change as the mapping task runs.

If the task uses the cache multiple times, the task uses the same data. If the cache is dynamic, the task updates the cache based on the actions in the task, so if the task uses the lookup multiple times, downstream transformations can use the updated data.

For information about lookup caching, see *Transformations* in the Data Integration documentation.

Rules and guidelines for configuring lookup transformations

Consider the following rules and guidelines when you configure an Amazon Redshift lookup transformation:

- You cannot use unconnected lookups for mappings in advanced mode.
- You cannot configure a dynamic lookup cache for mappings in advanced mode.
- A Lookup transformation with the **On Multiple Matches** property configured as `Report Error`, runs successfully without displaying an error message for mappings in advanced mode.
- You cannot use the SQL query property for uncached lookups.
- To use the IS NULL and IS NOT NULL operators in an uncached lookup, you need to set the -DENABLE_NULL_FLAG_FOR_UNCACHED_LOOKUP=true property as a JVM option under the DTM type in the Secure Agent properties. You can use only the = and != operators. The task fails if you use any other operator.

Mapping task with Oracle CDC sources example

Your organization needs to replicate real-time changed data from a mission-critical Oracle production system to minimize intrusive, non-critical work, such as offline reporting or analytical operations system. You can use Amazon Redshift V2 Connector to capture changed data from the Oracle CDC source and write the changed data to an Amazon Redshift target table. Add the Oracle CDC sources in mappings, and then run the associated mapping tasks to write the changed data to the target.

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

   The **New Mapping** dialog box appears.
2. Enter a name and description for the mapping.
3. On the Source transformation, specify a name and description in the general properties.
4. On the **Source** tab, select the configured Oracle CDC connection and specify the required source properties.
5. On the Target transformation, specify a name and description in the general properties.
6. On the **Target** tab, perform the following steps to configure the target properties:
   a. In the **Connection** field, select the Amazon Redshift V2 connection.
   b. In the **Target Type** field, select the type of the target object.
   c. In the **Object** field, select the required target object.
   d. In the **Operation** field, select **Data Driven** to properly handle insert, update, and delete records from the source.
   e. In the **Data Driven Condition** field, leave the field empty.
   f. In the **Advanced Properties** section, provide the values of the required target properties. You must select the **Preserve record order on write** check box and enter the value of the **Recovery Schema Name** property.
7. On the **Field Mapping** tab, map the incoming fields to the target fields. You can manually map an incoming field to a target field or automatically map fields based on the field names.
8. In the **Actions** menu, click **New Mapping Task**.

   The **New Mapping Task** page appears.
9. On the **Definition** tab, enter the task name and select the configured mapping.
10. On the **CDC Runtime** tab, specify the required properties.

    For more information about the **CDC Runtime** properties, see the help for Oracle CDC Connector.

11. On the **Runtime Options** tab, specify the following properties in the **Advanced Session Properties** section:
    a. Select **Commit on End of File** from the menu, and keep the property disabled.
    b. Select the **Commit Type** field as **Source**.
    c. Select **Recovery Strategy** and set **Resume from last checkpoint** as the value of the property.
12. Click **Save** > **Run** the mapping.

    Alternatively, you can create a schedule that runs the mapping task on a recurring basis without manual intervention. You can define the schedule to minimize the time between mapping task runs.

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

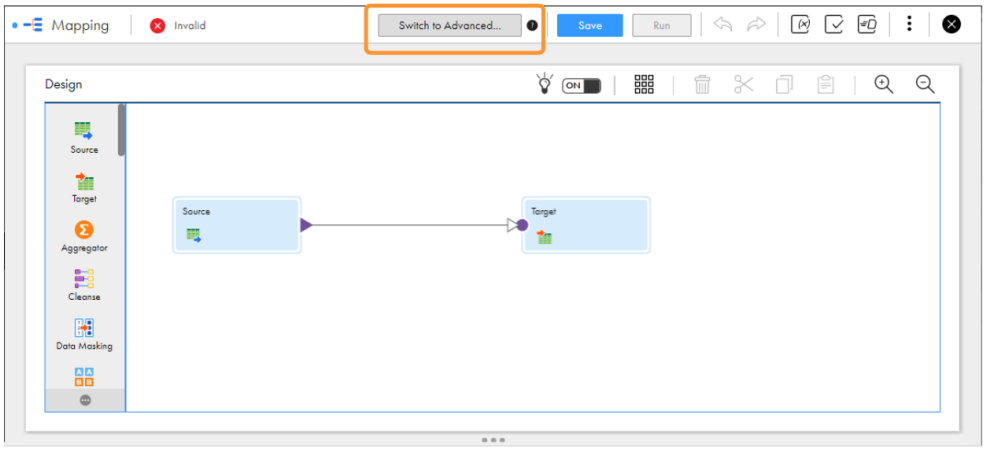
Mapping in advanced mode example

You work for an organization that stores large amount of purchase order details, such as customer ID, item codes, and item quantity in Amazon S3. You need to port the data from Amazon S3 to another cloud-based environment to quickly analyze the purchase order details and to increase future revenues.

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

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

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



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

   The Mapping Designer updates the mapping canvas to display the transformations and functions that are available in advanced mode.
4. Enter a name, location, and description for the mapping.
5. Add a Source transformation, and specify a name and description in the general properties.
6. On the **Source** tab, perform the following steps to provide the source details to read data from the Amazon S3 source:
    a. In the **Connection** field, select the Amazon S3 V2 source connection.

b. In the **Source Type** field, select the type of the source.

c. In the **Object** field, select the required object.

d. In the **Advanced Properties** section, provide the appropriate values.

7. On the **Fields** tab, map the Amazon S3 source fields to the target fields.
8. On the Target transformation, specify a name and description in the general properties.
9. On the **Target** tab, perform the following steps to provide the target details to write data to the Amazon Redshift target:

   a. In the **Connection** field, select the Amazon Redshift V2 target connection.

   b. In the **Target Type** field, select the type of the target.

   c. In the **Object** field, select the required object.

   d. In the **Operation** field, select the required operation.

   e. In the **Advanced Properties** section, provide appropriate values for the advanced target properties.
10. Map the Amazon S3 source and the Amazon Redshift target.
11. Click **Save** > **Run** to validate the mapping.

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

SQL transformation

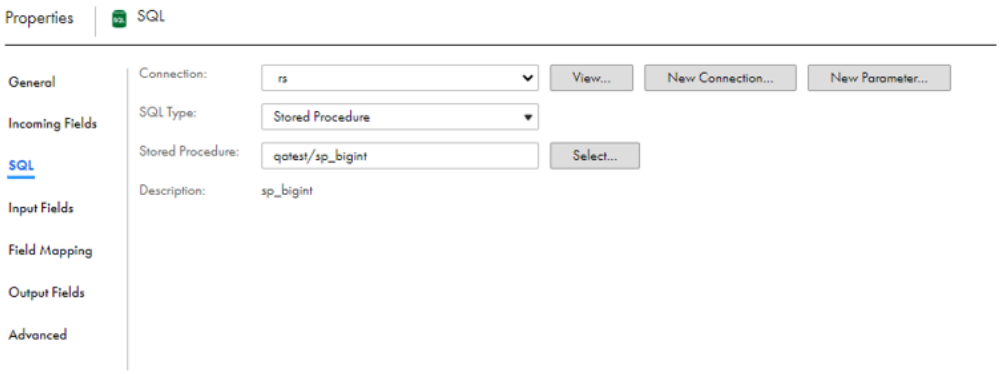
You can configure an SQL transformation to process SQL queries midstream in an Amazon Redshift V2 mapping. You cannot configure an SQL transformation for a mapping in advanced mode.

When you add an SQL transformation to the mapping, on the SQL tab, you define the connection and the type of SQL that the transformation processes.

Stored procedure

You can configure an SQL transformation to call a stored procedure in Amazon Redshift. The stored procedure must exist in the Amazon Redshift database before you create the SQL transformation. When the SQL transformation processes a stored procedure, it passes input parameters to the stored procedure. The stored procedure passes the return value to the output fields of the transformation.

The following image shows a stored procedure in a SQL transformation:



Consider the following rules and guidelines for stored procedures:

- When you use the same column names in the source table as in the table used in the stored procedure, the mapping task fails.
- When you specify the numeric data type, for example, `numeric(10,2)` for a column in a stored procedure, the mapping task fails.
- When you create a stored procedure in a mapping, the output port must not include the refcursor data type.
- You cannot read from multiple stored procedures that have the same name even if the input parameters that you specify for the stored procedures are different.

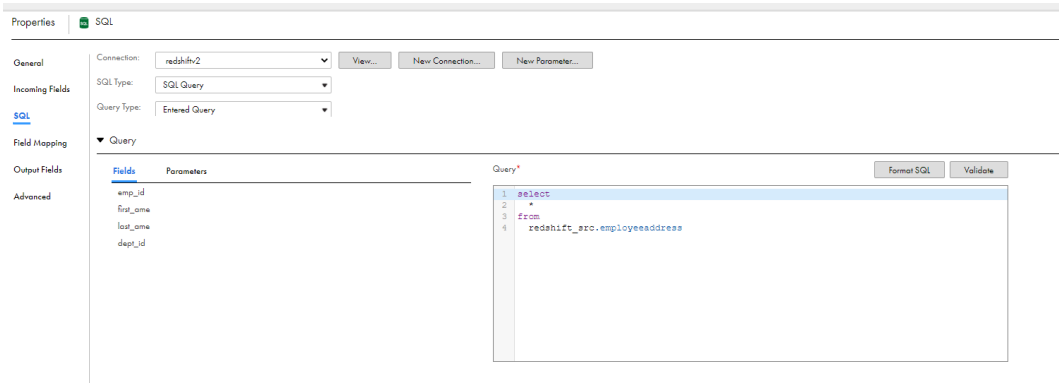
For more information about stored procedures, see *Transformations* in the Data Integration documentation.

SQL query

You can configure an SQL transformation to process an entered query that you define in the SQL editor. The SQL transformation processes the query and returns the rows. The SQL transformation also returns any errors that occur from the underlying database or the user syntax.

You can process an entered query to call a stored procedure that returns a value or no values.

The following image shows an entered query in a SQL transformation:



Consider the following rules and guidelines for SQL queries:

- You cannot use the saved query type in an SQL transformation.
- You can specify only a single query in an entered query. Do not specify multiple queries.
- Do not use the COMMIT, ROLLBACK, and EXPLAIN PLAN statements in an entered query.

For more information about SQL queries, see *Transformations* in the Data Integration documentation.

Dynamic schema handling for mappings

You can choose how Data Integration handles changes that you make to the data object schemas. To refresh the schema every time the mapping task runs, you can enable dynamic schema handling in the task.

A schema change includes one or more of the following changes to the data object:

- Fields added.
- Fields updated for data type, precision, or scale.

Configure schema change handling on the **Runtime Options** page when you configure the task.

The following table describes the schema change handling options:

Option	Description
Asynchronous	Default. Data Integration refreshes the schema when you edit the mapping or mapping task, and when Informatica Intelligent Cloud Services is upgraded.
Dynamic	Data Integration refreshes the schema every time the task runs. You can choose from the following options to refresh the schema: - **Alter and Apply Changes.** Data Integration alters the target schema and adds the new fields from the source. - **Drop Current and Recreate.** Drops the existing target table and then recreates the target table at runtime using all the incoming metadata fields from the source. - **Don't Apply DDL Changes.** Data Integration does not apply the schema changes to the target.

For more information, see the "Schema change handling" topic in *Tasks* in the Data Integration help.

Rules and guidelines for dynamic schema handling for mappings

Consider the following rules and guidelines when you enable dynamic schema change handling:

- Updates to the data types, precision, and scale are not applicable.
- Schema updates that involve a decrease in the precision of the Varchar data type are not applicable.
- When you use special characters in a target table column and select **Alter and Apply Changes**, the mapping task fails with an error.
- When you use a completely parameterized filter in a field mapping in a Source transformation and select the **Dynamic** option, the **Runtime Options** page does not display the options to refresh the schema.
- When you use an SQL override in a Source transformation and select **Alter and Apply Changes**, the mapping task fails with an error.
- When you add a field in the source as a primary key and select **Alter and Apply Changes**, the mapping task runs successfully and the field is propagated to target. However, the primary key is not considered.
- When you override the source table name or schema name and select the **Alter and Apply Changes** schema change handling option, and if new fields are added to the overridden source table, the mapping runs successfully. However, the Secure Agent does not fetch the newly added fields from the overridden source table.

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**.

Bulk processing for write operations

You can enable bulk processing to write large amounts of data to Amazon Redshift. Bulk processing utilizes minimal number of API calls and enhances performance of the write operation.

To enable bulk processing, specify the property *-DENABLE_WRITER_BULK_PROCESSING=true* in the Secure Agent properties:

Perform the following steps to configure bulk processing before you run a mapping:

1. In Administrator, select the Secure Agent listed on the **Runtime Environments** tab.
2. Click **Edit**.
3. In the **System Configuration Details** section, select **Data Integration Service** as the service and **DTM** as the type.
4. Edit the **JVMOption1** property, and enter *-DENABLE_WRITER_BULK_PROCESSING=true*.
5. Click **Save**.

Note: Not applicable to mapping tasks configured with SQL ELT optimization.

Optimize the staging performance for a mapping

Data Integration, by default, creates a flat file locally in a temporary folder to stage the data before reading from and writing to Amazon Redshift. You can set Data Integration to optimize the staging performance.

If you do not set the optimization property, Data Integration performs staging without the optimized settings, which might impact the performance of the task.

Consider the following rules when you enable the optimization property:

- If you run a mapping enabled for SQL ELT optimization, the mapping runs without SQL ELT optimization.
- For read operation, the staging applies only to source transformation and does not apply to lookup transformation.
- If the data contains timestamp data types with time zone, the job runs without staging the data in the local flat file.
- If the mapping contains Oracle CDC as a source and Amazon Redshift as the target, the job runs without staging the data in the local flat file.
- When you run a mapping to read multiline data from the source that contains CR or LF characters, the column data is split into multiple lines.

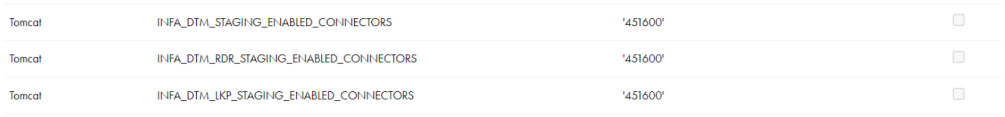
Enabling Amazon Redshift Connector to optimize the staging performance

Perform the following steps to enable the Amazon Redshift V2 Connector to optimize the staging performance:

1. In Administrator, click **Runtime Environments**.
2. Edit the Secure Agent for which you want to set the property.
3. In the **System Configuration Details** section, select the **Service** as **Data Integration Server** and the type as **Tomcat**.

4. Set the value of the Tomcat property *INFA_DTM_RDR_STAGING_ENABLED_CONNECTORS* for the read operation, *INFA_DTM_STAGING_ENABLED_CONNECTORS* for the write operation, and *INFA_DTM_LKP_STAGING_ENABLED_CONNECTORS* for the cached lookup operation to the plugin ID of Amazon Redshift Connector.
   You can find the plugin ID in the manifest file located in the following directory: `<Secure Agent installation directory>/downloads/<AmazonRedshift package>/CCIManifest`

   The following image shows the optimization property that you set for staging data in the DTM of the Secure Agent:



When you run the mapping, the flat file is created in the following directory in your machine: `C:\Windows\Temp\AmazonRedshift\stage\<AmazonRedshift_Target.txt>`

You can check the session logs to verify if the optimization property is enabled.

CHAPTER 8

Migrating a mapping

If you configure a connection and mapping in one environment and then migrate and run the mapping in another environment, Data Integration uses the configured runtime attributes from the earlier environment and runs the mapping successfully in the new environment.

After the migration, you can change the connection properties from the Administrator service, but you do not need to modify the mapping.

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

Use the same object path for the migrated mapping

If you want the migrated mapping in Org 2 to use the same object path as in Org 1, you must maintain the same schema and table in the Amazon Redshift account for Org 2.

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

Account1: SCHEMA1/TABLE1

Account2: SCHEMA1/TABLE1

In this scenario, you do not need to override the schema and table in the advanced properties.

Use a different object path for the migrated mapping

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

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

As a rule, when you specify the schema and table in the advanced properties or object properties, Data Integration honors the attributes in the following order of precedence:

1. **Runtime advanced attributes**. The advanced properties such as schema name and table name in the Source, Target, or Lookup transformation in a mapping.
2. **Object metadata**. The object type selected in the Source, Target, or Lookup transformation in a mapping.

Migration options

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

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

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

Option 2. Override the properties from the advanced properties

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

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

Option 3. Parameterize the properties in the mapping

You can choose to parameterize the advanced attributes, such as the schema and table name before the migration. You can configure input parameters, in-out parameters, and parameter files in the mapping.

After you migrate the mapping, do not edit or update the mapping. If you have used in-out parameters, you can change the schema and table attributes using the parameter file so that the changes are applied when the task runs.

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

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

Parameterizing both the object and the advanced properties

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

Parameterization rules

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

- Parameterization is not applicable for mappings that use the **Create Target** option.
- For the migration use case, input parameters and in-out parameters are applicable for mappings in advanced mode.
- If there are multiple pipelines configured in a mapping, do not parameterize the Amazon Redshift object. You must select a placeholder object while creating the mapping before you migrate.

Rules and guidelines for migrating a mapping

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

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

Transformations	Object Type	Advanced Fields
Source	Single object, multiple objects	Schema and table name
Lookup	Single object **Note:** Applicable for unconnected, and connected cached and uncached.	
Target	Single object	

- Before you migrate a mapping to Org2, map the connection1 of Org1 to Org2 that can access the schema and table name configured in the advanced properties in Org1.
- After you migrate the mapping to Org2, you must not edit the mapping.
- When you override the schema and table name in a mapping, the mapping fails with an error if the schema used in the connection in Org2 is not valid. The mapping fails with the following error:
  `An unexpected exception occured while fetching metadata:[The following schema is not found - invalidschema]`
- You cannot dynamically refresh the data object schema at runtime. You must maintain the same metadata for the table selected in the source, target, or lookup transformations and the corresponding advanced field overrides as schema change handling is not applicable.
- Consider the following guidelines for multiple sources:
  - If the mapping contains multiple sources, you can override the schema in the advanced properties only if the source objects have the same schema, as in the following example:
    object1: `<schema1.object1>`

    object2: `<schema1.object2>`

    If the schema is different, for example, if the schema for object1 is `<schema1.object1>` and object2 is `<schema2.object2>`, migration does not work.
  - Do not override the table from the advanced properties if the mapping contains multiple objects.

CHAPTER 9

Upgrade the connection type

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

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

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

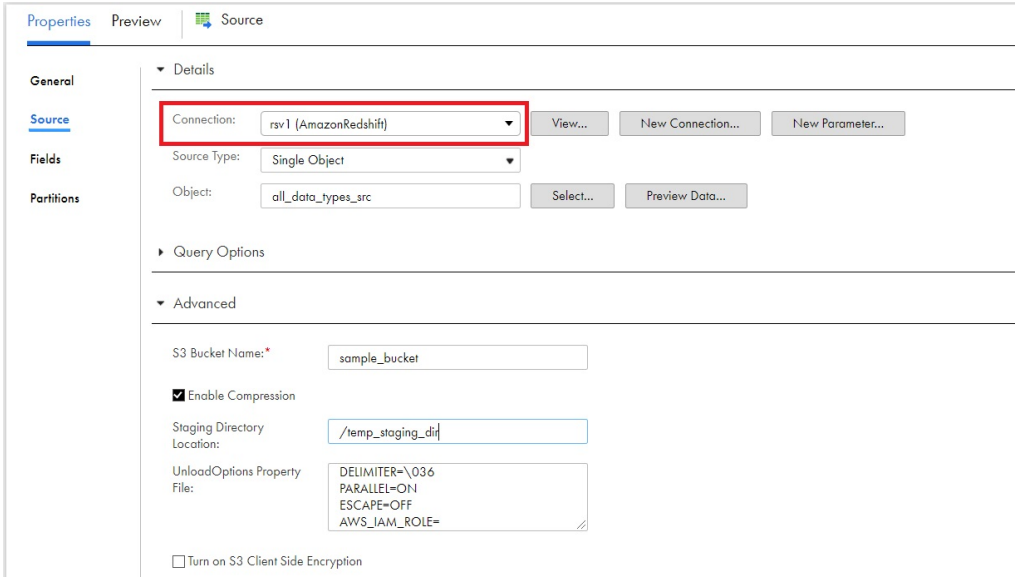
Note: If you are using the Amazon Redshift V1 connection in mappings to read from or write data to Amazon Redshift, Informatica recommends you to use the Amazon Redshift V2 connection to make use of the features that the enhanced connector offers. To get the license for Amazon Redshift V2 Connector,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

Connection switching example

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

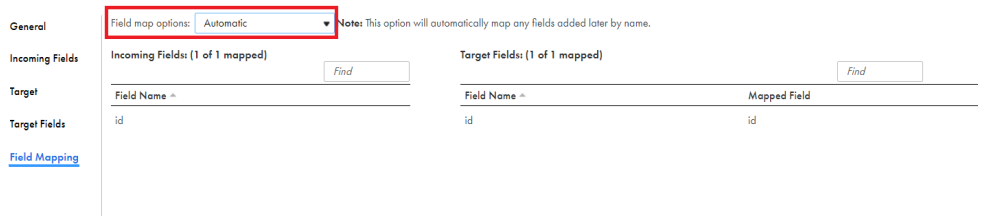
1. Open the existing Amazon Redshift V1 mapping that you want to upgrade to Amazon Redshift V2.

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



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

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

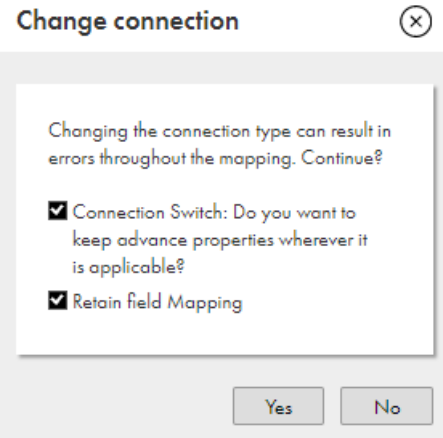


   - To manually map the retained fields after the switch, select **Manual**.

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

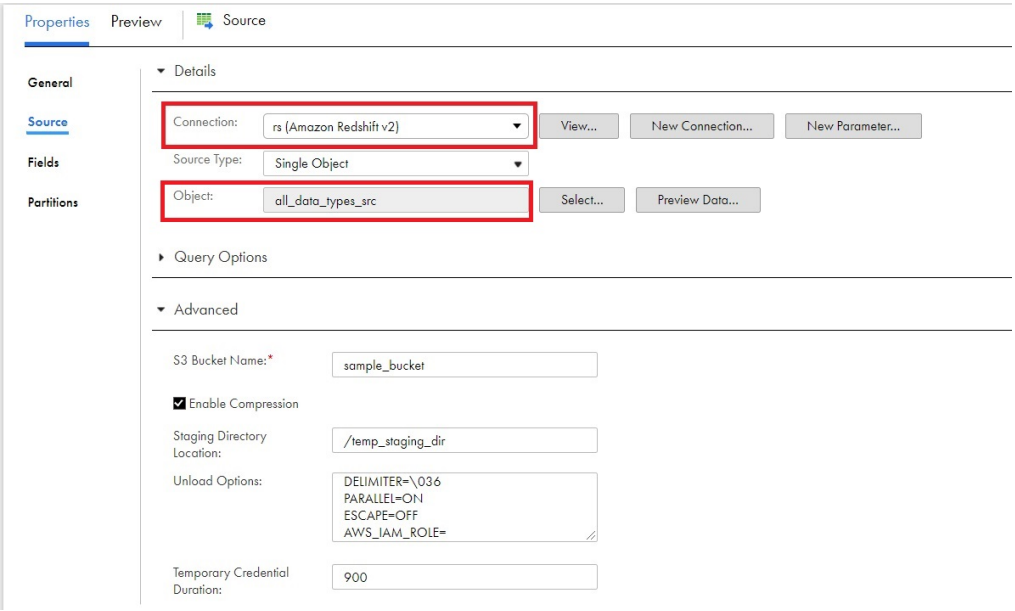
3. To switch the connection, in the **Connection** field, change the connection from Amazon Redshift V1 to Amazon Redshift V2.
4. In the **Change Connection** dialog box, select **Connection Switch** property, and click **Yes**:
   - **Connection switch**. Switches to the connection that you select.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

The following image shows the option that you must select:



5. Use the same object path in the mapping as Amazon Redshift V1.

   The following image shows the switched connection with the same object path as Amazon Redshift V1:



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

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



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

7. Click **Save**.

Advanced properties retained after the switch

When you replace the source or target connection type in existing mappings or mapping tasks that use the Amazon Redshift V1 connection or the Amazon Redshift ODBC connection with the Amazon S3 V2 connection, you have the option to retain the configured advanced properties.

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

Properties	Amazon Redshift ODBC properties retained in Amazon Redshift V2
Source	- Pre SQL - Post SQL - SQL Override
Target	- Pre SQL - Post SQL

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

Properties	Amazon Redshift V1 properties retained in Amazon Redshift V2
Source and Lookup	- S3 Bucket Name - Enable Compression - Staging Directory Location - Unload Options - Encryption Type - Download S3 Files In Multiple Parts - Source Table Name - Pre-SQL - Post-SQL - SQL-Query - Select Distinct
Target	- S3 Bucket Name - Enable Compression - Staging Directory Location - S3 Client Side Encryption - S3 Server Side Encryption - Pre-SQL - Post-SQL - Batch Size - Max Errors Per Batch Upload Insert - Truncate Target Table Before Data Upload - Require Null Value For Char And Varchar - Wait Time In Seconds For S3 File Consistency - Copy Options - Vacuum Target Table - Analyze Target Table - Prefix to retain staging files on S3 - Target Table Name - Minimum Upload Part Size - Success File Directory - Error File Directory - Transfer Manager Thread Pool Size

Rules and guidelines

Consider the following rules before you replace the Amazon Redshift ODBC connection or Amazon Redshift V1 connection in an existing mapping with the Amazon Redshift V2 connection:

- When you specify the pre-SQL, post-SQL, and SQL query commands in the source and target properties for Amazon Redshift and upgrade the connection to Amazon Redshift V2, the queries are truncated. The issue occurs when the number of characters in the queries in the Amazon Redshift ODBC connection or Amazon Redshift V1 connection exceeds 65535.
- Even after you replace the connection with the Amazon Redshift V2 connection, the object field displays the object that you selected in the earlier connection. You need to reimport the Amazon Redshift object and remap the fields in the mapping.

CHAPTER 10

SQL ELT with Amazon Redshift V2 Connector

You can enhance the mapping performance with SQL ELT.

You can use SQL ELT to read data from a cloud data warehouse and write to the same cloud data warehouse. You can also read data from a data lake in your cloud ecosystem and write it to a cloud data warehouse in the same ecosystem. Data Integration translates the transformation logic into ecosystem-specific commands and SQL statements that run in the underlying cloud infrastructure. This increases the data processing speed because the data isn't moved out of the cloud infrastructure for processing.

Example

You work in the sales and inventory department of a retail organization that operates multiple stores and an online platform. The management wants to optimize its inventory management, understand customer behavior, and improve sales strategies by analyzing data from various sources.

The organization uses Amazon S3 to store the data from multiple sources. To analyze the data, the organization plans to leverage the warehouse infrastructure of Amazon Redshift and load all its data to Amazon Redshift.

You can use SQL ELT to quickly load data from Amazon S3 to Amazon Redshift. SQL ELT is faster than the ETL method and eliminates any data egress costs as the data remains within your cloud infrastructure.

SQL ELT configuration options

You can either use the Mapping - SQL ELT mapping type to directly create a mapping in SQL ELT mode, or you can first create a mapping and then enable SQL ELT optimization from the mapping task properties.

Create a mapping in SQL ELT mode

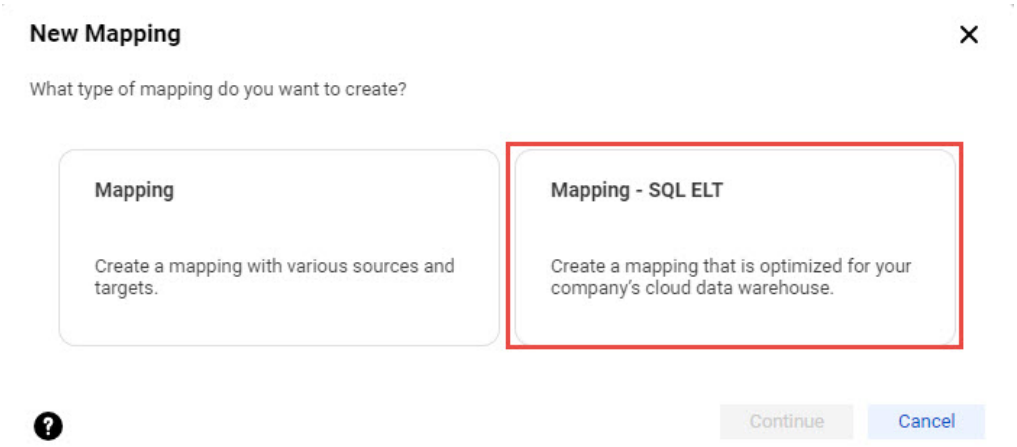
In a mapping in SQL ELT mode, you don't have to specify an optimization type. Data Integration pushes the transformation logic by default to the target database.

When you run a mapping in SQL ELT mode, you can load data from the following data sources to Amazon Redshift:

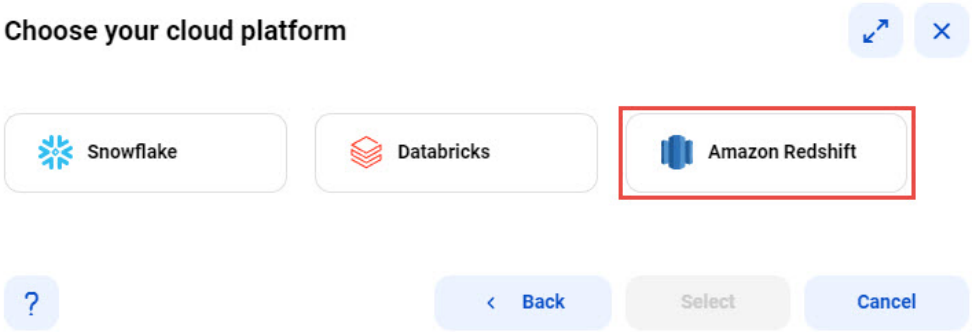
- Amazon Redshift

- Amazon S3

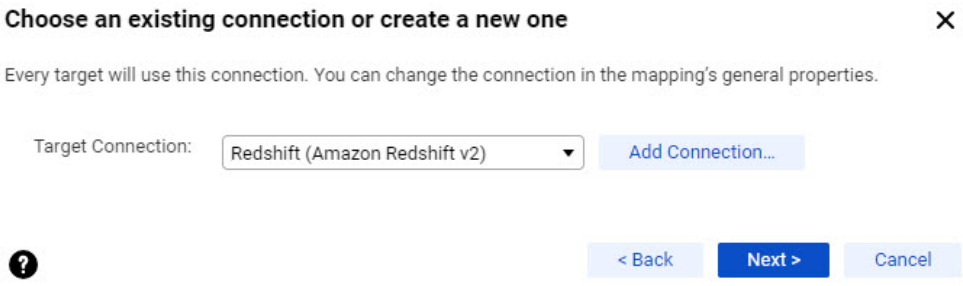
1. To create a mapping in SQL ELT mode, you create a mapping and select **Mapping - SQL ELT** as the mapping type.



2. Select the cloud ecosystem where you want all the mapping logic to be processed.



3. You're then prompted to select the target connection.



   The mapping is automatically configured with the target connection you chose.

   You can add additional transformations to the mapping. The transformations available in the transformation palette are transformations that the target endpoint can process. You can also preview data for individual transformations to validate the mapping logic.

For more information on mappings in SQL ELT mode, see Mappings in SQL ELT mode for Amazon Redshift on page 99.

Enable SQL ELT optimization for a mapping task

In a task enabled with SQL ELT optimization, you can select the optimization type to push some or all the transformation logic. Data Integration pushes the transformation logic to the source or target database based on the optimization type you specify in the task properties.

To enable SQL ELT optimization, add the configured mapping or mapping in advanced mode to a mapping task, and then enable SQL ELT optimization in the mapping task.

You can configure Full, Source, and other optimization configurations in the **SQL ELT optimization** advanced session properties on the **Runtime Options** tab. Data Integration processes any transformation logic that isn't pushed to the sources and targets on the on the Data Integration Server and on an advanced cluster.

When you create mappings and enable SQL ELT optimization in a mapping task, you can load data from the following data sources to Amazon Redshift:

- Amazon Redshift
- Amazon S3

For more information on SQL ELT optimization, see SQL ELT optimization for a mapping taskon page 102.

SQL ELT query preview

Before you run a mapping that is configured for SQL ELT optimization or a mapping in SQL ELT mode, you can preview SQL queries in the SQL ELT Query panel in the Mapping Designer.

After you configure a mapping and run the preview, Data Integration creates and runs a temporary SQL ELT preview mapping task. When the job completes, Data Integration displays the SQL queries to be executed and any warnings in the **SQL ELT Query** panel. The warning messages help you understand which transformations in the configured mapping are not applicable for SQL ELT optimization. If SQL ELT query preview fails, Data Integration lists any queries generated up to the point of failure. You can edit the mapping and fix the required transformations before you run the mapping for SQL ELT optimization.

You can also view the temporary job created under **My Jobs** and download the session log to view the queries generated.

For more information about how to preview SQL ELT query, see "SQL ELT query preview" in *Mappings* in the Data Integration documentation.

Mappings in SQL ELT mode for Amazon Redshift

You can create mappings in SQL ELT mode to read data from Amazon Redshift or Amazon S3, load it to Amazon Redshift, and perform all of the data transformation within Amazon Redshift.

To create a mapping in SQL ELT mode, you create a mapping, and then select **Mapping - SQL ELT** as the mapping type. You're then prompted to choose Amazon Redshift as the cloud ecosystem and an Amazon Redshift V2 target connection. If your organization doesn't have any Amazon Redshift V2 connections, you're prompted to create one.

After you choose the target connection, the Mapping Designer opens. When you create a mapping, a Source transformation and a Target transformation are already on the canvas for you to configure.

Sources in mappings in SQL ELT mode

When you configure the source connection in the Source transformation, you can choose an Amazon Redshift V2 or an Amazon S3 V2 connection.

Amazon Redshift V2 source properties

You can configure the following properties for an Amazon Redshift source:

- Source connection - Parameter, Allow parameter to be overridden at run time
- Source Type - Single, Query, Parameter
- Schema Name
- Source Table Name
- Pre-SQL
- Post-SQL
- SQL Query
- Select Distinct

Amazon S3 V2 source properties

You can configure the following properties for an Amazon S3 V2 source:

- Source connection - Parameter, Allow parameter to be overridden at run time
- Source Type - Single, parameter
- Format - Delimited
- Source Type - File and directory
- Folder Path
- File Name
- Compression Format - Gzip

For information on how to configure the supported properties, see the Amazon S3 V2 Connector documentation.

Targets in mappings in SQL ELT mode

When you configure a Target transformation in a mapping in SQL ELT mode, you need to use only a Amazon Redshift V2 connection.

You can configure the following properties in a Amazon Redshift V2 target transformation in a mapping in SQL ELT mode:

- Target Connection - Parameter, Allow parameter to be overridden at run time
- Operation - Insert, Update, Delete
- Truncate Target Table Before Data Upload
- Require Null Value For Char And Varchar
- Copy Options
- Treat Source Rows As
- Override Target Query
- Schema Name
- Target Table Name

- Pre-SQL
- Post-SQL

For information on how the targets in mappings in SQL ELT mode behaves differently from the targets in other types of mappings, see Targets in mappings in SQL ELT mode in *Mappings* in the Data Integration help.

Transformations in mappings in SQL ELT mode

A mapping in SQL ELT mode includes transformations that Amazon Redshift can process.

You can use the following transformations in a mapping in SQL ELT mode:

- Source
- Target
- Expression
- Filter
- Router
- Sorter

Functions in mappings in SQL ELT mode

When you create expressions within a mapping in SQL ELT mode, you must use the native functions and expression syntax of Amazon Redshift and not Informatica functions and expression syntax.

You can use the following native functions in a mapping in SQL ELT mode:

Math functions

Function	Function	Function
ABS()	DEGREES()	POW()
ACOS()	DEXP()	POWER()
ASIN()	DLOG1()	RADIANS()
ATAN()	DLOG10()	RANDOM()
ATAN2()	EXP()	ROUND()
CBRT()	FLOOR()	SIGN()
CEIL()	LN()	SIN()
CEILING()	LOG()	SQRT()
COS()	MOD()	TAN()
COT()	PI()	TRUNC()

For more information about the native functions and expression syntax of Amazon Redshift, see SQL functions reference in Amazon Redshift documentation.

Operators in mappings in SQL ELT mode

When you use mappings in SQL ELT mode, Data Integration converts the expression in the transformation by determining equivalent operators in the database. If there is no equivalent operator, Data Integration processes the transformation logic.

The table lists the operators that you can push to Amazon Redshift:

Operator	Operator
+	|/
-	||/
*	@
/	<<
%	>>
^	&

Rules and guidelines in mappings in SQL ELT mode

Consider the following rules and guidelines when you run mappings in SQL ELT mode:

General guidelines

- When you read from and write to Amazon Redshift, ensure that you use the same connection in the source and target transformations.
- You can't preview data in transformations configured midstream in the mapping flow.
- You cannot read Avro, ORC, and Parquet files from an Amazon S3 source.
- When you read from an Amazon S3 source, ensure that the source data doesn't contain the filename port.

Data types

- When you read from and write to Amazon Redshift, ensure that the source data doesn't contain the Boolean, Time, TIMESTAMPTZ, and Super data types.
- When you read from an Amazon S3 source, ensure that the source data doesn't contain hierarchical data types.

SQL ELT optimization for mapping tasks

You can configure SQL ELT optimization from a mapping task to enhance the mapping performance.

After you create a mapping, add the mapping to a mapping task, and then configure SQL ELT optimization in the mapping task. You can select how Data Integration handles SQL ELT optimization in the **SQL ELT Optimization Fallback Option** menu on the **Runtime Options** tab.

The task converts the transformation logic to Amazon Redshift queries, sends the queries to Amazon Redshift, and the mapping logic is processed in the Amazon Redshift database.

If your mapping contains multiple pipelines, you can define the flow run order to load the targets from the pipelines in a particular order.

You cannot use SQL ELT optimization when you run a mapping in advanced mode.

SQL ELT optimization types

When you apply SQL ELT optimization , the task pushes transformation logic to the source or target database based on the optimization type you specify in the task properties. Data Integration translates the transformation logic into SQL queries or Amazon Redshift commands to the Amazon Redshift database. The database runs the SQL queries or Amazon Redshift commands to process the transformations.

You can configure the following types of SQL ELT optimization in a mapping:

None

The task does not push down the transformation logic to the Amazon Redshift database.

Full

The task pushes as much of the transformation logic as possible to process in the target database. When a transformation is not supported in the mapping, the task partially pushes down the mapping logic to the point where the transformation is supported for SQL ELT optimization.

Source

The task pushes down as much as the transformation logic as possible to process in the source database. When you select source SQL ELT optimization, the task pushes the transformation logic for all the configured transformations downstream in the mapping until it reaches the Target transformation. It leaves out the target transformation logic and pushes down the rest of the mapping logic to the source database.

Data Integration behavior with source and full SQL ELT optimization

When you select full or source SQL ELT optimization for a mapping that reads from or writes to Amazon Redshift, Data Integration analyzes the mapping from the source to the target or until it reaches a downstream transformation to determine whether to push down only partial or the entire mapping logic for processing to the database.

If all the transformations in the mapping are supported for SQL ELT optimization, the task pushes down the entire mapping logic to the database.

When a transformation is not supported in the mapping, the task partially pushes down the mapping logic to the point where the transformation is supported for SQL ELT optimization. Data Integration generates and executes a SELECT statement for the transformation logic that needs to be pushed down. Then, it reads the results of this SQL query and processes the remaining transformations in the mapping.

SQL ELT optimization scenarios

You can configure SQL ELT optimization using the Amazon Redshift V2 Connector or the Amazon Redshift ODBC Connector in mappings.

You can use configure SQL ELT optimization for the following scenarios when you use Amazon Redshift V2 Connector in mappings:

Note: You cannot configure SQL ELT optimization for a mapping task that is based on mapping in advanced mode.

Source and target endpoints	Supported SQL ELT optimization scenarios in mappings	SQL ELT optimization type
Amazon S3 source Amazon Redshift target	Reads from Amazon S3 and writes to Amazon Redshift using the Amazon Redshift V2 connection. When data is read from the source to the target, this connection uses the AWS commands.	Full
Amazon Redshift source Amazon Redshift target	Reads from Amazon Redshift and writes to Amazon Redshift using the Amazon Redshift V2 connection. The source and target can be in the same or different cluster regions.	Source, Full **Note:** The Secure Agent pushes the entire mapping logic or only the partial mapping logic, as applicable, for processing to Amazon Redshift.
Amazon Redshift source	Reads from Amazon Redshift using the Amazon Redshift V2 connection and writes to other targets.	Source
Note: You can use the Secure Agent or the Hosted Agent to run mappings enabled with SQL ELT optimization.		

Note: You can configure SQL ELT for a mapping that uses an Amazon Redshift ODBC connection to read from and write to Amazon Redshift. Informatica recommends that you use the Amazon Redshift V2 connection in mappings to configure SQL ELT optimization. If you cannot push down specific transformation logic using the Amazon Redshift V2 connection, you can explore configuring SQL ELT optimization using the Amazon Redshift ODBC connection.

The Amazon Redshift ODBC connection uses the Amazon ODBC 64-bit drivers on Windows and Linux systems. For more information, see the How-To Library article, Configuring SQL ELT optimization for Amazon Redshift using the ODBC Connector.

SQL ELT compatibility

You can configure the task to push transformations, functions, and operators to the database.

When you use SQL ELT optimization, the Secure Agent converts the expression in the transformation by determining equivalent operators and functions in the database. If there is no equivalent operator and function, the Secure Agent processes the transformation logic.

Functions with Amazon Redshift V2

When you use SQL ELT optimization, the Secure Agent converts the expression in the transformation by determining equivalent functions in the database. If there is no equivalent function, the Secure Agent processes the transformation logic.

The following table summarizes the availability of SQL ELT functions in an Amazon Redshift database. Columns marked with an X indicate that the function can be pushed to the Amazon Redshift database. Columns marked with a dash (-) symbol indicate that the function cannot be pushed to the database.

Function	SQL ELT	Function	SQL ELT	Function	SQL ELT
ABORT()	-	INITCAP()	X	REG_MATCH()	X
ABS()	X	INSTR()	X	REG_REPLACE	X
ADD_TO_DATE()	X	IS_DATE()	-	REPLACECHR()	X
AES_DECRYPT()	-	IS_NUMBER()	X	REPLACESTR()	X
AES_ENCRYPT()	-	IS_SPACES()	X	REVERSE()	-
ASCII()	X	ISNULL()	X	ROUND(DATE)	-
AVG()	X	LAST()	-	ROUND(NUMBER)	X
CEIL()	X	LAST_DAY()	X	RPAD()	X
CHOOSE()	-	LEAST()	-	RTRIM()	X
CHR()	X	LENGTH()	X	SET_DATE_PART()	-
CHRCODE()	-	LN()	X	SIGN()	X
COMPRESS()	-	LOG()	X	SIN()	X
CONCAT()	X	LOOKUP()	X	SINH()	-
COS()	X	LOWER()	X	SOUNDEX()	-
COSH()	-	LPAD()	X	SQRT()	X
COUNT()	X	LTRIM()	X	STDDEV()	X
CRC32()	-	MAKE_DATE_TIME()	-	SUBSTR()	X
CUME()	-	MAX()	X	SUM()	X
DATE_COMPARE()	X	MD5()	X	SYSTIMESTAMP()	X
DATE_DIFF()	X	MEDIAN()	X	TAN()	X
DECODE()	X	METAPHONE()	-	TANH()	-
DECODE_BASE64()	-	MIN()	X	TO_BIGINT	X
DECOMPRESS()	-	MOD()	X	TO_CHAR(DATE)	X

Function	SQL ELT	Function	SQL ELT	Function	SQL ELT
ENCODE_BASE64()	-	MOVINGAVG()	-	TO_CHAR(NUMBER)	X
EXP()	X	MOVINGSUM()	-	TO_DATE()	X
FIRST()	-	NPER()	-	TO_DECIMAL()	X
FLOOR()	X	PERCENTILE()	-	TO_FLOAT()	X
FV()	-	PMT()	-	TO_INTEGER()	X
GET_DATE_PART()	X	POWER()	X	TRUNC(DATE)	X
GREATEST()	-	PV()	-	TRUNC(NUMBER)	X
IIF()	X	RAND()	-	UPPER()	X
IN()	X	RATE()	-	VARIANCE()	X
INDEXOF()	-	REG_EXTRACT()	-		

Operators with Amazon Redshift V2

When you use SQL ELT optimization, Data Integration converts the expression in the transformation by determining equivalent operators in the database. If there is no equivalent operator, the Data Integration processes the transformation logic.

These operator can be pushed to the Amazon Redshift database by using full SQL ELT optimization.

The following table lists the operators that you can push to an Amazon Redshift database:

Operator	Operator
+	=
-	>=
*	<=
/	!=
%	AND
||	OR
>	NOT
<	

Variables with Amazon Redshift V2

You can use full SQL ELT to push the SYSDATE variable to the Amazon Redshift database in an expression transformation.

Note: When you use full SQL ELT to push the SYSDATE variable, it corresponds to the date/time data type of the Amazon Redshift cluster, while for mappings without SQL ELT optimization the SYSDATE variable corresponds to the date/time data type of the agent.

Configuring SQL ELT optimization

To optimize a mapping, add the mapping to a task, and then configure SQL ELT optimization in the mapping task.

1. Create a mapping task.
2. In the SQL ELT optimization section on the **Runtime Options** tab, set the SQL ELT optimization value to **Full** or **To Source**.

   You can use Source SQL ELT optimization only when you read from an Amazon Redshift source.

   Note: The optimization context type option is not applicable for an Amazon Redshift V2 mapping task.
3. If full SQL ELT optimization is not available, select how Data Integration handles SQL ELT optimization in the **SQL ELT optimization Fallback Option** menu:
   - Partial SQL ELT optimization. Default. Data Integration pushes as much transformation logic as possible to the source and target database. The task processes any transformation logic that it can't push to a database. You can use Partial SQL ELT optimization only when you read from and write to Amazon Redshift.
   - Non SQL ELT optimization. The task runs without SQL ELT optimization.
   - Fail Task. Data Integration fails the task.

   Note: The fallback options are not applicable to mappings in advanced mode.

When you run the mapping task, the transformation logic is pushed to the Amazon Redshift database. To verify that the SQL ELT optimization has taken place, you can check the session log for the job. You can monitor the jobs that you initiated on the **My Jobs** page.

Transformations for Amazon Redshift V2 mappings

When you configure SQL ELT optimization, the Secure Agent tries to push the configured transformation to Amazon Redshift.

The following list summarizes the availability of transformations that you can push down to Amazon Redshift.

- Aggregator
- Expression
- Filter
- Joiner
- Lookup
- Sorter
- SQL
- Router
- Union

Aggregator transformation

You can configure full SQL ELT optimization to push an Aggregator transformation to process in Amazon Redshift.

Aggregate calculations

You can perform the following aggregate calculations:

- AVG
- COUNT
- MAX
- MIN
- MEDIAN
- SUM
- VARIANCE

Incoming fields

When you configure an Aggregator transformation and the incoming field is not used in an aggregate function or in a group by field, the output is not deterministic as the ANY_VALUE() function returns any value from the field.

Expression transformation

You can configure full SQL ELT optimization to push an Expression transformation to process in Amazon Redshift.

You can add an Expression transformation to each of the sources in the mapping, followed by a join downstream in the mapping. Additionally, you can add multiple Expression transformations that branch out from a transformation and then branch in into a transformation downstream in the mapping.

When you configure an Expression transformation, consider the following rules to include variables in the expression:

- You cannot use variables where you are using the value assigned while processing a previous row for calculations in the current row. if you do, the mapping runs without SQL ELT optimization.
- The variables can be nested, but you cannot refer to a variable before it is defined in the expression. If the variables are not defined in that order, the mapping runs without SQL ELT optimization.

  For example,

  ```
  var: AGEPLUS2 = AGEPLUS1 + 1
  var: AGEPLUS1 = AGE + 1
  out: NEXTAGE = AGEPLUS2 + 1
  ```

  Here, AGE +1 is defined later. AGEPLUS2 in the first variable refers to AGEPLUS1 and remains unresolved. To resolve this, specify the variables in the following order:

  ```
  var: AGEPLUS1 = AGE + 1
  var: AGEPLUS2 = AGEPLUS1 + 1
  out: NEXTAGE = AGEPLUS2 + 1
  ```

- The variables cannot have an expression that is cyclic or refers to itself:
  For example,

  ```
  var: AGEPLUS1 = AGEPLUS2 + 1
  var: AGEPLUS2 = AGEPLUS1 + 1
  out: NEXTAGE= AGEPLUS2
  ```

  Here, AGEPLUS1 refers to AGEPLUS2 and remains unresolved.

Joiner transformation

You can configure a Joiner transformation between two Amazon S3 sources or two Amazon Redshift sources.

Lookup transformation

You can configure full SQL ELT optimization to push a Lookup transformation to process in Amazon Redshift. You can push both a connected and an unconnected lookup.

Consider the following rules when you configure a lookup in a mapping:

- You can configure a lookup to Amazon S3 only when the source used is Amazon S3.
- You can configure a lookup to Amazon Redshift when the source used is Amazon S3 or Amazon Redshift.

You can nest the unconnected lookup function with other expression functions.For more information on specific rules for lookups, see "Rules and guidelines for Lookup transformation" on page 121.

Sorter transformation

You can configure full SQL ELT optimization to push a Sorter transformation to process in Amazon Redshift.

When you configure a Sorter transformation, only the following sorter advanced properties are applicable:

- Distinct. You can remove the duplicate rows and create a distinct query.
- Null Treated Low. You can sort the columns in a query based on the columns having null values. You can sort the columns only in a query. The sort order is not honored on the Redshift target.

SQL transformation

You can use an SQL transformation to push Redshift supported scalar functions to Amazon Redshift.

When you configure SQL ELT optimization for a mapping, you can use SQL user-defined functions (UDFs) in a SQL transformation and run queries with the Amazon Redshift target endpoint.

You can use only the SELECT clause SQL statement to push down a function. The following snippet demonstrates the syntax of a simple select SQL query:

```
SELECT <function_name1>(~Arg~), <function_name2> (~Arg~)...
```

You must provide the corresponding query in the following format:

```
select <Redshift_supported_scalar_function1> (~Arg~),
<Redshift_supported_scalar_function2> (~Arg~)
```

You can push a SQL transformation with the following restrictions:

- You can configure only a SQL query in the SQL transformation. You cannot enable a stored procedure when you configure SQL ELT optimization to Amazon Redshift.
- The SQL query must be a simple SELECT statement without 'FROM' and 'WHERE' arguments. The SQL transformation only supports functions with Simple SELECT statement.
- When you specify a SELECT query, you must also specify the column name and number of columns based on the functions. For example, when you specify the query `select square(~AGE~), sqrt(~SNAME~)`, you must specify two output columns for `AGE` and `SNAME` functions each, otherwise the mapping fails.
- You can only use a SQL transformation when the SELECT statement is present in the query property. Even if an entire query containing the SELECT statement comes from a parameterized input port, the SQL ELT optimization fails.

- If any SQL error occurs, the error is added to the `SQLError` field by default. However, when mapping runs in PDO mode, the `SQLError` field will remain as Null.
- The `NumRowsAffected` field number records the number of rows affected while computing the output buffer. However, for SQL transformation, the `NumRowsAffected` will be 0 since the query runs for all records in a single go and not for each row.
- Amazon Redshift offers only passive behavior of SQL transformations where the support for dynamic queries are limited.

Router transformation

When you configure a Router transformation, you must connect or map only one output group to the target transformation.

Features

You can configure SQL ELT optimization for a mapping that reads from the following sources and writes to an Amazon Redshift target:

- Amazon Redshift V2 source
- Amazon S3 V2 source

When you configure a mapping, some parameters are not supported for a mapping enabled for SQL ELT optimization. You can refer to the list of parameters that each source supports.

Amazon Redshift V2 sources, targets, and lookups

You can configure an Amazon Redshift V2 connection in the source transformation with basic, IAM, and assume role authentication and enable SQL ELT optimization in the mapping task.

Connection properties

When you configure SQL ELT optimization, you can use the following advanced properties for an Amazon Redshift V2 source in the Amazon Redshift V2 connection:

- User name
- Password
- Access key and secret key
- JDBC URL
- Region
- Master symmetric key
- Customer master key ID

Source properties

When you configure SQL ELT optimization, you can use the following properties for an Amazon Redshift V2 source in the Source transformation:

- Source connection- Parameter, Allow parameter to be overridden at run time
- Source type - Single object, multiple objects, query, and parameter
- Filter
- Read mode - Only staging mode is applicable in mappings.

- S3 bucket name
- Enable compression
- Unload options
- Encryption type - SSE-S3, SSE-KMS, SSE-SMK

  Note: You must specify the same master symmetric key in the target properties.
- Schema name
- Source table name
- Pre-SQL
- Post-SQL
- SQL query
- Select distinct

Target properties

When you configure SQL ELT optimization, you can use the following properties for an Amazon Redshift V2 target:

- Target connection- Parameter, Allow parameter to be overridden at run time
- Target type- Single object, parameter
- Allow parameter to be overridden at run time
- Target object- Existing target, Create new at runtime
- Operation- Insert, update, upsert, delete, or data driven
- Analyze target table
- Assume Role
- COPY command
  - Region
  - Truncatecolumn
  - AWS_IAM_Role, only for Parquet and ORC files
- Pre-SQL
- Post-SQL
- Require null value for Char and Varchar
- Schema name
- Target table name
- Truncate target table before data upload
- Vacuum table
- Override Target Query
- Treat Source Rows As (Insert, Update, Upsert, Delete, and None). Select None when you use data driven operation.
- Max Errors per Upload Batch for INSERT

Note: If you configure source and target advanced properties that are not applicable, the mappings run in the Informatica runtime environment.

Lookup properties

When you configure SQL ELT optimization, you can use the following properties for Amazon Redshift V2 lookups:

- Lookup connection- Parameter, Allow parameter to be overridden at run time
- Source type - Single object, query, parameter
- Multiple matches - Report error. Applicable for unconnected lookups.
- Multiple matches- Return all rows. Applicable for connected lookups.
- S3 Bucket Name
- Enable Compression
- Unload Options
- Encryption Type
- Schema Name
- Source Table Name
- Pre-SQL
- Post-SQL
- SQL Query
- Select Distinct

Note: You can specify a parameter file in the mapping task to override the Amazon Redshift V2 source, lookup, and target connections and objects in a mapping.

Amazon S3 V2 source

You must configure an Amazon S3 V2 connection with basic, assume role, and IAM authentication when you enable SQL ELT optimization in a mapping task.

Source properties

When you configure SQL ELT optimization, you can use the following properties for an Amazon S3 V2 source in a mapping:

- Source connection
- Source type- Single object, parameter
- Parameter
- Encryption type- Client-side encryption (applicable to flat files), Server-side encryption, Server-side encryption with KMS
- Data compression- Gzip (applicable to flat and Parquet files), Deflate (applicable to Avro files), Snappy (applicable to Avro, Parquet, and ORC files), and Zlib (applicable to ORC files)
- File format
  - Delimiter
  - Qualifier
  - Code Page
  - Header Line Number
  - First Data Row
  - Source Type

- Folder Path

- File source type
- File name
- Format type
  - Avro
  - Parquet
  - ORC
  - JSON
  - Delimited

SQL ELT optimization for multiple targets

You can add multiple Amazon Redshift V2 targets in a mapping.

When you configure a mapping to write to multiple Amazon Redshift V2 targets, you can further optimize the write operation when you configure full SQL ELT optimization.

To optimize, you can choose to configure an insert, update, upsert, delete, or data driven operation for multiple targets individually.

You can select the same or different Amazon Redshift V2 target table in multiple Target transformations and perform different operations for each of the Target transformations to run independent of each other.

If your mapping contains multiple pipelines, you can define the flow run order to load the targets from the pipelines in a particular order.

When you define the flow run order for multiple pipelines and also set `-DEnableSingleCommit=true` for a single pipeline, the `-DEnableSingleCommit=true` property in given precedence.

Rules and guidelines

Consider the following rules and guidelines when you optimize full SQL ELT optimization for multiple targets:

- When you run a mapping and use the same target table in all the targets, the row count is different from the row count of the mapping that runs without SQL ELT optimization. Applicable when you use an Amazon S3 source and perform an upsert operation on multiple Amazon Redshift V2 targets.
- When you run a mapping and use the same target table in all the targets, the Secure Agent writes a different set of data to the target than the data of the mapping that runs without SQL ELT optimization. Applicable when you use an Amazon S3 source, perform an insert operation on multiple Amazon Redshift V2 targets, and enable the **Truncate Target Table Before Data Load** advanced target property.

Determine the order of processing for multiple targets

When you configure a mapping to write to multiple targets in a single pipeline, with each target configured for any write operation, the target operations are processed in a predefined order of delete, update, upsert, and insert.

To process the target operations in the predefined order, you need to set certain properties in the Secure Agent and in the task properties.

Set -DEnableSingleCommit=true in the Secure Agent properties

Perform the following tasks to set the property for the Secure Agent:

1. Open Administrator and select **Runtime Environments**.

2. Select the Secure Agent for which you want to set the property.
3. On the upper-right corner of the page, click **Edit**.
4. In the **System Configuration Details** section, select the **Type** as **DTM** for the Data Integration Service.
5. Edit the JVM options and set the property to *-DEnableSingleCommit=true*.



Set the EnableSingleCommit property in the task properties

Perform the following tasks to set the property in the task:

1. On the **Runtime Options** tab in the mapping task properties, navigate to the Advanced Session Properties section.
2. From the **Session Property Name** list, select **Custom Properties**, and set the **Session Property Value** to **Yes**.



Consider the following guidelines when the target operations are processed in a predefined order:

- After you run a mapping, the results are not displayed in the order of query execution. You can view the queries and the affected rows for each query in the session log.
- When you select the data driven target operation and set the **Treat Source Rows As** option to **None** for a target, consider the following guidelines:
  - When you enter a single value, such as, DD_INSERT, DD_UPDATE, or DD_DELETE in the data driven condition, the target operation specified in the data driven condition is considered and the target operations are processed in a predefined order of delete, update, upsert, and insert.
  - When you enter a Nested IFF statement or DD_REJECT in the data driven condition, the data driven target operation is considered and the target operations are not processed in a predefined order.
- The order of execution of post SQL and pre SQL queries after the -DEnableSingleCommit=true property is set differs from the order of execution before the property was set.

Clean stop a SQL ELT optimization job

When a task enabled for SQL ELT optimization is running, you can clean stop the job to terminate all the issued statements and processes spawned by the job.

Use the **Clean Stop** option on the My Jobs page in Data Integration and the All Jobs and Running Jobs page in Monitor.

See the following exceptions before you clean stop a SQL ELT optimization task:

- When you clean stop a task enabled for SQL ELT optimization and the target or source properties in the mapping contains SQL statements, all the SQL statements in progress are attempted to be reverted. All SQL statements that are committed cannot be reverted.
- When you run a mapping configured to create a new target at runtime and clean stop the job immediately, Data Integration creates the target table even if the job is terminated.

Rules and guidelines for SQL ELT optimization

Use the following rules and guidelines when configuring SQL ELT optimization for an Amazon Redshift database:

- To copy data from Amazon S3 to Amazon Redshift, you must use multiple data files by splitting large files. For more information, see the Amazon documentation.
- When you use an upper case in the column name of a JSON file and configure SQL ELT optimization, the mapping fails with the following error:

  ```
  Error Reason: Invalid JSONPath format: Member is not an object.
  ```
- When you configure a mapping with SQL ELT optimization and define an ORC source with date/time values, the Secure Agent might not write a few values correctly to the Redshift target.
- When you define a Parquet source with decimal values having precision greater than 19,0 and use **COPY** command to write to a Redshift target, the mapping fails with SQL ELT optimization.
- When the data in delimited, Avro, or JSON files has values that are greater than the precision values, specify the attribute **TRUNCATECOLUMNS=ON**.
- For the ORC and Parquet file types, specify **AWS_IAM_ROLE** in the COPY command, to enable full SQL ELT optimization.
- You cannot use the assume role when the source has a Parquet or ORC file format and enable full SQL ELT optimization for the mapping task.
- You cannot enable full SQL ELT optimization for a mapping task when the task contains a mapping with a single transformation connected to multiple transformations downstream and vice-versa.
- You cannot enable full SQL ELT optimization for an Avro source with Date, Decimal, and Timestamp data types.
- When you read data from an Amazon S3 flat file with Shift-JIS encoding, write to an Amazon Redshift target, and enable full SQL ELT optimization, the mapping fails.
- When you configure a mapping that reads an Amazon S3 Avro source with column names in uppercase letters and uses an Amazon Redshift V2 connection to write data to an Amazon Redshift target, the COPY command writes blank rows to the target.
- When you configure an update operation for data that contains the Time column for a mapping enabled with full SQL ELT optimization and you override the target query using the value specified in the Override Target Query field from the advanced target properties, the task runs but data is not processed.
- If the Amazon S3 bucket region and the Amazon Redshift region are different, specify the **REGION** attribute in the **COPY** command to enable full SQL ELT optimization. Applies to delimited, Avro, and JSON files.
- When you perform an update, upsert, or delete operation, ensure that you specify a primary key or an update column for the target table.
- When you perform an upsert operation and set the JVM option `-DuseDeleteForUpsert=true`, the target statistics of processed rows shows an additional row as compared to the case when you do not set the JVM option. The number of rows in the target table are the same in both the cases.
- When you perform an update, upsert, or delete operation on an Amazon Redshift target and specify the update column as a Date/Time data type in a different time zone, data is not written to the target. However, the task runs successfully. The issue occurs when you use an Amazon S3 Parquet source.
- When you perform an update operation, you cannot map an id column of a target table in the field mapping.

- Even if you configure a condition using a DD_INSERT, DD_UPDATE, or DD_DELETE data driven operation on a target object, the log contains queries for the remaining data driven operations as well. The mapping runs successfully.
- When the data driven condition contains only the DD_REJECT operation, the mapping runs without generating a query.
- When you parameterize a transformation in a mapping enabled for full SQL ELT optimization and configure a parameter file to override the input parameters, the Secure Agent ignores the overridden parameters.
- If the Union transformation has inputs from a few source columns and you do not map the rest of the columns to the target, the columns that you do not map show null data in target.
- If the input fields contain a decimal field and you do not map the decimal field to the target in a Union transformation, the mapping that runs with full SQL ELT optimization and uses an Amazon Redshift V2 connection fails with an error.
- If the mapping enabled for SQL ELT optimization contains Union and Aggregator transformations, include the incoming field from the aggregate function or group by field in the field mapping, or remove the field from the aggregate function or group by field altogether. Otherwise, the mapping fails.
- A mapping fails if the column of the Date or Datetime data type is not of the `YYYY-DD-MM` format.
- You must map all the fields from the SQL query to the target for the mappings enabled for SQL ELT optimization to run successfully.
- If the custom query contains duplicate columns, the mapping runs without SQL ELT optimization.
- When you run a mapping enabled for SQL ELT optimization that uses an Amazon Redshift V2 connection to update data with the float or integer data types in an Amazon Redshift target, the mapping might fail.
- When you parameterize an Expression transformation in a mapping task and configure a parameter file to override the parameters, the Secure Agent does not read the overridden parameters. The issue occurs when you configure full SQL ELT optimization for a mapping that uses an Amazon Redshift V2 connection.
- A mapping fails at runtime when you specify an advanced native filter in the following format: `schema_name.table_name.column_name`. While defining the advance filter, the condition should not have table name qualified with a schema name.
- When you use the query source type to read from Amazon Redshift, you can choose to retain the field metadata and save the mapping. Even if you edit the query and run the mapping, the field metadata specified at design time is retained.

Rules for SQL ELT optimization in mappings that read from and write to Amazon Redshift

Consider the following guidelines when you configure full SQL ELT optimization for mappings that read from or write to Amazon Redshift:

- When the Amazon Redshift source and target are in the same cluster, the Secure Agent does not use the UNLOAD and COPY commands to stage the data on Amazon S3 and then write to Amazon Redshift. The Secure Agent directly performs an insert, update, upsert, delete, or data driven operation to write the data to Amazon Redshift and the performance of the task is improved.
- When you configure an update operation for data that contains the Time column for a mapping enabled with full SQL ELT optimization and you override the target query using the value specified in the **Override Target Query** field from the advanced target properties, the task runs but data is not processed.
- When you assume an IAM role and run mappings enabled with source or full SQL ELT optimization to read data from Redshift, the assume role is not honored.

To honor the assume role and run mappings successfully, you must specify the AWS_IAM_ROLE property and set its value in the **Unload Options** field of the Source transformation.

For example, `AWS_IAM_ROLE=arn:aws:iam::0093:role/redshift_role`

- If the source and target objects in a mapping point to Amazon S3 buckets in different regions, specify the **REGION** attribute in the **COPY** command to set the Amazon S3 bucket for target in the same region as Amazon S3 bucket for source.
- If the data that you read contains delimiters (|), quotes ("), and escape (\) characters and you set the **Unload Options** field in the Source transformation to `ADDQUOTES;DELIMITER = \174;`, the mapping fails. To avoid this error, set the following properties in the **Copy Options** field in the Target transformation: `ESCAPE;REMOVEQUOTES;CSV=OFF;QUOTE=OFF;DELIMITER = \174;`
- When you read data from a single column table of Bigint, Datetime, and Boolean data type that contains NULL values, the null values are not written to the target.
  To avoid this error, you must set `IGNOREBLANKLINES` in the **Copy Options** field of the Target transformation.
- When you configure a Source transformation in a mapping to read data and you set `ESCAPE=OFF` or ON and the `ADDQUOTES` in the **Unload Options** field to add escape characters and quotes in the data, null values are written to the target.
  To avoid this, you must add `QUOTE=\042` in the **Copy Options** field of the Target transformation and then run the mapping.
- If the data contains an escape (\) character when you read data, you must specify ESCAPE=OFF in the **Unload Options** field of the Source transformation. If you do not set the property, the escape character gets duplicated in the target. For example, the data output 12\12\2012 appears as 12\\12\\2012.
- When you read from and write data that contains float4 or double values, the float values show a slight change in the target.
- When you specify a user defined function in a transformation, the function name is not displayed correctly in the session log.
- If the mapping contains a Sequence Generator transformation, ensure that you map the NEXTVAL output field to the target. Do not map the CURRVAL output field to the target. Else, the mapping task does not partially push down the mapping logic to the point where the transformation is supported and runs without SQL ELT optimization.
- When you configure a mapping with multiple objects at source where the table name and column name are the same or the table name is substring of the column name, the mapping fails.
- You cannot perform data driven operations on target objects of the timestamptz data type.
- When you run a mapping in full SQL ELT optimization with multi-objects at source using advance filters, the UNLOAD and COPY commands do not work across different clusters and the mapping fails.
- When you define pre-SQL or post-SQL queries that modify the same object in a mapping where the source and target belong to different clusters, the task fails at runtime in full SQL ELT optimization.
- When a mapping contains multiple pipelines, the logic of only one pipeline is pushed to the source if an unsupported function is encountered. The other pipelines run without SQL ELT optimization.

Rules for source SQL ELT optimization in mappings that read from Amazon Redshift

Consider the following guidelines when you configure source SQL ELT optimization for mappings that read from Amazon Redshift:

- You can configure an SQL query or custom query in the advanced source property to push the mapping to Amazon Redshift.

- When you use a custom query as a source or an SQL override, the table name alias is not generated as expected. The name starts with `"INF"` in the SQL ELT optimization query.
- When you run a mapping by overriding the SQL query in full SQL ELT optimization, where the source column names are aliased, the mapping fails. Ensure that the alias names and the source column names are the same.
- You cannot push a Router transformation with multiple output groups to the Amazon Redshift source.
- When you use the query source type to read from Amazon Redshift, you can choose to retain the field metadata and save the mapping. Even if you edit the query and run the mapping, the field metadata specified at design time is retained.
- You cannot set the read mode to **Direct** in Source and Lookup transformations.
- COPY command options are not applicable for the timestamptz data type.
- When you use the `[^\x00-\x7F]+` pattern in a REG_REPLACE() function and run a mapping with SQL ELT optimization, both the non-ASCII characters and spaces are replaced in the target. But, when you run the same mapping without SQL ELT optimization, only the non-ASCII characters are replaced and the spaces are retained in the target.
  For more information about the patterns you can use in an expression, see Pattern-matching conditions.

Rules and guidelines for adding multiple source objects

Consider the following rules and guidelines when you add multiple source objects:

- You must specify double quotes for the table name when you use a reserved word for the table name and the column name.
- You cannot use a self join when you add multiple source objects.
- When you use special characters in column names for an advanced relationship, the query formed is not correct and the mapping task fails.
- You can use the full outer-join condition only with the `=`, `,` and `AND` operators.
- When you override the schema name and configure an advanced filter on a related source object, the Secure Agent applies the advanced filter only on the parent object and not on the related source object.
- When you select parent and child objects that have a primary key and foreign key relationship, and the foreign key of the related object is also a primary key in the table, the mapping task fails when you create a target.
- When you select the Multiple Objects source type, add a source object, for example, `emp`, and define a primary key and foreign key relationship on different columns, for example, `emp.id` and `dept.d_id`, the mapping fails with the following error:
  `[FATAL] Unload/Copy command failed with error: Invalid operation: column emp.d_id does not exist.`

  The Select Related Objects list shows the join condition for the `dept` related object `asemp.d_id=dept.d_id`, even though `theemp` table does not have `d_id` column.
- When you select the Multiple Objects source type, you cannot use a period(.) in the table name.

Rules and guidelines for functions

Use the following rules and guidelines when pushing functions to an Amazon Redshift database:

- To push TO_DATE() and TO_CHAR() to Amazon Redshift, you must define the string and format arguments.
- If you use the NS format as part of the ADD_TO_DATE() function, the agent does not push the function to Amazon Redshift.

- If you use any of the following formats as part of the TO_CHAR() and TO_DATE() functions, the agent does not push the function to Amazon Redshift:

  - - NS

  - - SSSS

  - - SSSSS

  - - RR

- To push TRUNC(DATE), GET_DATE_PART(), and DATE_DIFF() to Amazon Redshift, you must use the following formats:

  - - D

  - - DDD

  - - HH24

  - - MI

  - - MM

  - - MS

  - - SS

  - - US

  - - YYYY

- You can use **REPLACESTR()** only to replace a single string value with a new string value.
  Syntax

  ```
  REPLACESTR (CaseFlag, InputString, OldString, NewString).
  ```

- To push **SUBSTR()** to Amazon Redshift, you must define an integer value for the length argument.
- When you push MD5() to Amazon Redshift, the Secure Agent returns all the MD5 fields in the lower case. However, when you run a mapping without SQL ELT optimization, the Secure Agent returns the MD5 fields in the upper case.
- When you use the IS_NUMBER function in a transformation and the input data contains d or D, for example in formats such as +3.45d+32 or +3.45D-32 , the function returns False or 0.
- When you use the IN function in an expression, you must not include the `CaseFlag` attribute.
- When you use the IN function and the arguments contain date and timestamp values, you must include the TO_DATE function in the expression.
- When you use the REG_REPLACE function in an Expression transformation, ensure that the expressions used in the argument are supported by AWS.

- When you use the ISNULL() function in an Expression transformation, Data Integration parses the column values differently when the source and target in the mapping are from the same cluster environment or across different cluster environments. The UNLOAD and COPY commands parse NULL and empty string values differently when the source and target are in different clusters.

  You can enable the use `TempTableForRedshiftAPDO` property when you push down functions from mappings where the source and target are across different clusters. When you set the property and run the mapping, the mapping considers NULL and empty string values in the columns as NULL only. However, when the source and target is in the same cluster, Data Integration considers NULL as NULL and empty strings as empty strings in the columns.

  Mappings without SQL ELT optimization parses the ISNULL() function differently. When you select the Treat NULL Value as NULL option, the mapping considers NULL and empty string values as NULL. When you don't select the Treat NULL Value as NULL option, the mapping considers NULL and empty strings values as empty strings.

Rules and guidelines for aggregate functions

Use the following rules and guidelines when pushing aggregate functions to an Amazon Redshift database:

- You cannot use conditional clauses in the aggregate expression functions.
- You can use non-aggregate functions in aggregate expressions.
- You cannot use nested aggregate functions directly. For example, you cannot specify `SUM(AVG(col1))` in the aggregate expression function columns. Use nested aggregations by adding multiple aggregate transformations in the mapping.
- You can parameterize the `GROUP BY` fields and aggregate functions in a mapping task.
- When you use STDDEV or VARIANCE functions for an expression that consists of a single value and run the mapping, the result of the function is NULL. When you run the mapping without pushing it down, the result of the function is 0.
- During the field mapping, you must map only the fields that you added in the `GROUP BY` port. Else, the mapping runs in the non-PDO mode with an error message.
- When you do not specify a port from an Amazon S3 flat file source in the `GROUP BY` clause of the aggregate function and map the port to a Redshift target, the mapping task runs successfully without SQL ELT optimization with the following message:

  ```
  Pushdown optimization to the source stops before transformation [Aggregator] because [f_varchar] is a non-group by passthrough port, which is not allowed.
  ```

  The mapping fails when you push down a mapping with an Amazon S3 Avro or Parquet source.
- A mapping enabled for full or source SQL ELT optimization which contains a table column with a regular expression argument in the REG_MATCH() function fails. However, the mapping runs successfully when you do not enable SQL ELT optimization.
- When you configure the REG_MATCH() expression:
  - Use double-slashes for parsing expressions in full SQL ELT optimization. For example, use REGMATCH(searchstring,'\\D W').
  - Use single-slash to parse expressions without SQL ELT optimization. For example, use REGMATCH(searchstring,'\D\W').

- When you pass timestamp and date data values through an ASCII() function, the ASCII() function parses the values differently with and without SQL ELT optimization.
  In a mapping without SQL ELT optimization, the ASCII() function returns the first digit of the day, while in full and source SQL ELT optimization, the function returns the first digit of the year.

  For example, to read a date time MM/DD/YYYY HH24:MI:SS.US for a session, the ASCII() value returns the first character of MM without SQL ELT optimization and the first character of YYYY with SQL ELT optimization.

  As a workaround, you can set the DateTime Format String property value to yyyy-mm-dd in the **Advanced Session Properties** when you enable mappings without SQL ELT optimization.

Rules and guidelines for Router transformation

Consider the following rules and guidelines for a Router transformation:

- For a source field, do not edit the metadata for an id column from `string` to `int`.
- When the source has a field of the binary data type, the mapping fails with the following error: `[Full Pushdown Optimization Failed Due To PARQUET : Not Supported data type binary.]`
- When you create conditions for router groups and connect the router groups to target tables, and one of the condition fails, the mapping does not process the rest of the conditions as well. The mapping runs without SQL ELT optimization.

Rules and guidelines for Lookup transformation

When you configure a Lookup transformation, adhere to the following guidelines:

- You cannot configure dynamic lookup cache and persistent cache.
- You cannot configure the advanced lookup properties for connected and unconnected lookups.
- Even if you select an advanced lookup property for a Lookup transformation, the Secure Agent ignores the advanced property. An error message does not appear in the logs.
- If you add the ADDQUOTE option in the Unload command for an Amazon Redshift lookup, you must also add the QUOTE option as `QUOTE="` in the Copy command for the Amazon Redshift target.
- If the source and target in a mapping point to Amazon S3 buckets in different regions, specify the **REGION** attribute in the **COPY** command to set the Amazon S3 bucket for target in the same region as Amazon S3 bucket for source.
- If you configure the CSE-SMK encryption type for an Amazon Redshift source or lookup object, ensure that you specify a master symmetric key in the target properties.
- You can only specify the `=` operator in a completely parameterized lookup condition. If you specify the operators such as `<`, `<=`, `>`, `>=`, and `!=` in a complex condition, the mapping fails.
- When you configure a lookup for an Amazon S3 source, remove the FileName field from both the Amazon S3 source and lookup object. The FileName field is not applicable.

Connected lookups

Consider the following rules and guidelines for a connected Lookup transformation:

- You must select the **Multiple Matches** property value as **Return all rows** in the connected lookup properties for SQL ELT optimization to work.
- When an Amazon S3 location contains source files having the same prefix, for example, `abc.txt.1` and `abc.txt.2`, the COPY command tries to load both the files to an Amazon Redshift target and the mapping might fail.

Unconnected lookups

Consider the following rules and guidelines for an unconnected Lookup transformation:

- You must select the **Multiple Matches** property value as **Report error** in the unconnected lookup properties for SQL ELT optimization to work. However, when multiple matches are encountered in the lookup table, the Secure Agent does not report an error. Hence, ensure that multiple matches are not encountered in the lookup table.
- If you select the **Return Any** property, the mapping task runs without SQL ELT optimization.
- You cannot use operators for unconnected lookups in a lookup expression. Use an additional expression transformation to include the operator.
- When you use the same column names in the source and unconnected lookup tables, and enable full SQL ELT optimization, the mapping fails with a duplicate column error. To run the mapping successfully, rename the incoming fields using a prefix or postfix and use the fields in the lookup expression.

Rules and guidelines for SQL transformation

Consider the following rules and guidelines for a SQL transformation:

- When you configure a SQL transformation in a mapping enabled for SQL ELT optimization and use the DATE_PART function in a query, the task fails. You can only configure a SQL transformation using the simple SELECT statement for any Redshift supported scalar function with the prescribed format.
- When you configure a SQL transformation in a mapping with user-defined functions that have date, decimal, or smallint data types, the mapping fails. As a workaround, configure user-defined functions in Redshift only with the corresponding transformation data types supported in Amazon Redshift.
- When you run a mapping with SQL transformation, and define user-defined functions (UDFs) with Unicode or special characters, enclose the schema and UDF in double quotes.
- When you run a mapping with SQL transformation having multiple select queries, the mapping fails. Amazon Redshift only supports SQL transformations with a single simple select query.
- When you use user-defined functions and scalar functions together in the same query to partially push down a mapping logic, the mapping fails if the target data type does not match with the source type. As a workaround, you can enable full SQL ELT optimization or define only scalar functions in the query.

CHAPTER 11

Data type reference

Data Integration uses the following data types in mappings and mapping 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 and Range
Bigint	Bigint	Signed eight-byte integer -9223372036854775808 to 9223372036854775807
Boolean	Small Integer	Logical Boolean (true/false)
Char	String	Fixed-length character string Maximum value is 4096 bytes.
Date	Timestamp	Calendar date (year, month, day) 4713 BC to 294276 AD

Amazon Redshift Data Type	Transformation Data Type	Description and Range
Decimal	Decimal	Exact numeric of selectable precision For mappings: Maximum precision is 28, scale is 27. For mappings in advanced mode: Maximum precision is 38, scale is 37. For more information about the behavior of decimal data types in mappings and mappings in advanced mode, see "Rules and guidelines for data types" on page 125.
Double Precision	Double	Double precision floating-point number 15 significant digits of precision
Integer	Integer	Signed four-byte integer -2147483648 to +2147483647
Real	Double	Single precision floating-point number 6 significant digits of precision
Smallint	Small Integer	Signed two-byte integer -32768 to +32767
Super	String	Stores semistructured data or documents as values. Maximum value is 16 MB For more information on super data type, see the AWS documentation.
Time[1]	Timestamp	Time (without time zone) 00:00:00.000000 to 23:59:59.999999
Timestamp	Timestamp	Date and time (without time zone) 4713 BC to 294276 AD (with up to a maximum of six digits of precision for fractional seconds)
Timestamptz[1]	String	Date, time, and time zone 4713 BC to 294276 AD (with up to a maximum of six digits of precision for fractional seconds)
Varchar	String	Variable-length character string with a user-defined limit Maximum value is 65535 bytes.

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

Rules and guidelines for data types

You need to follow certain rules and guidelines for some of the data types.

Time, datetime, timestamp, and timestampz

Consider the following rules and guidelines for data that contains the time, datetime, timestamp, or timestampz data types:

- When you run a mapping using the time data type, the task runs successfully, but the decimal values are truncated from the output. For the time data type, the SQL query `CallableStatement .setTime(java.sql.Time)` has precision support till seconds only.
- When you create a mapping for data that contains a time column and you edit the data type from time to timestamp, the mapping fails at run time.
- When you read from or write to Amazon Redshift, the Time data type appears appended with dummy values in the data preview for all the transformations in the mapping. All the downstream transformations also receive the appended dummy date values. However, the time values appear correctly only in an existing Amazon Redshift target.
- When you create a target at runtime and the incoming values contain time columns, the time column is converted to the timestamp data type in the target. If a time column is added to the source and you dynamically refresh the schema, the column is created with the timestamp data type in the target.
- If you configure an SQL transformation to call a stored procedure from Amazon Redshift, and you specify the time data type with fractional seconds as a parameter in the SQL transformation, the precision is truncated.
- When you specify SESSSTARTTIME variable in a query in a mapping task to return the datetime values, specify the query in the following format:
  `:select to_timestamp('$$$SESSSTARTTIME','YYYY-MM-DD HH24:MI:SS.MS')::TIMESTAMP as xz;`

  You must set the **DateTime Format String** in the Advanced session properties to *YYYY-MM-DD HH24:MI:SS.MS*.

  If you want to use the default format *MM/DD/YYY* for SESSIONSTARTTIME in the mapping task, you must use the following format in the query:

  `select to_timestamp('$$$SESSSTARTTIME','MM/DD/YYYY HH24:MI:SS.MS')::TIMESTAMP as xz;`
- When you use a Filter transformation to apply a filter to the date, timestamp, or time column, use the to_date() expression in following format. For example, `to_date('1970-01-01 23:05:06.123457','YYYY-MM-DD HH24:MI:SS.US'`. The source string format and the format string must match. Change the fractional part of the string based on the source values. For more information on the TO_DATE format strings, see the Transformation Language Reference Guide.
- When you read from an Amazon Redshift source using a simple filter on the time data type column and run a mapping, the task fails at runtime with the following error: `Operation failed: Invalid expression string for filter condition []`.
  As a workaround, use the advance filter to read from an Amazon Redshift source.
- When you refresh the dynamic schema on a timestampz column, the varchar data type is only propagated in the column value and not timestampz.
- When you write data of Date data type to an Amazon Redshift target and the Secure Agent and the Amazon Redshift cluster are in different time zones, the date logged in the target is one day behind of the actual date.
  For example, the date 2014-12-07 is written as 2014-12-06 in the target.

Decimal

Consider the following rules and guidelines for data that contains the decimal data type:

- The behavior of the decimal data type differs in mappings and mappings in advanced mode. Mappings support decimal max precision 28, while mappings in advanced mode supports max decimal precision of 38. In mappings, if the decimal data type exceeds 28 precision in the source, the numeric value for the decimal is rounded off after the16th precision and the remaining digits are replaced with zeroes in the target.
  For example, the value 2.31213563489578908888986998989899999990 from the source is rounded off to 2.31213563489578900000000000000000000000 in the target. However, in mappings in advanced mode, the decimal data from the source remains the same in the target.

  To resolve the issue in mappings, specify a precision that is less than or equal to 28 for the Decimal data type in the source table.
- When you read data that contains columns of the decimal data type, any scale changes that you configure for the decimal data type columns is considered by default. If you do not want to consider the changes in scale, you can set the JVM option *-honorDecimalScaleRedshift* value to *false* in the Secure Agent properties. By default, this property is set to true.
- When you write to an Amazon Redshift V2 target with the staging optimization property enabled for staging performance, and the source values have data of float data type, the decimal values in the target get truncated and rounded off to the nearest integer. To avoid this issue, in the system configuration details section, select the Type as DTM for the Data Integration service, and set the value for `DisableInfaDoubleHandlingForStaging` as yes.

Super

Consider the following guidelines for data that contains the super data type:

- You cannot condition in a simple filter or an uncached lookup for the super data type columns.
- By default, the precision of the super data type is 256. If you do not set the precision correctly in the mapping, the mapping fails due to data truncation or target overflow.
  When you enable the mapping for source SQL ELT optimization to write to a flat file target and you do not set the precision, the data gets truncated.
- When you create a new target at run time to write data that contains the super data type from a mapping enabled for SQL ELT optimization, the mapping fails if the precision value exceeds 65536 characters. Ensure that the incoming super data type is mapped to the super data type in the target. Otherwise, the Amazon Redshift super data type is mapped to the string transformation data type. Hence, the new target with the super type column is created as varchar. For varchar columns, the precision needs to be less than 65536.
  The same limitation for the super data type precision is applicable for the dynamic schema refresh operation.
- When you map the following data types to the super data type in the target, the COPY command fails:

  - char
  - varchar
  - date
  - time
  - timestamp
  - timestamptz

  The issue occurs because the super column in a Amazon Redshift table always expects the string data to be enclosed within double quotes while loading the data using COPY command. When you map the Char

and Varchar data types to the Super data type in the target, ensure that the incoming data contains double quotes (").

Note: The same issue is also observed for an Amazon S3 V2 source object in a mapping to write to Amazon Redshift target columns of the super data type.

- When you read from a single object source that contains super data type column values and define a simple filter on that column, the mapping is successful, but no records are generated for the columns in a mapping enabled with full SQL ELT optimization.
- When you run a mapping in advanced mode to read from a Amazon Redshift source and write to a target and the columns of Super data type contain null values, the null values are written as the string value "null" in the target.

CHAPTER 12

Troubleshooting

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

Troubleshooting for Amazon Redshift V2 Connector

Using the JVM option for upsert operation

For an upsert operation, you can set the JVM option `-DuseDeleteForUpsert=true` and connect all the fields of the target table. Use the following rules when you set the JVM option:

- When you set the JVM option, the Secure Agent deletes the records of the target table based on the primary key match between the source table and the target table. All the records of the source table are inserted into the target table. When you do not set the JVM option, the Secure Agent updates existing rows and inserts other rows as if marked for insert.
- For mappings in advanced mode, configure the JVM option in the **Spark Session Properties** when you create a mapping task.
- When you use a target table that has all the columns forming a composite key and you do not set an update column in the target, the mapping might fail with a warning. When you set the JVM option, the mapping task runs successfully.
- When you use a target table that has duplicate values in an update column, the target table has lesser number rows after you run the mapping. When you do not set the JVM option, the target table does not have lesser number of rows.
- When you use a source table that has duplicate values in a column and select the column as an update column on the target, an additional row is added to the target. When you do not set the JVM option, an additional row is not added to the target.
- For an upsert operation, when you use a table with column names that contain special characters and do not set the JVM option, the mapping fails.

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

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

NOT NULL columns with default values

Even when you do not map the NOT NULL columns that have default values in an Amazon Redshift target table, the insert, update, or upsert operation is successful and the default values for NOT NULL columns are used.

If you set the JVM option `-DRetainUnmappedNotNullColumnValidation` value to `true` in the Secure Agent properties, the operation is not successful and the default values for NOT NULL columns are not used.

How to implement the upsert operation using Amazon Redshift V2 Connector?

For information about implementing the upsert operation, see Implementing the upsert operation using Amazon Redshift V2 Connector.

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

For information about configuring AWS IAM authentication, see Configuring AWS IAM Authentication for Amazon Redshift and Amazon Redshift V2 Connectors.

How to connect to Amazon Redshift Serverless offered by Amazon Web Services (AWS) using the Amazon Redshift V2 connector?

For information about connecting to Amazon Redshift Serverless, see Using Amazon Redshift Serverless with Cloud Data Integration.

Invalid timestamp error occurs when a string data type is mapped to a time data type in a mapping

When you run a mapping enabled for SQL ELT optimization to write data from a string column that contains date, timestamp and time information and you want to process it with default date/time format to write to Redshift, we can make use of JVM property `-DHonorInfaDateFormat=true` for the Secure Agent.

To configure the JVM option in **Administrator**, perform the following steps:

1. Select **Administrator** > **Runtime Environments**.
2. On the **Runtime Environments** page, select the Secure Agent machine that runs the mapping.
3. Click **Edit**.
4. In the **System Configuration Details** section, select **Data Integration Server** as the Service and DTM as the **Type**.
5. Edit the JVM Option system property and set the value to `-DHonorInfaDateFormat=true`.
6. Click **Ok** to save the changes.

We also need to set `DateTime Format String` session property as per the input string format under advanced session properties in a mapping task.

Part III: Data Integration with Amazon Redshift Connector

This part contains the following chapters:

- Introduction to Amazon Redshift Connector, 131
- Amazon Redshift connections, 136
- Amazon Redshift sources and targets, 139
- Synchronization tasks with Amazon Redshift, 150
- Mappings and mapping tasks with Amazon Redshift, 157
- Amazon Redshift SQL ELT optimization, 168
- Data type reference, 178
- Troubleshooting, 180

CHAPTER 13

Introduction to Amazon Redshift Connector

You can use Amazon Redshift Connector to securely read data from or write data to Amazon Redshift. Amazon Redshift sources and targets represent records in Amazon Redshift.

You can create an Amazon Redshift connection and use the connection in synchronization tasks, mappings, and mapping tasks. When you use Amazon Redshift objects in synchronization tasks, mappings, and mapping tasks, you must configure properties specific to Amazon Redshift.

Create an Amazon Redshift connection to specify the location of Amazon Redshift sources, lookups, and targets you want to include in a task. You can switch the mapping to advanced mode to include transformations and functions that enable advanced functionality. When you run an Amazon Redshift synchronization task, mapping, or mapping task, the agent writes data to Amazon Redshift based on the workflow and Amazon Redshift connection configuration.

The 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 agent issues a copy command that copies data from Amazon S3 to the Amazon Redshift target table.

You can also read data from or write data to the Amazon Redshift cluster that reside in a Virtual Private Cloud (VPC). When you read data from or write data to Amazon Redshift, you can specify the Hosted Agent or the Secure Agent.

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

Note: Informatica recommends to use Amazon Redshift V2 Connector as the new features and enhancements are provided for Amazon Redshift V2 Connector.

Amazon Redshift Connector assets

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

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

- Mapping
- Synchronization task

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

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 the Amazon driver can communicate with Amazon Redshift.

Amazon Redshift Connector 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.

Administration of Amazon Redshift Connector

As a user, you can use Amazon Redshift 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 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 connection. Instead, you must create an Redshift Role Amazon Resource Name (ARN), add the minimal Amazon IAM 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 Connector, see “IAM authentication” on page 134

- 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 IAM policy for Amazon Redshift Connector.
- When you create a temporary table for an upsert, update, or delete operation in the local staging area, you must create the temporary table in the following format:

  ```
  RecordName + "_" + time-stamp + ProcessID + PartitionId
  ```

  Note: By default, you have the permission to create the temporary tables as you have the PUBLIC group membership. To deny the permission, revoke the TEMP permission from the PUBLIC group and allow the TEMP permission to specific or groups of individuals.

Configure Amazon Redshift Connector 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 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 a minimal Amazon IAM policy

Create an Amazon IAM policy and define the required permissions to stage the data in Amazon S3 when you want to read data from and write data to Amazon Redshift.

Use the following minimum required permissions to stage the data in Amazon S3:

- PutObject
- GetObject
- DeleteObject
- ListBucket
- GetBucketPolicy

You can use the following sample Amazon IAM policy:

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

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

The supported regions are:

- Asia Pacific (Mumbai)
- Asia Pacific (Seoul)
- Asia Pacific (Singapore)
- Asia Pacific (Sydney)
- Asia Pacific (Tokyo)
- AWS GovCloud
- Canada (Central)
- China (Bejing)
- EU (Ireland)
- EU (Frankfurt)
- 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 on an Amazon Elastic Compute Cloud (EC2) system. Use IAM authentication for secure and controlled access to Amazon Redshift resources when you run

Use IAM authentication when you want to run a on an EC2 system. Perform the following steps to configure IAM authentication:

1. Create a minimal Amazon IAM policy. For more information, see "Create a minimal Amazon IAM policy" on page 133.
2. Create the Amazon EC2 role. Associate the minimal Amazon IAM policy while creating the 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.

3. Create an EC2 instance. Assign the Amazon EC2 role that you created in step #2 to the EC2 instance.
4. Create the Amazon Redshift Role ARN for secure access to Amazon Redshift resources. Associate the minimal Amazon IAM policy while creating the Amazon Redshift role. 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.
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.
6. Install the on the EC2 system.

CHAPTER 14

Amazon Redshift connections

Create an Amazon Redshift connection to securely read data from or write data to Amazon Redshift. You can use Amazon Redshift connections to specify sources and targets in synchronization tasks, mappings, and mapping tasks.

Create a connection and associate it with a synchronization task, mapping, or mapping task. Define the source and target properties to read data from or write data to Amazon Redshift.

You can create an Amazon Redshift connection on the **Connections** page and use it in the Mapping Designer when you create a mapping or in the Synchronization Task wizard when you create a task. The connection becomes available to the entire organization.

Amazon Redshift connection properties

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

The following table describes the Amazon Redshift 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.
Schema	Amazon Redshift schema name. Default is public.
AWS Access Key ID	Optional. Amazon S3 bucket access key ID. To run tasks on Secure Agent installed on an EC2 system, you might leave the Access Key ID blank. To run tasks on Secure Agent that is not installed on an EC2 system, you must provide the Access Key ID.
AWS Secret Access Key	Optional. Amazon S3 bucket secret access key ID. To run tasks on Secure Agent installed on an EC2 system, you might leave the Secret Access Key blank. To run tasks on Secure Agent that is not installed on an EC2 system, you must provide the Secret Access Key.

Connection property	Description
Master Symmetric Key	Optional. Amazon S3 encryption key. Provide a 256-bit AES encryption key in the Base64 format.
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 connection URL.
Number of bytes needed to support multibytes for varchar	Applicable to Create Target. Reads the Varchar precision of the source table and creates the target table with 1x/2x/3x/4x times of the source precision to successfully write multibyte characters in the target table. **Note:** You cannot create a target table if the Varchar precision exceeds 65535 that is maximum allowed.

Note: When you test a connection, Secure Agent validates Redshift connection. Validation of AWS Access key and AWS Secret key requires the Amazon S3 bucket name present in the advanced source and target properties. Therefore, Secure Agent validates AWS Access key and AWS Secret key when a synchronization or mapping task is run.

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.

Note: You can only use an unauthenticated proxy server to connect to Amazon Redshift.

Contact your network administrator for the correct proxy settings.

Configuring proxy settings on Windows

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

Configuring proxy settings through Secure Agent Manager

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

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

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

4. Configure the following proxy server details:

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

5. Click **OK**.

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

Configuring the proxy settings through JVM Options

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

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

   For example,

   ```
   JVMOption1=-Dhttps.proxyHost=<proxy_server_hostname>
   JVMOption2=-Dhttps.proxyPort=8081
   ```

6. Click **Save**.

   The Secure Agent restarts to apply the settings.

Configuring proxy settings on Linux

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

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

1. Navigate to the following directory:

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

2. Update the `proxy.ini` file using the following command:

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

3. Restart the Secure Agent.

CHAPTER 15

Amazon Redshift sources and targets

You can assign the source and target properties for an Amazon Redshift connection.

Amazon Redshift sources

You can use an Amazon Redshift object as a source in a synchronization task, mapping, or mapping task. You can also use multiple related Amazon Redshift standard objects as sources in a synchronization task.

When you use Amazon Redshift source objects, you can select a standard object as the primary source, and then add child objects.

When you configure the advanced source properties, you configure properties specific to Amazon Redshift. 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 sources

The agent creates a staging file in the directory that you specify in the source properties. The synchronization tasks, mapping and mapping tasks stage data in a staging directory before reading data from Amazon Redshift. The agent deletes the staging files from the staging directory when the task completes.

You cannot configure a directory on Hosted Agent. The Hosted Agent creates a directory to stage data at a temporary location and deletes the staging files from the temporary location when the task completes.

To improve task performance, enable compression for staging files. Specify a staging directory with an appropriate amount of disk space for the volume of data that you want to process. Specify a directory path that is available on each Secure Agent machine in the runtime environment.

The applications create subdirectories in the staging directory based on the time that the task runs. Subdirectories use the following naming convention:

```
<staging directory>/infaRedShiftStaging<MMddHHmmssSSS+xyz>
```

Server-side encryption for Amazon Redshift sources

If you want Amazon Redshift to encrypt data while fetching the file from Amazon Redshift and staging the file to Amazon S3, you must enable server-side encryption.

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, you must specify the customer master key ID in the connection properties.

Note: The staging files in the Amazon S3 are deleted after the task is complete.

Data encryption in Amazon Redshift sources

You can encrypt data using the customer master key ID generated by AWS Key Management Service (AWS KMS) for server-side encryption.

You can select the type of the encryption in the **Encryption Type** field under the Amazon Redshift 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:

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.

If you enable the **Turn on S3 Client Side Encryption** property and select the **Encryption Type** as **SSE-S3**, the Amazon S3 encrypts the data using the master symmetric key for client-side encryption.

If you enable the **Turn on S3 Client Side Encryption** property and select the **Encryption Type** as **SSE-KMS**, the Amazon S3 encrypts the data using the customer master key ID generated by AWS Key Management Service for server-side encryption.

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

Client-side encryption for Amazon Redshift sources

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 sources, Amazon Redshift unloads the data in encrypted format, and then pushes the data to the Secure Agent. The Secure Agent writes the data to the target based on the task or mapping logic.

To enable client-side encryption, you must provide a master symmetric key in the connection properties and select the **Turn on S3 Client Side Encryption** option in the advanced target 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 Connector 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.

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 **UnloadOptions Property File** field. Enter the options in uppercase and delimit the options by using a The Unload command has the following options and default values:

The property file contains the Unload command options. Include the property file path in the **UnloadOptions Property File** field. For example:

```
C:\Temp\Redshift\unloadoptions.txt
```

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

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

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 **UnloadOptions 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 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 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 occurrences of the following characters:

- Linefeed \n
- Carriage return \r
- Delimiter character specified for the unloaded data
- Escape character \
- Single- or double-quote character

Default is 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 on 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`

ADDQUOTES

ADDQUOTES is implemented with the UNLOAD command by default. Do not specify the ADDQUOTES option in the advanced source properties. The Unload command adds quotation marks to each data field. With added quotation marks, the UNLOAD command can read data values that contain the delimiter. If double quote (") is a part of data, use ESCAPE to read the double quote as a regular character.

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: The recommended maximum number of partitions is 32. If you configure more than 32 partitions, the mapping task might fail with a memory buffer error.

Amazon Redshift targets

You can use an Amazon Redshift object as a single target in a synchronization task, mapping, or mapping task. You can also create an Amazon Redshift target based on the input source. When you use Amazon Redshift target objects, you can select a standard object as the primary source.

You can insert, update, upsert, and delete data from Amazon Redshift targets. An update or insert task writes an entire batch to an Amazon Redshift target if no errors occur within the batch. If an error occurs within a batch, the Secure Agent writes the entire batch to the error rows file.

When you configure the advanced target properties, you configure properties specific to Amazon Redshift. 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.

If a mapping includes a flat file or an Amazon Redshift target, you can choose to use an existing target or create a new target at run time. You must specify Amazon Redshift target object names in lowercase letters.

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 targets

The agent creates a staging file in the directory that you specify in the target properties. The synchronization tasks, mappings, and mapping tasks stage data in a staging directory before writing data to Amazon Redshift. You can configure the task to retain or delete staging files.

You cannot configure a directory on Hosted Agent. The Hosted Agent creates a directory to stage data at a temporary location and deletes the staging files from the temporary location when the task completes.

To improve task performance, enable compression for staging files. Specify a staging directory with an appropriate amount of disk space for the volume of data that you want to process. Specify a directory path that is available on each Secure Agent machine in the runtime environment.

The applications creates subdirectories in the staging directory based on the time that the task runs. 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 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 targets

If you want Amazon Redshift to encrypt data while uploading the .csv files to Amazon Redshift, you must enable server-side encryption. To enable server-side encryption, select Server Side Encryption as the encryption type in the advanced target properties on the **Schedule** page.

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 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 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.

To enable client-side encryption, you must provide a master symmetric key in the connection properties and select the **Turn on S3 Client Side Encryption** option in the advanced target 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 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:

Retain staging files

You can retain staging files on Amazon S3 after the 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 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 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 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 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 source to Amazon Redshift.

You can specify the Copy command options directly in the **CopyOptions Property File** field. Enter the options in uppercase and delimit the options by using a The Copy command has the following options and default values:

The property file contains the Copy command options. Include the property file path in the **CopyOptions Property File** field. For example:

```
C:\Temp\Redshift\copyoptions.txt
```

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 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 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.

AWS_IAM_ROLE

Specify the Amazon Redshift Role Resource Name (ARN) to run the on 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`

STATUPDATE

Optional. Governs computation and refresh of optimizer statistics at the end of a successful COPY command.

The behavior of STATUPDATE depends on how it is configured:

- If you specify STATUPDATE ON, the COPY command updates the statistics even if the table is empty or not. Ensure that the current user is either the table owner or a superuser for the COPY command to work.
- If you specify STATUPDATE OFF, the COPY command does not update the statistics.
- If you do not specify STATUPDATE, the COPY command requires the insert permissions.

Default is OFF.

Field mappings

The field mapping page displays key icons for primary key fields. When you configure field mappings, map all key fields and NOT NULL fields to successfully insert or upsert data to Amazon Redshift targets. Though Amazon Redshift enforces NOT NULL fields, it does not enforce key constraints.

The field mapping page displays key icons for primary key fields. Other Amazon Redshift key types are not marked. You must map a non-key field for update operation. If you use Amazon Redshift Identity fields in field mappings, map all available Identity fields or none. The Identity fields contain data that is automatically generated by Amazon Redshift.

You cannot map identity columns in a field map, if the identity column is not part of a key. If an identity column is part of a key, you must map the identity column in field map. However, you cannot set a value on the identity column from source.

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.

Working with large tables

You can upload or download a large object as a set of multiple independent parts.

Amazon Redshift Connector uses the AWS TransferManager API to upload a large object in multiple parts to Amazon S3. While downloading a large object, the Secure Agent downloads the object in multiple parts from the Amazon S3.

When the file size is more than 5 MB, you can configure multipart upload to upload object in multiple parts in parallel. You can choose to download the object in multiple parts in parallel when the file size of an Amazon S3 object is greater than 12 MB.

You can configure **Enable Downloading S3 Files in Multiple Parts** option in the advanced source properties. You can configure the **Part Size** and **TransferManager Thread Pool Size** options in the advanced target properties.

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 session. Success and error files are .csv files that contain row-level details. The Hosted Agent does not create success and error files after you run a session.

The Secure Agent generates a success file after you run a session. 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 cannot access the error rows file from the **All Jobs** page. You can manually delete the files that you no longer need.

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

- By default, a success rows file is generated in the following directory: `<Secure Agent installation directory>/apps/Data_Integration_Server/data/success`. You can specify a different directory with the **Success File Directory** advanced target option.
- The success rows file uses the following naming convention: `infa_rs_<operation>_<schema.table_name>.batch_<batch_number>_file_<file_number>_<timestamp>_success.csv`.

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

- By default, an error rows file is generated in the following directory: `<Secure Agent installation directory>/apps/Data_Integration_Server/data/error`. You can specify a different directory with the **Error File Directory** advanced target option.

- 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`.
- For insert tasks, the error rows file uses the following naming convention: `infa_rs_<operation>_<schema.table>.batch_<batch_number>_file_<file_number>_<timestamp>_error.csv`. For upsert tasks, the error rows file uses the following naming convention: `infa_rs_<operation>_<schema.table>_<timestamp_inLong>.batch_<batch_number>_file_<file_number>_<timestamp>_error.csv`.

CHAPTER 16

Synchronization tasks with Amazon Redshift

In this section, you can understand how to configure source, target, and lookup properties for synchronization tasks.

Amazon Redshift sources in synchronization tasks

You configure Amazon Redshift source properties on the **Source** page of the Synchronization Task wizard.

To optimize performance, you can configure a filter in the **Data Filters** page. Configure a simple filter or an advanced filter to remove rows at the source. You can improve efficiency by filtering early in the data flow. A simple filter includes a field name, operator, and value. Use an advanced filter to define a more complex filter condition, which can include multiple conditions using the AND or OR logical operators.

The following table describes the Amazon Redshift source properties:

Property	Description
Connection	Name of the source connection.
Source Type	Type of the source object. Select Single or Multiple.
Source Object	Name of the source object. Select the source object for a single source or multiple related sources.

When you configure a synchronization task to use an Amazon Redshift source, you can configure advanced source properties. Advanced source properties appear on the **Schedule** page of the Synchronization Task wizard.

The following table describes the Amazon Redshift advanced source properties:

Advanced Property	Description
S3 Bucket Name	Amazon S3 bucket name for the Amazon Redshift source data. Use an S3 bucket in the same region as your Amazon Redshift cluster.
Enable Compression	Compresses staging files before writing the files to Amazon Redshift. 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 each Secure Agent machine in the runtime environment. When you run a task in Hosted Agent runtime environment, leave the staging directory location blank. The Hosted Agent creates a directory at a temporary location.
UnloadOptions Property File	Unload command options. Add options to the Unload command to write data from an Amazon Redshift object to an S3 bucket. You can add the following options: - DELIMITER - PARALLEL - ESCAPE - AWS_IAM_ROLE When you run a task in the Secure Agent runtime environment, either specify the path of the property file that contains the unload options or specify the unload options directly in the **UnloadOptions Property File** field. When you run a task in the Hosted Agent runtime environment, specify options directly in the **UnloadOptions Property File** field.
Turn on S3 Client Side Encryption	Indicates that the Secure Agent encrypts data by using a private encryption key.
Encryption Type	Select the source encryption type. You can select from the following encryption types: - SSE-S3 - SSE-KMS Default is **SSE-S3**. For more information, see "Data encryption in Amazon Redshift sources" on page 140 .
Enable Downloading 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 5 MB, you can choose to download the object in multiple parts in parallel.
Part Size	Specifies the part size of an object. Default is 5 MB.
Infa Advanced Filter	Not applicable for Amazon Redshift Connector.
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	Overrides the default query. Enclose column names in double quotes. The SQL query is case sensitive. Specify an SQL statement supported by the Amazon Redshift database.

Advanced Property	Description
Number of Sorted Ports	Number of columns used when sorting rows queried from the source. The agent adds an ORDER BY clause to the default query when it reads source rows. The ORDER BY clause includes the number of ports specified, starting from the top of the transformation. When you specify the number of sorted ports, the database sort order must match the session sort order. Default is 0.
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.
Source Table Name	You can override the default source table name.

Amazon Redshift targets in synchronization tasks

You can use an Amazon Redshift object as a target in a synchronization task.

When you use Amazon Redshift target objects, you can select a standard object as the primary source.

You can configure Amazon Redshift target properties on the **Target** page of the Synchronization Task wizard.

The following table describes the Amazon Redshift target properties:

Property	Description
Connection	Name of the target connection.
Target Object	Name of the target object. Select the primary target object.
Create Target	Creates a target. Enter a name for the target object and select the source fields that you want to use. Default name is the source object name and by default, all source fields are used. Optionally, enter a file extension for the target object.

When you configure a synchronization task to use Amazon Redshift targets, you can configure advanced target properties.

The following table shows the Amazon Redshift advanced target properties:

Property	Description
S3 Bucket Name	Amazon S3 bucket name for the Amazon Redshift target data. Use an S3 bucket in the same region as your Amazon Redshift cluster.
Enable Compression	Compresses staging files before writing the files to Amazon Redshift. The performance of the synchronization task improves when the Secure Agent compresses the staging files. Default is selected.

Property	Description
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 each Secure Agent machine in the runtime environment. When you run a task in Hosted Agent runtime environment, leave the staging directory blank. The Hosted Agent creates a directory at a temporary location.
Batch Size	Minimum number of rows in a batch. Enter a number greater than 0. Default is 2000000.
Max Redshift 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 rows file. Default is 1.
Truncate Target Table Before Data Load	Truncates an Amazon Redshift target before writing data to the target.
Null value for CHAR and VARCHAR data types	String value used to represent null values in CHAR and VARCHAR fields in Amazon Redshift targets, such as NULL or a space character. Default is an empty string.
Wait time in seconds for file consistency on S3	Number of seconds to wait for the Secure Agent to make the staging files available. Default is 5.
CopyOptions Property File	Copy command options. Add options to the Copy command to write data from an Amazon S3 bucket to Amazon Redshift target. You can add the following options: - DELIMITER - ACCEPTINVCHARS - QUOTE - COMPUPDATE - AWS_IAM_ROLE When you run a task in the Secure Agent runtime environment, either specify the path of the property file that contains the copy options or specify the copy options directly in the **CopyOptions Property File** field. When you run a task in the Hosted Agent runtime environment, you must specify options directly in the **CopyOptions Property File** field.
Turn on S3 Server Side Encryption	Indicates that Amazon S3 encrypts data during upload and decrypts data at the time of access.
Turn on 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 runtime environment ignores the server side encryption.
Vacuum Target Table	Recovers disk space and sorts rows 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.

Property	Description
Analyze Target Table	Improve the efficiency of the read and write operations. The query planner on Amazon Redshift updates the statistical metadata to build and choose optimal plans to improve the efficiency of queries.
Prefix for Retaining 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`.
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.
Target Table Name	You can override the default target table name.
Part Size	Specifies the part size of an object. Default is 5 MB.
TransferManager Thread Pool Size	Specifies the number of the threads to write data in parallel. Default is 10.
Number of Files per Batch	Provide the number of files to calculate the number of the target staging file per batch. If you do not provide a value, the number of the target staging file is calculated internally. **Note:** Specify a minimum value based on the cluster type and number of nodes in the Amazon Redshift cluster. To avoid errors, specify a value lesser than 1500.
Success File Directory	Directory for the Amazon Redshift success rows file. Specify a directory path that is available on each Secure Agent machine in the runtime environment. By default, Data Integration writes the success file to the following directory: `<Secure Agent installation directory>/apps/Data_Integration_Server/data/success` The Hosted Agent does not create a success rows file. Leave the **Success File Directory** field blank when you run a task in the Hosted Agent runtime environment.
Error File Directory	Directory for error rows file. Specify a directory path that is available on each Secure Agent machine in the runtime environment. By default, Data Integration writes the error rows file to the following directory: `<Secure Agent installation directory>/apps/Data_Integration_Server/data/error` When you specify the default error file directory you can download the error file from the schedule tab. When the Error File Directory is other than the default error directory, you cannot download the error file from the Schedule tab. You must go to the specified directory to retrieve the error file. The Hosted Agent does not create an error rows file. Leave the **Error File Directory** field blank when you run a task in the Hosted Agent runtime environment.

Amazon Redshift lookups in synchronization tasks

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

Note: Amazon Redshift Connector does not support un-connected lookup transformation.

Rules and guidelines for synchronization tasks

Consider the following rules and guidelines when you use synchronization tasks for Amazon Redshift sources and targets:

- If you use multiple Amazon Redshift source objects in a synchronization task, the **Refresh Fields** option in the **Field Mapping** tab does not display the metadata for the source objects. You can save and run the task, but the task fails at run time. Hence, you must create a new synchronization task with a single source object, or use the mapping task.
- Do not use single quotes in the `WHERE` clause of a filter or query.

Synchronization task example

You work for an e-commerce organization that stores sales order details in a MySQL database. Your organization needs to move the data from the MySQL database to an Amazon Redshift target.

Configure a synchronization task to write to Amazon Redshift.

You perform the following synchronization tasks:

Define the synchronization task.

Configure a synchronization task to use the insert operation.

Use a MySQL source object.

The source for the mapping is a MySQL connection that contains the sales order details. The MySQL object is a single source in the synchronization task. You can include the Customer ID, Item_codes, Item_quantity, and Price columns. Specify *sales_order_details* as the resource for the source object.

Create an Amazon Redshift target object.

Select the fields *Customer_ID*, *Item_codes*, *Item_quantity*, and *Price* from the source object that you want to insert into the target object. Provide a name *sales_order_details* for the target object and specify the connection type as MySQL. The synchronization task writes the data to Amazon Redshift. You can also use an existing target object.

Configure a field mapping.

Map all the fields under *sales_order_details* source data to all the fields in the target *sales_order_details*. The synchronization application writes the mapped source data to Amazon Redshift.

Configure the advanced target properties.

In the advanced target properties, you choose properties that are specific to Amazon Redshift. Specify an Amazon S3 bucket name for the Amazon Redshift target data. Use an S3 bucket in the same region

as your Amazon Redshift cluster. You can also specify options for the copy command, and turn on server side and client side encryption. Click **Save** and **Finish** the task.

Open Amazon Redshift to visualize the exported data.

Schedule the task.

You can schedule the task for each requirement and save. You can select the synchronization task from the **Explore** page and run the task. In **Monitor**, you can monitor the status of the logs after you run the task.

CHAPTER 17

Mappings and mapping tasks with Amazon Redshift

Create a mapping task to process data based on the data flow logic defined in a mapping or integration template. You can also create a mapping task to capture changed data from the Oracle CDC source and write the changed data to an Amazon Redshift target table. You can switch the mapping to advanced mode to include transformations and functions that enable advanced functionality.

Caution: The pre-SQL and post-SQL commands configured in the Source, Target, and Lookup transformations are not validated at run time. Ensure that the command syntax is free from potentially unsafe content and the database user you configured in the connection object has appropriate privileges. It is recommended to use TLS- enabled database connections for secure network communications.

Amazon Redshift sources in mappings

In a mapping, you can configure a Source transformation to represent a single Amazon Redshift source or multiple Amazon Redshift sources.

You can use multiple related Amazon Redshift standard objects as a source. You can select a standard object as the primary source, then you add one or more child objects.

The following table describes the Amazon Redshift 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. Select the source object for a single source.

The following table describes the Amazon Redshift query options that you can configure in a Source transformation:

Property	Description
Filter	Filter value in a read operation. Click **Configure** to add conditions to filter records and reduce the number of rows that the Secure Agent reads from the source. You can specify the following filter conditions: - **Not parameterized**. Use a basic filter to specify the object, field, operator, and value to select specific records. - **Completely parameterized**. Use a parameter to represent the field mapping. - **Advanced**. Use an advanced filter to define a more complex filter condition that uses the Amazon Redshift query format.
Sort	Not applicable.

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

Advanced Property	Description
S3 Bucket Name	Amazon S3 bucket name for the Amazon Redshift target data. Use an S3 bucket in the same region as your Amazon Redshift cluster.
Enable Compression	Compresses staging files before writing the files to Amazon Redshift. Task performance improves when the runtime environment 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 each Secure Agent machine in the runtime environment. When you run a task in Hosted Agent runtime environment, leave the staging directory blank. The Hosted Agent creates a directory at a temporary location.
UnloadOptions Property File	Unload command options. Add options to the unload command to write data from an Amazon Redshift object to an S3 bucket. You can add the following options: - DELIMITER - PARALLEL - ESCAPE - AWS_IAM_ROLE When you run a task in the Secure Agent runtime environment, either specify the path of the property file that contains the unload options or specify the unload options directly in the **UnloadOptions Property File** field. When you run a task in the Hosted Agent runtime environment, specify options directly in the **UnloadOptions Property File** field.
Turn on S3 Client Side Encryption	Indicates that the Secure Agent encrypts data by using a private encryption key.
Encryption Type	Select the source encryption type. You can select from the following encryption types: - SSE-S3 - SSE-KMS Default is **SSE-S3**.

Advanced Property	Description
Enable Downloading 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 5 MB, you can choose to download the object in multiple parts in parallel.
Part Size	Specifies the part size of an object. Default is 5 MB.
Infa Advanced Filter	Not applicable for Amazon Redshift Connector.
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	Overrides the default query. Enclose column names in double quotes. The SQL query is case sensitive. Specify an SQL statement supported by the Amazon Redshift database.
Number of Sorted Ports	Number of columns used when sorting rows queried from the source. The agent adds an ORDER BY clause to the default query when it reads source rows. The ORDER BY clause includes the number of ports specified, starting from the top of the transformation. When you specify the number of sorted ports, the database sort order must match the session sort order. Default is 0.
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.
Source Table Name	You can override the default source table name.
Tracing Level	Sets the amount of detail that appears in the log file. You can choose terse, normal, verbose initialization, or verbose data. Default is normal.

Note: Do not use single quotes in the `WHERE` clause of a filter or query.

Configuring key range partitioning

Configure key range partitioning 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**.

   Note: The key range that you specify for partitioning must not contain a string value with a single quote.

Amazon Redshift targets in mappings

In a mapping, you can configure a Target transformation to represent a single Amazon Redshift target. You can also create an Amazon Redshift target at runtime based on the input fields.

When you use an Amazon Redshift target object, select a standard object as the primary target, and then add a child object. You can use a custom object as a single target.

The following table describes the Amazon Redshift 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 and select the source fields that you want to use. Default name is the source object name and by default, all source fields are used. Optionally, enter a file extension for the target object.

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

Property	Description
S3 Bucket Name	Amazon S3 bucket name for the Amazon Redshift target data. Use an S3 bucket in the same region as your Amazon Redshift cluster.
Enable Compression	Compresses staging files before writing the files to Amazon Redshift. Task performance improves when the runtime environment compresses the staging files. Default is selected.
Staging Directory Location	Amazon Redshift staging directory. For Secure Agent runtime environment, specify a directory path that is available on each Secure Agent machine in the runtime environment. For Hosted Agent runtime environment, leave the staging directory blank. The Hosted Agent creates a directory at a temporary location.
Batch Size	Minimum number of rows in a batch. Enter a number greater than 0. Default is 2000000.
Max Redshift 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 runtime environment writes the entire batch to the error rows file. Default is 1.

Property	Description
Truncate Target Table Before Data Load	Truncates an Amazon Redshift target before writing data to the target.
Null value for CHAR and VARCHAR data types	String value used to represent null values in CHAR and VARCHAR fields in Amazon Redshift targets, such as NULL or a space character. Default is an empty string.
Wait time in seconds for file consistency on S3	Number of seconds to wait for the runtime environment to make the staging files available. Default is 5.
CopyOptions Property File	Copy command options. Add options to the Copy command to write data from an Amazon S3 bucket to Amazon Redshift target. You can add the following options: - DELIMITER - ACCEPTINVCHARS - QUOTE - COMPUPDATE - AWS_IAM_ROLE When you run a task in the Secure Agent runtime environment, either specify the path of the property file that contains the copy options or specify the copy options directly in the **CopyOptions Property File** field. When you run a task in the Hosted Agent runtime environment, you must specify options directly in the **CopyOptions Property File** field.
Turn on S3 Server Side Encryption	Indicates that Amazon S3 encrypts data during upload and decrypts data at the time of access.
Turn on S3 Client Side Encryption	Indicates that the runtime environment encrypts data by using a private encryption key. If you enable both server side and client side encryption, the runtime environment ignores the server side encryption.
Vacuum Target Table	Recovers disk space and sorts rows 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.
Prefix for Retaining 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`.
Analyze Target Table	Improve the efficiency of the read and write operations. The query planner on Amazon Redshift updates the statistical metadata to build and choose optimal plans to improve the efficiency of queries.
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.

Property	Description
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.
Target Table Name	You can override the default target table name.
Part Size	Specifies the part size of an object. Default is 5 MB.
TransferManager Thread Pool Size	Specifies the number of the threads to write data in parallel. Default is 10.
Number of Files per Batch	Provide the number of files to calculate the number of the target staging file per batch. If you do not provide a value, the number of the target staging file is calculated internally. **Note:** Specify a minimum value based on the cluster type and number of nodes in the Amazon Redshift cluster. To avoid errors, specify a value lesser than 1500.
Success File Directory	Directory for the Amazon Redshift success rows file. Specify a directory path that is available on each Secure Agent machine in the runtime environment. By default, Data Integration writes the success file to the following directory: `<Secure Agent installation directory>/apps/Data_Integration_Server/data/success` The Hosted Agent does not create a success rows file. Leave the **Success File Directory** field blank when you run a task in the Hosted Agent runtime environment.
Error File Directory	Directory for the Amazon Redshift error rows file. Directory for error rows file. Specify a directory path that is available on each Secure Agent machine in the runtime environment. By default, Data Integration writes the error rows file to the following directory: `<Secure Agent installation directory>/apps/Data_Integration_Server/data/error` When you specify the default error file directory you can download the error file from the schedule tab. When the Error File Directory is other than the default error directory, you cannot download the error file from the Schedule tab. You must go to the specified directory to retrieve the error file. The Hosted Agent does not create an error rows file. Leave the **Error File Directory** field blank when you run a task in the Hosted Agent runtime environment.
Forward Rejected Rows	Determines whether the transformation passes rejected rows to the next transformation or drops rejected rows. By default, the mapping application forwards rejected rows to the next transformation.

When you edit a target task, selecting a different Amazon Redshift connection clears the advanced target properties. Enter the S3 bucket name and other advanced properties applicable to the selected Amazon Redshift connection.

Amazon Redshift lookups in mappings

In a mapping, you can configure a Lookup transformation to represent an Amazon Redshift object.

When you use an Amazon Redshift object as a lookup, you need to configure the Amazon S3 bucket name in Amazon Redshift properties.

When you use a cache lookup with Amazon Redshift connection, the lookup condition is ignored if the lookup condition contains a NULL value.

Note: Amazon Redshift Connector does not support un-connected lookup transformation.

Amazon Redshift objects in mapping tasks

When you configure a mapping task based on an integration template, you can configure advanced properties for Amazon Redshift sources and targets.

Amazon Redshift sources in mapping tasks

For Amazon Redshift source connections used in template-based mapping tasks, you can configure advanced properties in the **Sources** page of the Mapping Task wizard.

You can configure the following advanced properties:

Advanced Property	Description
S3 Bucket Name	Amazon S3 bucket name for the Amazon Redshift target data. Use an S3 bucket in the same region as your Amazon Redshift cluster.
Enable Compression	Compresses staging files before writing the files to Amazon Redshift. Task performance improves when the runtime environment 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 each Secure Agent machine in the runtime environment. When you run a task in Hosted Agent runtime environment, leave the staging directory blank. The Hosted Agent creates a directory at a temporary location.
UnloadOptions Property File	Unload command options. Add options to the unload command to write data from an Amazon Redshift object to an S3 bucket. You can add the following options: - DELIMITER - PARALLEL - ESCAPE - AWS_IAM_ROLE When you run a task in the Secure Agent runtime environment, either specify the path of the property file that contains the unload options or specify the unload options directly in the **UnloadOptions Property File** field. When you run a task in the Hosted Agent runtime environment, specify options directly in the **UnloadOptions Property File** field.

Advanced Property	Description
Turn on S3 Client Side Encryption	Indicates that the Secure Agent encrypts data by using a private encryption key.
Encryption Type	Select the source encryption type. You can select from the following encryption types: - SSE-S3 - SSE-KMS Default is **SSE-S3** For more information, see "Data encryption in Amazon Redshift sources" on page 140 .
Enable Downloading 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 5 MB, you can choose to download the object in multiple parts in parallel.
Part Size	Specifies the part size of an object. Default is 5 MB.
Infa Advanced Filter	Not applicable for Amazon Redshift Connector.
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	Overrides the default query. Enclose column names in double quotes. The SQL query is case sensitive. Specify an SQL statement supported by the Amazon Redshift database.
Number of Sorted Ports	Number of columns used when sorting rows queried from the source. The agent adds an ORDER BY clause to the default query when it reads source rows. The ORDER BY clause includes the number of ports specified, starting from the top of the transformation. When you specify the number of sorted ports, the database sort order must match the session sort order. Default is 0.
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.
Source Table Name	You can override the default source table name.
Tracing Level	Sets the amount of detail that appears in the log file. You can choose terse, normal, verbose initialization, or verbose data. Default is normal.

Amazon Redshift targets in mapping tasks

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

You can configure the following advanced properties:

Property	Description
S3 Bucket Name	Amazon S3 bucket name for the Amazon Redshift target data. Use an S3 bucket in the same region as your Amazon Redshift cluster.
Enable Compression	Compresses staging files before writing the files to Amazon Redshift. Task performance improves when the runtime environment compresses the staging files. Default is selected.
Staging Directory Location	Amazon Redshift staging directory. Specify a directory on the machine that hosts the runtime environment.
Batch Size	Minimum number of rows in a batch. Enter a number greater than 0. Default is 2000000.
Max Redshift 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 runtime environment writes the entire batch to the error rows file. Default is 1.
Truncate Target Table Before Data Load	Truncates an Amazon Redshift target before writing data to the target.
Null value for CHAR and VARCHAR data types	String value used to represent null values in CHAR and VARCHAR fields in Amazon Redshift targets, such as NULL or a space character. Default is an empty string.
Wait time in seconds for file consistency on S3	Number of seconds to wait for the runtime environment to make the staging files available. Default is 5.
CopyOptions Property File	Path to the property file. Enables you to add options to the copy command to write data from Amazon S3 to an Amazon Redshift target. You can add the following options: - DELIMITER - ACCEPTINVCHARS - QUOTE - COMPUPDATE When you run a task in the Secure Agent runtime environment, either specify the path of the property file that contains the copy options or specify the copy options directly in the **CopyOptions Property File** field. When you run a task in the Hosted Agent runtime environment, you must specify options directly in the **CopyOptions Property File** field.
Turn on S3 Server Side Encryption	Indicates that Amazon S3 encrypts data during upload and decrypts data at the time of access.
Turn on S3 Client Side Encryption	Indicates that the runtime environment encrypts data by using a private encryption key. If you enable both server side and client side encryption, the runtime environment ignores the server side encryption.

Property	Description
Vacuum Target Table	Recovers disk space and sorts rows 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.
Analyze Target Table	Improve the efficiency of the read and write operations. The query planner on Amazon Redshift updates the statistical metadata to build and choose optimal plans to improve the efficiency of queries.
Prefix for Retaining 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`.
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.
Target Table Name	You can override the default target table name.
Part Size	Specifies the part size of an object. Default is 5 MB.
TransferManager Thread Pool Size	Specifies the number of the threads to write data in parallel. Default is 10.
Number of Files per Batch	Provide the number of files to calculate the number of the target staging file per batch. If you do not provide a value, the number of the target staging file is calculated internally. **Note:** Specify a minimum value based on the cluster type and number of nodes in the Amazon Redshift cluster. Do not specify a large value. Else, the task fails with an `OutOfMemoryError` error message.

Property	Description
Success File Directory	Directory for the Amazon Redshift success rows file. Specify a directory path that is available on each Secure Agent machine in the runtime environment. By default, Data Integration writes the success file to the following directory: `<Secure Agent installation directory>/apps/Data_Integration_Server/data/success` The Hosted Agent does not create a success rows file. Leave the **Success File Directory** field blank when you run a task in the Hosted Agent runtime environment.
Error File Directory	Directory for the Amazon Redshift error rows file. Directory for error rows file. Specify a directory path that is available on each Secure Agent machine in the runtime environment. By default, Data Integration writes the error rows file to the following directory: `<Secure Agent installation directory>/apps/Data_Integration_Server/data/error` When you specify the default error file directory you can download the error file from the schedule tab. When the Error File Directory is other than the default error directory, you cannot download the error file from the Schedule tab. You must go to the specified directory to retrieve the error file. The Hosted Agent does not create an error rows file. Leave the **Error File Directory** field blank when you run a task in the Hosted Agent runtime environment.

CHAPTER 18

Amazon Redshift SQL ELT optimization

You can use SQL ELT optimization to push transformation logic to source databases or target databases. Use SQL ELT optimization when using database resources can improve task performance.

When you run a task configured for SQL ELT optimization, the task converts the transformation logic to an SQL query. The task sends the query to the database, and the database executes the query.

Amazon Redshift Connector supports **Full** and **Source** SQL ELT optimization for the ODBC connection type that uses Amazon ODBC Redshift drivers for mapping.

Note: You cannot configure an upsert operation in a mapping enabled for full SQL ELT optimization. You cannot configure SQL ELT optimization for a mapping in advanced mode.

Example

You work for a rapidly growing data science organization. Your organization develops software products to analyze financials, building financial graphs connecting people profiles, companies, jobs, advertisers, and publishers. The organization uses infrastructure based on Amazon Web Services and stores its data in Amazon Redshift, a petabytescale data warehouse. The organization plans to implement a business intelligence service to build visualization and perform real-time analysis. Therefore, you need to port the vast amount of data stored in Amazon Redshift to the business intelligence service. You can use Amazon Redshift Connector to read data from Amazon Redshift. To read this large amount of data, you can use source SQL ELT optimization for the ODBC connection type. Using the ODBC connection type with SQL ELT optimization enhances the performance.

Supported functions and operators for Amazon Redshift mappings

The following table summarizes the availability of functions in an Amazon Redshift database. Columns marked with an X indicate that the function can be pushed to the Amazon Redshift database by using source-side or full SQL ELT optimization. Columns marked with S indicate that the function can be pushed to the

Amazon Redshift database only by using source-side SQL ELT optimization. Columns marked with a dash (-) symbol indicate that the function cannot be pushed to the database.

Function	SQL ELT optimization	Function	SQL ELT optimization	Function	SQL ELT optimization
ABORT()	-	INSTR()	X	REG_REPLACE	-
ABS()	X	IS_DATE()	-	REPLACECHR()	-
ADD_TO_DATE()	X	IS_NUMBER()	-	REPLACESTR()	-
AES_DECRYPT()	-	IS_SPACES()	-	REVERSE()	-
AES_ENCRYPT()	-	ISNULL()	S	ROUND(DATE)	-
ASCII()	-	LAST()	-	ROUND(NUMBER)	X
AVG()	S	LAST_DAY()	X	RPAD()	X
CEIL()	X	LEAST()	-	RTRIM()	X
CHOOSE()	-	LENGTH()	X	SET_DATE_PART()	-
CHRCODE()	-	LN()	X	SIGN()	X
COMPRESS()	-	LOG()	-	SIN()	X
CONCAT()	X	LOOKUP	-	SINH()	-
COS()	X	LOWER()	X	SOUNDEX()	-
COSH()	-	LPAD()	X	SQRT()	X
COUNT()	S	LTRIM()	X	STDDEV()	S
CRC32()	-	MAKE_DATE_TIME()	-	SUBSTR()	X
CUME()	-	MAX()	S	SUM()	S
DATE_COMPARE()	X	MD5()	X	SYSTIMESTAMP()	S
DATE_DIFF()	X	MEDIAN()	-	TAN()	S
DECODE()	X	METAPHONE()	-	TANH()	-
DECODE_BASE64()	-	MIN()	S	TO_BIGINT	X
DECOMPRESS()	-	MOD()	S	TO_CHAR(DATE)	S
ENCODE_BASE64()	-	MOVINGAVG()	-	TO_CHAR(NUMBER)	X
EXP()	X	MOVINGSUM()	-	TO_DATE()	X
FIRST()	-	NPER()	-	TO_DECIMAL()	X

Function	SQL ELT optimization	Function	SQL ELT optimization	Function	SQL ELT optimization
FLOOR()	X	PERCENTILE()	-	TO_FLOAT()	X
FV()	-	PMT()	-	TO_INTEGER()	X
GET_DATE_PART()	X	POWER()	X	TRUNC(DATE)	S
GREATEST()	-	PV()	-	TRUNC(NUMBER)	S
IIF()	X	RAND()	-	UPPER()	X
IN()	S	RATE()	-	VARIANCE()	S
INDEXOF()	-	REG_EXTRACT()	-		
INITCAP()	X	REG_MATCH()	-		

The following table lists the operators that can be used in an Amazon Redshift database. Columns marked with an X indicate that the operator can be pushed to the Amazon Redshift database by using source-side, or full SQL ELT optimization.

Operator	SQL ELT optimization
+	X
-	X
*	X
/	X
%	X
||	X
>	X
=	X
>=	X
<=	X
!=	X
AND	X
OR	X
NOT	X
^=	X

Configuring Amazon Redshift ODBC connection

Amazon Redshift supports Amazon ODBC Redshift drivers on Windows and Linux systems. You must install the Amazon ODBC Redshift 64-bit driver based on your system requirement.

Note: Informatica certifies Amazon Redshift ODBC driver version, `AmazonRedshiftODBC-64-bit-1.4.8.1000-1.x86_64`, to use for SQL ELT optimization.

Configuring Amazon Redshift ODBC connection on Windows

Before you establish an ODBC connection to connect to Amazon Redshift on Windows, you must configure the ODBC connection.

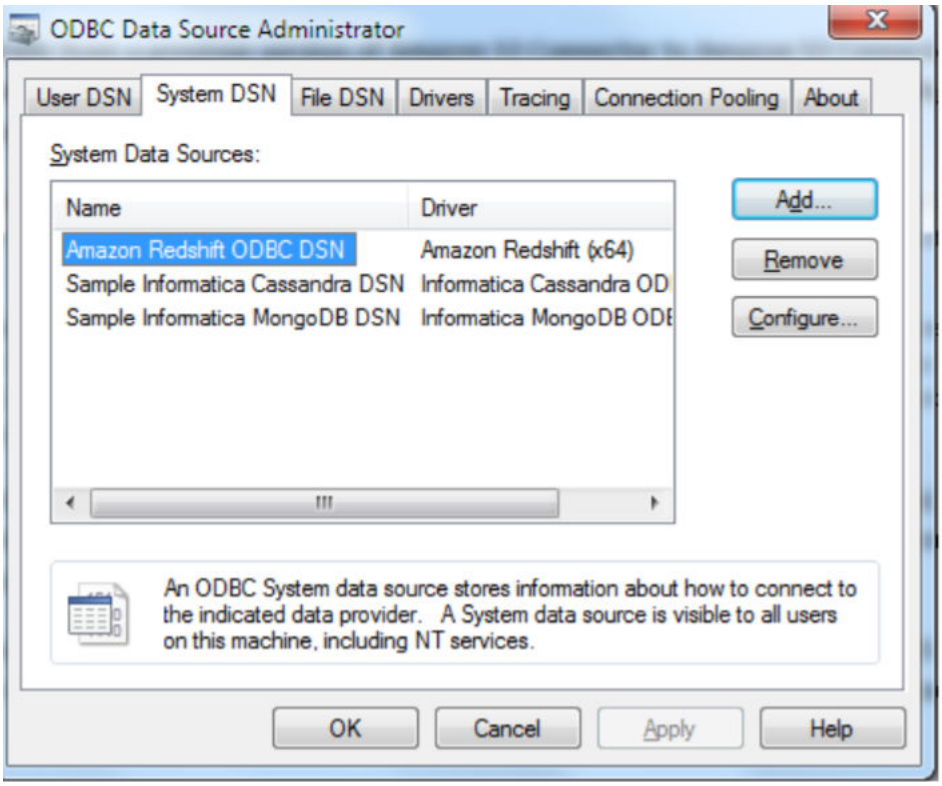
Perform the following steps to configure an ODBC connection on Windows:

1. Download the Amazon Redshift ODBC drivers from the AWS website.

   You must download the Amazon Redshift ODBC 64-bit driver.
2. Install the Amazon Redshift ODBC drivers on the machine where the Secure Agent is installed.
3. Open the folder in which ODBC data source file is installed.
4. Run the `odbcad32.exe` file.

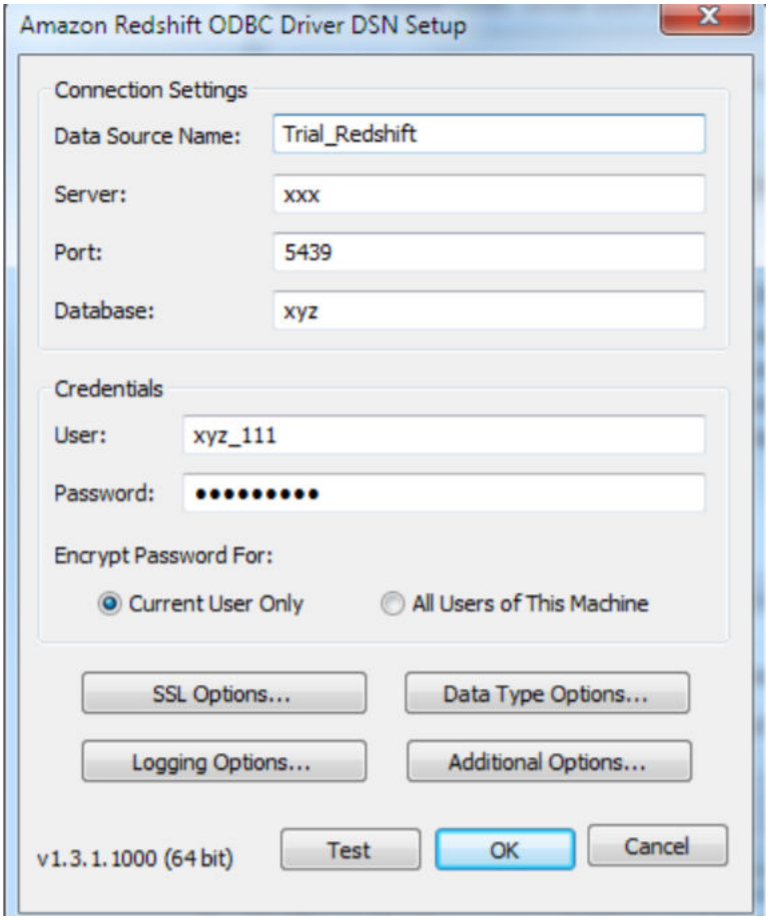
   The **ODBC Data Source Administrator** dialog box appears.
5. Click **System DSN**.

   The **System DSN** tab appears. The following image shows the **System DSN** tab on the **ODBC Data Source Administrator** dialog box:



6. Click **Configure**.

The **Amazon Redshift ODBC Driver DSN Setup** dialog box displays. The following image shows the **Amazon Redshift ODBC Driver DSN Setup** dialog box where you can configure the **Connection Settings** and **Credentials** section:



7. Specify the following connection properties in the **Connection Settings** section:

Property	Description
Data Source Name	Name of the data source.
Server	Location of the Amazon Redshift server.
Port	Port number of the Amazon Redshift server.
Database	Name of the Amazon Redshift database.

Note: You must specify the **Server**, **Port**, and **Database** values from the JDBC URL.

8. Specify the following credentials in the **Credentials** section:

Property	Description
User	User name to access the Amazon Redshift database.
Password	Password for the Amazon Redshift database.
Encrypt Password For	Encrypts the password for the following users: - **Current User Only** - **All Users of This Machine** Default is **Current User Only**.

9. Click **Test** to test the connection in the **Amazon Redshift ODBC Driver DSN Setup** box.
10. Click **OK**.

The Amazon Redshift ODBC connection is configured successfully on Windows.

After you configure the Amazon Redshift ODBC connection, you must create an ODBC connection to connect to Amazon Redshift.

For more information about how to create an ODBC connection to connect to Amazon Redshift, see "Creating an ODBC connection" on page 174

Configuring Amazon Redshift ODBC connection on Linux

Before you establish an ODBC connection to connect to Amazon Redshift on Linux, you must configure the ODBC connection.

Perform the following steps to configure an ODBC connection on Linux:

1. Download the Amazon Redshift ODBC drivers from the AWS website.

   You must download the Amazon Redshift ODBC 64-bit driver.
2. Install the Amazon Redshift ODBC drivers on the machine where the Secure Agent is installed.
3. Configure the `odbc.ini` file properties in the following format:

```
[ODBC Data Sources]
driver_name=dsn_name

[dsn_name]
Driver=path/driver_file

Host=cluster_endpoint
Port=port_number
Database=database_name
```

4. Specify the following properties in the `odbc.ini` file:

Property	Description
ODBC Data Sources	Name of the data source.
Driver	Location of the Amazon Redshift ODBC driver file.
Host	Location of the Amazon Redshift host.

Property	Description
Port	Port number of the Amazon Redshift server.
Database	Name of the Amazon Redshift database.

Note: You must specify the **Host**, **Port**, and **Database** values from the JDBC URL.

5. Set the following environment variables for the operating system:

Variable	Description
LD_LIBRARY_PATH	Directory where the Amazon Redshift ODBC driver is installed.
ODBCINI	Directory that contains the odbc.ini file.
ODBCHROME	ODBC installation directory.

6. Run the following command to export the `odbc.ini` file.

```
Export ODBCINI=/<odbc.ini file path>/odbc.ini
```

7. Restart the Secure Agent.

   The Amazon Redshift ODBC connection on Linux is configured successfully.

After you configure the Amazon Redshift ODBC connection, you must create an ODBC connection to connect to Amazon Redshift.
For more information about how to create an ODBC connection to connect to Amazon Redshift, see "Creating an ODBC connection" on page 174

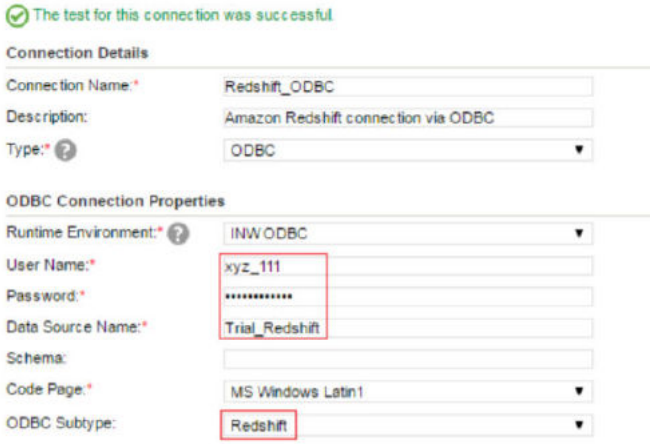
Creating an ODBC connection

You must create an ODBC connection to connect to Amazon Redshift after you configure the ODBC connection.

Perform the following steps to create an Amazon Redshift ODBC connection on the **Connections** page:

1. In Administrator, click **Connections**.
2. In the upper right corner, click **New Connections**.

   The **New Connection** page appears. The following image shows the **New Connection** page:



3. Configure the following connection details in the **Connection Details** section:

Property	Description
Connection Name	Name of the ODBC connection.
Description	Description of the connection.
Type	Type of the connection. Select the type of the connection as **ODBC**.

4. Configure the following connection details in the **Connection Properties** section:

Property	Description
Runtime Environment	The name of the runtime environment where you want to run the tasks.
User Name	User name of the Amazon Redshift account.
Password	Password for the Amazon Redshift account.
Data Source Name	Enter the name of the ODBC data source name that you created for the Amazon Redshift database.
Schema	Amazon Redshift schema name.
Code Page	Select the code page that the Secure Agent must use to read or write data.
ODBC Subtype	Enter the value of the **ODBC Subtype** field as **Redshift**.

The Amazon Redshift ODBC connection is created successfully.

Cross-schema SQL ELT optimization

You can configure cross-schema SQL ELT optimization for a mapping task that uses a Amazon Redshift ODBC connection to read or write data to Amazon Redshift objects of different schemas in the same database.

To use cross-schema SQL ELT optimization, create Amazon Redshift ODBC connections and specify the schema for the source and target connections. The source and target schemas must be different but must belong to the same database. Configure SQL ELT optimization for the mapping task and enable cross-schema SQL ELT optimization in the advanced session properties. By default, the **Enable cross-schema SQL ELT optimization** check box is selected.

Configuring cross-schema SQL ELT optimization for an Amazon Redshift mapping task

Perform the following steps to configure cross-schema SQL ELT optimization for an Amazon Redshift mapping task:

1. Create Amazon Redshift ODBC source and target connections, each defined with a different schema.

   For example,

   - Create a `rs_odbc1` Amazon Redshift ODBC connection and specify `CQA_SCHEMA1` schema in the connection properties.
   - Create a `rs_odbc2` Amazon Redshift ODBC connection and specify `CQA_SCHEMA2` schema in the connection properties.
2. Create an Amazon Redshift mapping.

   For example, create a `m_rs_pdo_crossSchema` Amazon Redshift mapping.
3. Add a Source transformation. Include an Amazon Redshift source object and connection to read data using the schema specified in the connection.

   For example, add a Source transformation. Include an Amazon Redshift source object and connection `rs_odbc1` to read data using `CQA_SCHEMA1`.
4. Add a Target transformation. Include an Amazon Redshift target object and connection to write data using the schema specified in the connection.

   For example, add a Target transformation. Include an Amazon Redshift target object and connection `rs_odbc2` to write data using `CQA_SCHEMA2`.
5. Create an Amazon Redshift mapping task, and perform the following tasks:

   a. Select the configured Amazon Redshift mapping.

      For example, select the `m_rs_pdo_crossSchema` Amazon Redshift mapping.

   b. On the **Runtime Options** tab, set **SQL ELT Optimization** to **Full**.

   c. Then, select **Enable cross-schema SQL ELT optimization**.

   d. Save the task and click **Finish**.

   When you run the mapping task, the Secure Agent reads data from the Amazon Redshift source object associated with the `CQA_SCHEMA1` schema and writes data to the Amazon Redshift target object associated with `CQA_SCHEMA2` schema.

Rules and guidelines for functions in SQL ELT optimization

Use the following rules and guidelines when pushing functions to an Amazon Redshift database:

- To push TRUNC(DATE) to Amazon Redshift, you must define the date and format arguments. Otherwise, the agent does not push the function to Amazon Redshift.
- The aggregator functions for Amazon Redshift accept only one argument, a field set for the aggregator function. The filter condition argument is not honored. In addition, make sure that all fields mapped to the target are listed in the GROUP BY clause.
- Do not specify a format for SYSTIMESTAMP() to push the SYSTIMESTAMP to Amazon Redshift. The Amazon Redshift database returns the complete time stamp.

- To push INSTR() to Amazon Redshift, you must only define string, search_value, and start arguments. Amazon Redshift does not support occurrence and comparison_type arguments.
- The flag argument is ignored when you push TO_BIGINT and TO_INTEGER to Amazon Redshift.
- The CaseFlag argument is ignored when you push IN() to Amazon Redshift.
- If you use the NS format as part of the ADD_TO_DATE() function, the agent does not push the function to Amazon Redshift.
- If you use any of the following formats as part of the TO_CHAR() and TO_DATE() functions, the agent does not push the function to Amazon Redshift:
  - NS
  - SSSS
  - SSSSS
  - RR
- To push TRUNC(DATE), GET_DATE_PART(), and DATE_DIFF() to Amazon Redshift, you must use the following formats:
  - D
  - DDD
  - HH24
  - MI
  - MM
  - MS
  - SS
  - US
  - YYYY

CHAPTER 19

Data type reference

Data Integration uses the following data types in synchronization tasks, mappings, and mapping 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 runtime environment uses to move data across platforms. Transformation data types appear in all transformations in synchronization tasks, mappings, and mapping tasks.

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 the runtime environment 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.
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.

Amazon Redshift Data Type	Transformation Data Type	Description
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 20

Troubleshooting

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

Troubleshooting for Amazon Redshift Connector

How to solve the task failure issue when you use ODBC connection to connect to Amazon Redshift to read UTF characters and Secure Agent is installed on Linux?

For information about the issue, see https://kb.informatica.com/solution/23/Pages/62/516325.aspx?myk=516325

What must be the maximum size of the local staging area, when the compression option for an Amazon Redshift connection to perform a read or write operation is enabled?

For information about the issue, see https://kb.informatica.com/faq/7/Pages/16/497540.aspx?myk=497540

How to configure AWS IAM Authentication for Amazon Redshift Connector?

For information about configuring AWS IAM authentication, see Configuring AWS IAM Authentication for Amazon Redshift and Amazon Redshift V2 Connectors.

Does the result of an Amazon Redshift task vary based on whether you map the target field that contains identity and primary key to the same column or different column?

For information about the issue, see https://kb.informatica.com/faq/7/Pages/21/535693.aspx?myk=535693

Troubleshooting Amazon Redshift connection

When you run a task to write data to an Amazon Redshift, the task fails with the following error:

```
Amazon_RedshiftWriter_30007 [ERROR] Copy command on record 'public.basic_data_types'
failed due to [ERROR: S3ServiceException:The bucket you are attempting to access
must be addressed using the specified endpoint. Please send all future requests to
this endpoint.,Status 301,Error PermanentRedirect,Rid A8BA401CC765AC53,ExtRid
NAbd1uxKirJVjDas1zo3WONdQ/+6p674RYkO
```

This issue occurs because the Amazon Redshift user and cluster in the connection properties are in a different region from the S3 bucket in the task.

You must configure the task to use an S3 bucket in the same region as the user and cluster in the connection. You can also use a different connection to write to the S3 bucket.

INDEX

A

administration
  IAM authentication 134
  minimal Amazon IAM policy 133
Amazon Redshift
  connection properties 136
  introduction 20
  mapping tasks 163
  spectrum 20, 60
  SQL ELT optimization overview 168
  SSL configuration 133
  troubleshooting connection 180
Amazon Redshift and transformation
  data types 123
Amazon Redshift connections
  administration 132
Amazon Redshift Connections
  overview 136
Amazon Redshift Connector
  data flow 12
Amazon Redshift Data Types
  data types 178
Amazon Redshift lookups
  mapping tasks 75, 78
Amazon Redshift ODBC connection
  configuration 171
Amazon Redshift sources
  mapping 157
  server-side encryption 140
  staging directory 139
Amazon Redshift targets
  mappings 160
  staging directory 143
Amazon Redshift V2
  connection properties 32
  sources 48
  supported task types 19
  targets 52
Amazon Redshift V2 connection
  configuration 107
Amazon Redshift V2 connections
  administration 59
  overview 21
Amazon Redshift V2 Connector
  overview 19
Amazon Redshift V2 sources
  mapping 62
Amazon Redshift V2 targets
  mappings 69
at-scale mapping 17

C

cache
  enable lookup cache 81
CDC source
  Amazon Redshift mapping task 82
Cloud Application Integration community
  URL 8
Cloud Developer community
  URL 8
Configuring Amazon Redshift ODBC connection
  Linux 173
Configuring Amazon Redshift ODBC Connection
  Windows 171
connections
  Amazon Redshift 136
  Amazon Redshift V2 32
copy command
  option 53
  overview 53
Creating an ODBC
  connection 174
cross-schema
  SQL ELT optimization 175

D

Data Integration community
  URL 8
data type reference
  overview 123, 178
dynamic schema handling 85

E

encryption type 49, 140

F

field mapping
  Amazon Redshift lookups 155
flat file
  staging data 87

I

IAM authentication
  administration 134
Informatica Global Customer Support
  contact information 9
Informatica Intelligent Cloud Services
  web site 8

introduction
amazon redshift introduction 132

K

Key Range Partition
configuration 86
key range partitioning 51

L

lookup cache
dynamic 81
persistent 81
Lookup transformation
lookup caching 81

M

maintenance outages 9
mapping
Oracle CDC Sources 16
mapping and mapping task 16
Mapping task
target properties 164
targets 164
source properties 163
sources 163
mappings
Amazon Redshift lookups 163
mappings in advanced mode
example 83

O

octal values
DELIMITER 57, 147
QUOTE 57, 147

P

preserve record order on write
target property 57

R

recovery and restart processing 56

S

source partitioning 51
SQL ELT optimization
functions 105, 106, 168
preview 99
Rules and guidelines for functions 115
transformations 105–107, 168
SQL ELT optimization preview 99
SQL transformations
configuration 84, 85
staging data
flat file 87
status
Informatica Intelligent Cloud Services 9
success and error files 58
synchronization
example 155
synchronization task 16
Synchronization task
source 150
source properties 150
target 152
target properties 152
synchronization tasks
Amazon Redshift lookups 155
system status 9

T

transformations
SQL ELT optimization 107
troubleshooting
Amazon Redshift V2 Connector 128
trust site
description 9

U

unload command
options 50
overview 49
upgrade notifications 9

W

web site 8