



Informatica® PowerExchange for Amazon
Redshift

10.5.6

User Guide for PowerCenter

© Copyright Informatica LLC 2014, 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.

Informatica, the Informatica logo, PowerCenter, and PowerExchange 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.

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.

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

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: 2024-05-29

Table of Contents

Preface	5
Informatica Resources.	5
Informatica Network.	5
Informatica Knowledge Base.	5
Informatica Documentation.	5
Informatica Product Availability Matrices.	6
Informatica Velocity.	6
Informatica Marketplace.	6
Informatica Global Customer Support.	6
 Chapter 1: Introduction to PowerExchange for Amazon Redshift.....	 7
PowerExchange for Amazon Redshift Overview.	7
PowerCenter Integration Service and Amazon Redshift Integration.	8
Introduction to Amazon Redshift.	9
 Chapter 2: PowerExchange for Amazon Redshift Configuration.....	 10
PowerExchange for Amazon Redshift Configuration Overview.	10
Prerequisites.	10
Configuring Custom Property.	11
IAM Authentication.	11
Create a minimal Amazon IAM policy.	12
Registering the Plug-in.	13
Configure the Java Heap Memory.	13
 Chapter 3: Amazon Redshift Sources and Targets.....	 14
Amazon Redshift Sources and Targets.	14
Import Amazon Redshift Objects.	14
Amazon Redshift Lookup Transformation.	16
 Chapter 4: Amazon Redshift Mappings.....	 17
Amazon Redshift Mappings Overview.	17
Configuring the Source Qualifier.	17
Amazon Redshift Mapping Example.	18
 Chapter 5: Amazon Redshift Pushdown Optimization.....	 19
Amazon Redshift Pushdown Optimization Overview.	19
Configuring Amazon Redshift ODBC Connection.	19
Configuring Amazon Redshift ODBC Connection on Windows.	20
Configuring Amazon Redshift ODBC Connection on Linux.	23
Creating an Amazon Redshift ODBC Connection.	24

Pushdown Optimization Functions.	26
Rules and Guidelines in Pushdown Optimization.	28
Chapter 6: Amazon Redshift Sessions.....	30
Amazon Redshift Sessions Overview.	30
Amazon Redshift Connections.	31
PowerExchange for Amazon Redshift Connections.	31
Configuring an Amazon Redshift Connection.	32
Configuring the Source Qualifier.	32
Amazon Redshift Source Sessions.	32
Amazon Redshift Staging Directory for Amazon Redshift Sources.	33
Identity Columns.	34
Unload command.	35
Partitioning.	36
Amazon Redshift Source Session Configuration.	38
Amazon Redshift Target Sessions.	40
Amazon Redshift Staging Directory for Amazon Redshift Targets.	40
Data Encryption in Amazon Redshift Targets.	40
Vacuum Tables.	41
Retain staging files.	42
Copy command.	42
Amazon Redshift Target Session Configuration.	44
Working with Large Tables.	47
Octal values as DELIMITER and QUOTE.	48
Success and Error Files.	48
Success Files.	48
Error Files.	49
Appendix A: Amazon Redshift Data Type Reference.....	50
Data Type Reference Overview.	50
Amazon Redshift and Transformation Data Types.	50
Appendix B: Troubleshooting.....	52
Troubleshooting Overview.	52
Troubleshooting for PowerExchange for Amazon Redshift.	52
Index.....	53

Preface

Use the Informatica® PowerExchange® for Amazon Redshift User Guide to learn how to read from or write to Amazon Redshift by using PowerCenter Client. Learn to create an Amazon Redshift connection, develop mappings, and run sessions in an Informatica domain.

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 Network

The Informatica Network is the gateway to many resources, including the Informatica Knowledge Base and Informatica Global Customer Support. To enter the Informatica Network, visit <https://network.informatica.com>.

As an Informatica Network member, you have the following options:

- Search the Knowledge Base for product resources.
- View product availability information.
- Create and review your support cases.
- Find your local Informatica User Group Network and collaborate with your peers.

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 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 Product Availability Matrices

Product Availability Matrices (PAMs) indicate the versions of the operating systems, databases, and types of data sources and targets that a product release supports. You can browse the Informatica PAMs at <https://network.informatica.com/community/informatica-network/product-availability-matrices>.

Informatica Velocity

Informatica Velocity is a collection of tips and best practices developed by Informatica Professional Services and based on real-world experiences from hundreds of data management projects. Informatica Velocity represents the collective knowledge of Informatica consultants who work with organizations around the world to plan, develop, deploy, and maintain successful data management solutions.

You can find Informatica Velocity resources at <http://velocity.informatica.com>. If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at ips@informatica.com.

Informatica Marketplace

The Informatica Marketplace is a forum where you can find solutions that extend and enhance your Informatica implementations. Leverage any of the hundreds of solutions from Informatica developers and partners on the Marketplace to improve your productivity and speed up time to implementation on your projects. You can find the Informatica Marketplace at <https://marketplace.informatica.com>.

Informatica Global Customer Support

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

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link:

<https://www.informatica.com/services-and-training/customer-success-services/contact-us.html>.

To find online support resources on the Informatica Network, visit <https://network.informatica.com> and select the eSupport option.

CHAPTER 1

Introduction to PowerExchange for Amazon Redshift

This chapter includes the following topics:

- [PowerExchange for Amazon Redshift Overview, 7](#)
- [PowerCenter Integration Service and Amazon Redshift Integration, 8](#)
- [Introduction to Amazon Redshift, 9](#)

PowerExchange for Amazon Redshift Overview

You can use PowerExchange for Amazon Redshift to read data from or write data to Amazon Redshift. You can also read data from or write data to the Amazon Redshift cluster that reside in a Virtual Private Cloud (VPC).

Use PowerExchange for Amazon Redshift to read data from Amazon Redshift views. An Amazon Redshift views contain information about the functioning of the Amazon Redshift system. You can run a query on views like you run a query on database tables.

You can use Amazon Redshift objects as sources and targets in mappings. When you use Amazon Redshift objects in mappings, you must configure properties specific to Amazon Redshift.

You can also configure an SSL connection to connect to Amazon Redshift. The PowerCenter Integration Service uses the Amazon driver to communicate with Amazon Redshift.

Note: PowerExchange for Amazon Redshift does not support real-time processing.

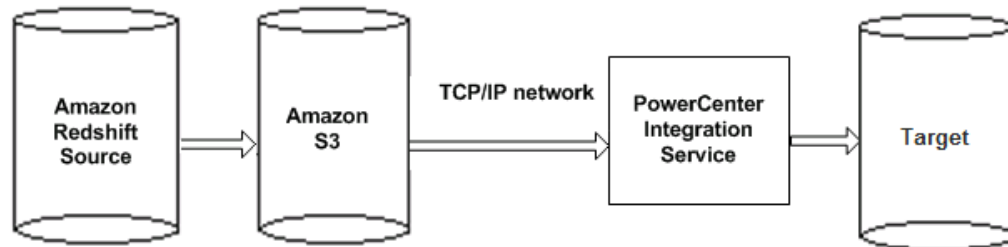
Example

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.

PowerCenter Integration Service and Amazon Redshift Integration

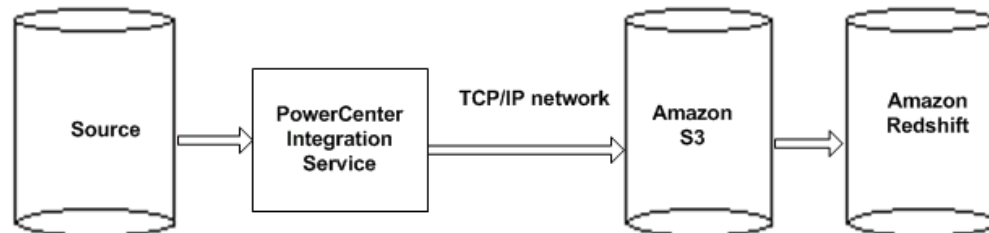
The PowerCenter Integration Service uses the Amazon Redshift connection to connect to Amazon Redshift.

The following image shows how PowerCenter connects to Amazon Redshift to read data:



When you run the Amazon Redshift session, the PowerCenter Integration Service reads data from Amazon Redshift based on the workflow and Amazon Redshift connection configuration. The PowerCenter Integration Service connects and reads data from Amazon Simple Storage Service (Amazon S3) through a TCP/IP network. The PowerCenter Integration Service then stores data in a staging directory on the PowerCenter machine. Amazon S3 is a storage service in which you can copy data from source and simultaneously move data to any target. The PowerCenter Integration Service issues a copy command that copies data from Amazon S3 to the target.

The following image shows how PowerCenter connects to Amazon Redshift to write data:



When you run the Amazon Redshift session, the PowerCenter Integration Service writes data to Amazon Redshift based on the workflow and Amazon Redshift connection configuration. The PowerCenter Integration Service stores data in a staging directory on the PowerCenter machine. The PowerCenter Integration Service then 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 source and simultaneously move data to Amazon Redshift clusters. The PowerCenter Integration Service issues a copy command that copies data from Amazon S3 to the Amazon Redshift target table.

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.

CHAPTER 2

PowerExchange for Amazon Redshift Configuration

This chapter includes the following topics:

- [PowerExchange for Amazon Redshift Configuration Overview, 10](#)
- [Prerequisites, 10](#)
- [Configuring Custom Property, 11](#)
- [IAM Authentication, 11](#)
- [Registering the Plug-in, 13](#)
- [Configure the Java Heap Memory, 13](#)

PowerExchange for Amazon Redshift Configuration Overview

You can use PowerExchange for Amazon Redshift on Windows or Linux. You must configure PowerExchange for Amazon Redshift before you can extract data from or load data to Amazon Redshift.

Prerequisites

Before you can use PowerExchange for Amazon Redshift, perform the following tasks:

1. Install or upgrade to PowerCenter.
2. Verify that you can connect to Amazon Redshift with an SQL client that uses the PostgreSQL driver. For example, you can use SQL Workbench/J to connect to Amazon Redshift.
3. Verify that you have read, write, and execute permissions on the following directories:
`<PowerCenter Installation Directory>/server/bin`
4. Verify that you have read and write permissions on the following directories:
`<PowerCenter Client Installation Directory>/Clients/PowerCenterClient/client/bin`

The organization administrator must also perform the following tasks:

- Mandatory. Get the user name and password to connect to the Amazon Redshift database.
- Mandatory. Get the Amazon Redshift JDBC URL.
- Mandatory. Manage Authentication. Use either of the following two methods:
 - Create an Access Key ID and Secret Access Key. Applicable when the PowerCenter client is not installed on Amazon Elastic Compute Cloud (EC2) system. Provide the values for Access Key ID and Secret Access Key when you configure the Amazon Redshift connection.
 - Configure AWS Identity and Access Management (IAM) Authentication. For enhanced security. Applicable when you install the PowerCenter client on Amazon EC2 system and you want to run sessions on the EC2 system. 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 S3 bucket policy to the Redshift Role ARN, and add the Redshift Role ARN to the Redshift cluster. For more information, see *IAM Authentication* section. Provide the Redshift Role ARN in the AWS_IAM_ROLE option in the UNLOAD and COPY commands when you create a session.

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.
- Optional. Create an Amazon Redshift master symmetric key to enable client-side encryption.
- Optional. Create an AWS Key Management Service (AWS KMS)-managed customer master key to enable server-side encryption.

Configuring Custom Property

You can configure the custom properties for Amazon Redshift targets.

You can set the custom property to run the copy command serially. Set the value of the `explicitCopyCommit` custom property to true. For example, you can wait for the previous copy command to run successfully before running another copy command.

IAM Authentication

Optional. You can configure Amazon Identity and Access Management (IAM) authentication when the PowerCenter Integration Service runs on an Amazon Elastic Compute Cloud (EC2) system. Use IAM authentication for secure and controlled access to Amazon Redshift resources when you run a session.

Use IAM authentication when you want to run a session 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 12](#).
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 the PowerCenter Integration Service on the EC2 system.

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 (Beijing)
- EU (Ireland)
- EU (Frankfurt)
- South America (Sao Paulo)
- US East (N. Virginia)
- US East (Ohio)
- US West (N. California)
- US West (Oregon)

Registering the Plug-in

After you install or upgrade to PowerCenter, you must register the plug-in with the PowerCenter repository.

A plug-in is an XML file that defines the functionality of PowerExchange for Amazon Redshift. To register the plug-in, the repository must be running in exclusive mode. Use the Administrator tool or the `pmrep RegisterPlugin` command to register the plug-in.

The plug-in file for PowerExchange for Amazon Redshift is `AmazonRSCloudAdapterPlugin.xml`. When you install PowerExchange for Amazon Redshift, the installer copies the `AmazonRSCloudAdapterPlugin.xml` file to the following directory:

```
<Informatica Installation Directory>\server\bin\Plugin
```

Note: If you do not have the correct privileges to register the plug-in, contact the user who manages the PowerCenter Repository Service.

Configure the Java Heap Memory

When the source or target contains a large amount of data, configure the memory for the Java heap size on the node that runs the PowerCenter Integration Service. You must ensure that the minimum physical memory available on the server machine for each session is in the range of 300 to 500 MB.

You must set a maximum heap size value based on the amount of data that you want to process.

1. In the Administrator tool, navigate to the PowerCenter Integration Service for which you want to change the Java heap size.
2. On the **Processes** tab, edit the **General Properties** section.
3. Specify the maximum heap size in Java SDK Maximum Memory limit based on the amount of data you want to process.

The default value of the maximum heap size is 64 MB and the minimum size is 32 MB.

4. Click **OK**.
5. Recycle the PowerCenter Integration Service.

CHAPTER 3

Amazon Redshift Sources and Targets

This chapter includes the following topics:

- [Amazon Redshift Sources and Targets, 14](#)
- [Import Amazon Redshift Objects, 14](#)
- [Amazon Redshift Lookup Transformation, 16](#)

Amazon Redshift Sources and Targets

Create a mapping with an Amazon Redshift source to read data from Amazon Redshift. Create a mapping with any source and an Amazon Redshift target to write data to Amazon Redshift. You cannot use the **Create Target** option from the advanced target session properties on an Amazon Redshift target using the PowerCenter Client.

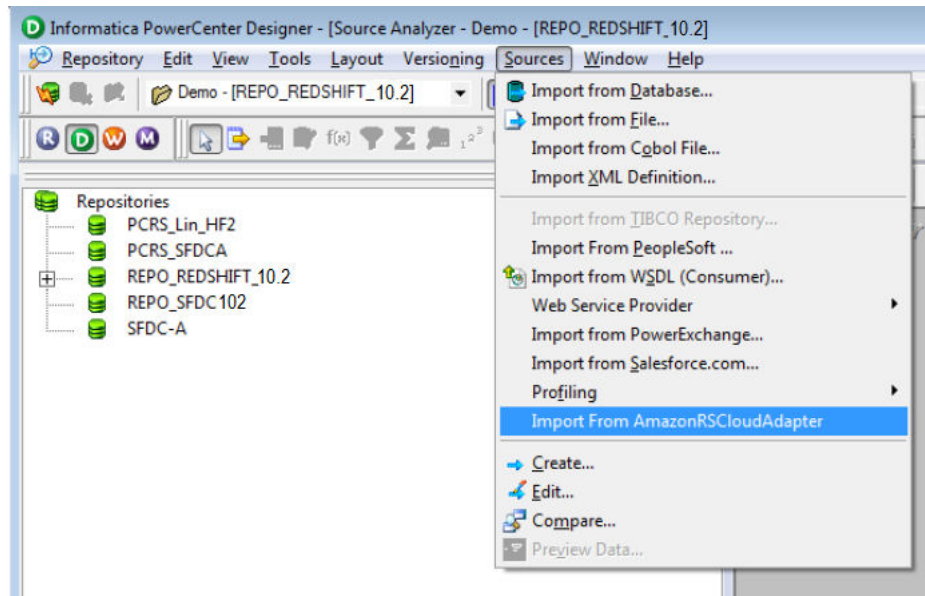
Note: Dynamic partitioning does not work for PowerExchange for Amazon Redshift.

Import Amazon Redshift Objects

You can import Amazon Redshift source and target objects before you create a mapping.

1. Start PowerCenter Designer, and connect to a PowerCenter repository configured with an Amazon Redshift instance.
2. Open a source or target folder.
3. Select **Source Analyzer** or **Target Designer**.

- Click **Sources** or **Targets**, and then click **Import from AmazonRSCloud Adapter**.



Note: When you configure pushdown optimization, click **Import from Database** to import the source or target object.

For information about pushdown optimization, see [“Amazon Redshift Pushdown Optimization Overview” on page 19](#).

The **Establish Connection** dialog box appears.

- Specify the following information and click **Connect**.

Property	Description
Schema	<p>Schema name for the Amazon Redshift tables.</p> <p>When you import objects from AmazonRSCloudAdapter in the PowerCenter Designer, the table names are listed in an alphabetical order.</p> <p>Default is public.</p> <p>Note: The public schema might not work for all the Amazon Redshift tables.</p>
AWS Access Key ID	Amazon S3 bucket access key ID.
AWS Secret Access Key	Amazon S3 bucket secret access key ID.
Master Symmetric Key	<p>Optional. Amazon S3 encryption key.</p> <p>Provide a 256-bit AES encryption key in the Base64 format.</p>

Property	Description
Customer Master Key ID	<p>Optional. Specify the customer master key ID or alias name generated by AWS Key Management Service (AWS KMS).</p> <p>You must generate the customer master key ID for the same region where Amazon S3 bucket reside. You can specify any of the following values:</p> <p>Customer Generated Customer Master Key</p> <p>Enables client-side or server-side encryption.</p> <p>Default Customer Master Key</p> <p>Enables client-side or server-side encryption. Only the administrator user of the account can use the default customer master key ID to enable client-side encryption.</p>
JDBC URL	Amazon Redshift connection URL.
Number of bytes needed to support multibytes for varchar	<p>Not applicable.</p> <p>This property is not supported as you cannot use the Create Target option from the target session properties on an Amazon Redshift target using the PowerCenter Client.</p>

- Click **Next**.
- Select the table that you want to import, and then click **Finish**. If you want to see the table metadata, select the table, and click the table name.

When you import objects from Amazon Redshift, you cannot import related objects as a source.

Amazon Redshift Lookup Transformation

You can use the imported Amazon Redshift source in a lookup transformation.

You cannot configure the **Lookup Caching Enabled** option. By default, the **Lookup Caching Enabled** option is selected.

For more information, see "Lookup Transformation" in the *PowerCenter Transformation Guide*.

CHAPTER 4

Amazon Redshift Mappings

This chapter includes the following topics:

- [Amazon Redshift Mappings Overview, 17](#)
- [Configuring the Source Qualifier, 17](#)
- [Amazon Redshift Mapping Example, 18](#)

Amazon Redshift Mappings Overview

After you import an Amazon Redshift source or target definition into the PowerCenter repository, you can create a mapping to read data from an Amazon Redshift source or write data to an Amazon Redshift target.

You can read data from a single Amazon Redshift source and write data to a multiple Amazon Redshift targets.

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.

For information on the Performance Tuning and Sizing Guidelines, see <https://docs.informatica.com/data-integration/powerexchange-adapters-for-powercenter/h2l/1110-performance-tuning-and-sizing-guidelines-for-powerexchange-/abstract.html>

Configuring the Source Qualifier

When you import a source to create a mapping for Amazon Redshift source , you must configure the source qualifier to create the mapping.

1. In the mapping, click **Source Qualifier**.
2. Select the **Configure** tab.
3. Specify the Amazon Redshift connection details.
4. Save the mapping.

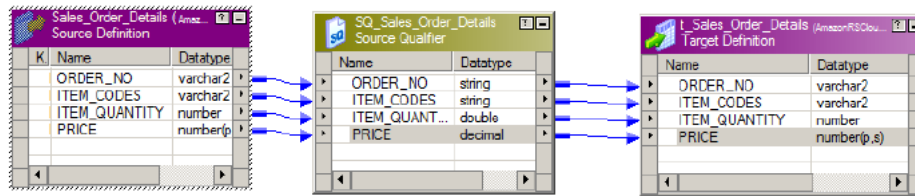
Amazon Redshift Mapping Example

Your e-commerce organization stores sales order details in an Oracle database. Your organization needs to move the data from the Oracle database to an Amazon Redshift target.

The following procedure shows how to move data from the Oracle database to Amazon Redshift:

1. Import the Oracle source.
2. Import an Amazon Redshift target.
3. Create a mapping with a source and an Amazon Redshift target.

The following image shows the example mapping:



4. Create a session and configure it to write the data to the Amazon Redshift target.

The mapping contains the following objects:

Source Definition

The mapping source definition is a relational Oracle database. In the Source Analyzer, import the Oracle source. The PowerCenter Integration Service reads the sales order details from the Oracle source.

The following table describes the structure of the source definition called Sales_Order_Details:

Field	Data Type
Order_No	Varchar
Item_Codes	Varchar
Item_Quantity	Number
Price	Number (p,s)

Mapping Target

The mapping contains an Amazon Redshift target definition.

In the Target Designer, import an Amazon Redshift target definition.

The following image shows the Amazon Redshift target definition t_Sales_Order_Details:

The screenshot shows the 't_Sales_Order_Details' target definition in the Amazon Redshift target designer. The table has columns: Name, Datatype, and Length/Precision. The data is as follows:

Name	Datatype	Length/Precision
ORDER_NO	varchar2	18
ITEM_CODES	varchar2	10
ITEM_QUANTITY	number	15
PRICE	number(p,s)	15

CHAPTER 5

Amazon Redshift Pushdown Optimization

This chapter includes the following topics:

- [Amazon Redshift Pushdown Optimization Overview, 19](#)
- [Configuring Amazon Redshift ODBC Connection, 19](#)
- [Pushdown Optimization Functions, 26](#)
- [Rules and Guidelines in Pushdown Optimization, 28](#)

Amazon Redshift Pushdown Optimization Overview

You can use pushdown optimization to push transformation logic to source or target databases. Use pushdown optimization when you use database resources to improve mapping performance.

When you run a mapping configured for pushdown optimization, the mapping converts the transformation logic to an SQL query. The mapping sends the query to the database, and the database executes the query.

Amazon Redshift supports source-side, target-side, or full pushdown optimization for mappings. You can perform insert, update, or delete operation in a pushdown optimization.

Note: PowerExchange for Amazon Redshift does not support upsert operation in a pushdown optimization.

When you configure full pushdown optimization for a mapping, you can include all the commands in one transaction and then use a single commit command to increase the performance. Set the value of the `FullPushdownInOneTransaction` custom property to yes to avoid multiple commit commands. For example, instead of running separate transactions for create view, insert, update, delete, and drop view commands, you can include all the commands in a single transaction and commit the transaction.

Configuring Amazon Redshift ODBC Connection

You can set the pushdown optimization for the ODBC connection type that uses Amazon ODBC Redshift drivers to enhance the mapping performance. To use an ODBC connection to connect to Amazon Redshift, you must configure the ODBC connection.

After you configure the ODBC connection, select the value of the **Pushdown Optimization** property as **Full**, **To Source**, or **To Target** accordingly in the session properties.

Note: If you want to perform an update operation, you must select the value of the **Allow Temporary View for Pushdown** as **Yes** in the session properties.

Amazon Redshift supports Amazon ODBC Redshift drivers on Windows and Linux systems. You must install the Amazon ODBC Redshift 32-bit or 64-bit driver based on your system requirement.

Note: Informatica has certified the Amazon Redshift ODBC pushdown optimization using the following driver version: AmazonRedshiftODBC-64-bit-1.4.8.1000-1.x86_64

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 32-bit or 64-bit driver based on your Windows system.
2. Install the Amazon Redshift ODBC drivers on the machine that hosts the PowerCenter Integration Service.
3. Open the following folder in which ODBC data source file is installed.

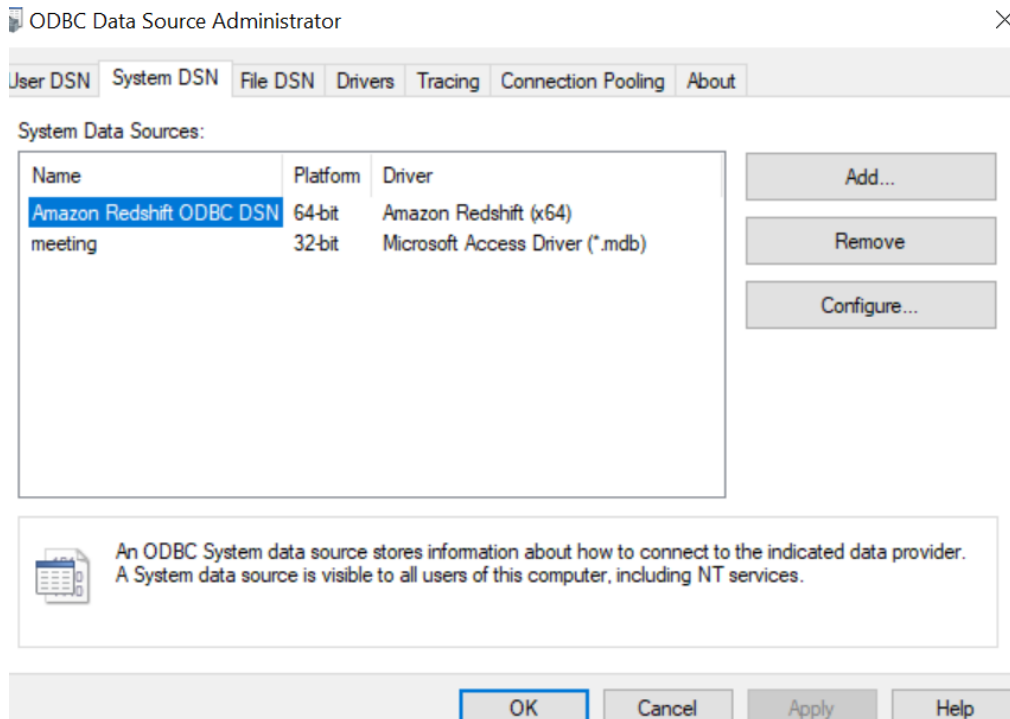
- For 32-bit driver: C:\WINDOWS\syswow64
- For 64-bit driver: C:\WINDOWS\system32

4. Run the `odbcad32.exe` file.

The **ODBC Data Sources Administrator** box appears.

5. Click **System DSN**.

The **System DSN** tab appears. The following image shows the **System DSN** tab on the **ODBC Data Sources Administrator** box:



- Click **Configure**.

The **Amazon Redshift ODBC Driver DSN Setup** box appears. The following image shows the **Amazon Redshift ODBC Driver DSN Setup** box where you can configure the connection settings and authentication:

Amazon Redshift ODBC Driver DSN Setup

Connection Settings

Data Source Name: Amazon_Redshift_ODBC_DSN

Server: i46sw.us-west-2.redshift.amazonaws.com

Port: 5439

Database: rsqa

Authentication

Auth Type: Standard

User: User_Name

Password:

Encrypt Password For:

☒ Current User Only ☐ All Users of This Machine

SSL Options... Additional Options...

Logging Options... Data Type Options...

v1.3.7.1000 (64 bit) Test OK Cancel

- 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.

Property	Description
Port	Port number of the Amazon Redshift server.
Database	Name of the Amazon Redshift database.

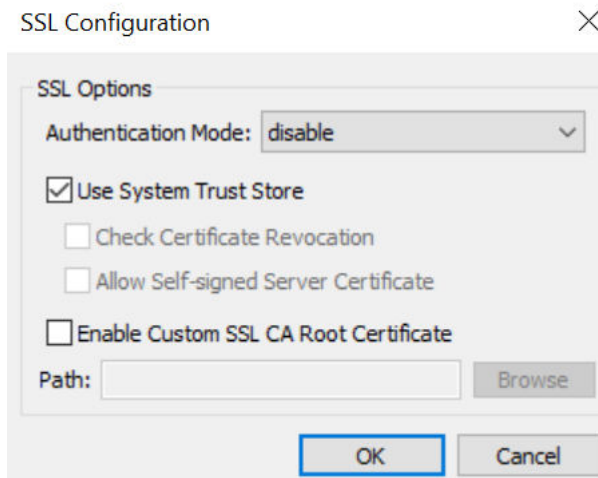
Note: You must specify the **Server**, **Port**, and **Database** values from the JDBC URL.

8. Specify the following authentication properties in the **Authentication** section:

Property	Description
Auth Type	Type of the authentication. Default is Standard .
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: <ul style="list-style-type: none"> - Current User Only - All Users of This Machine Default is Current User Only .

9. Click **SSL Options** in the **Amazon Redshift ODBC Driver DSN Setup** box.

The **SSL Configuration** box appears. The following image shows the **SSL Configuration** box:



10. Select **disable** to disable the authentication in the **Authentication Mode** field.
11. Click **OK** in the **SSL Configuration** box.
The **SSL Configuration** box closes.
12. Click **Test** to test the connection in the **Amazon Redshift ODBC Driver DSN Setup** box.
13. 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 Amazon Redshift ODBC Connection” on page 24](#)

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 32-bit or 64-bit driver based on your Linux system.
2. Install the Amazon Redshift ODBC drivers on the machine that hosts the PowerCenter Integration Service.
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.
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. Add the `odbc.ini` file path in your source file in the following format:

```
ODBCINI=/<odbc.ini file path>/odbc.ini
```

6. Restart the PowerCenter Server.

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 Amazon Redshift ODBC Connection” on page 24](#)

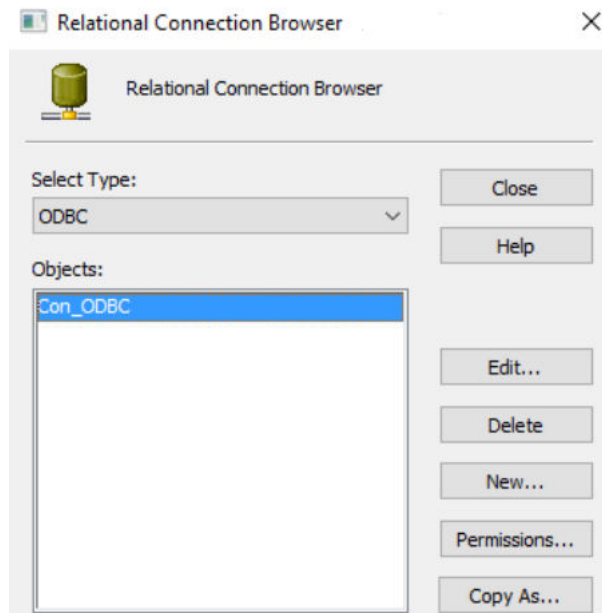
Creating an Amazon Redshift ODBC Connection

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:

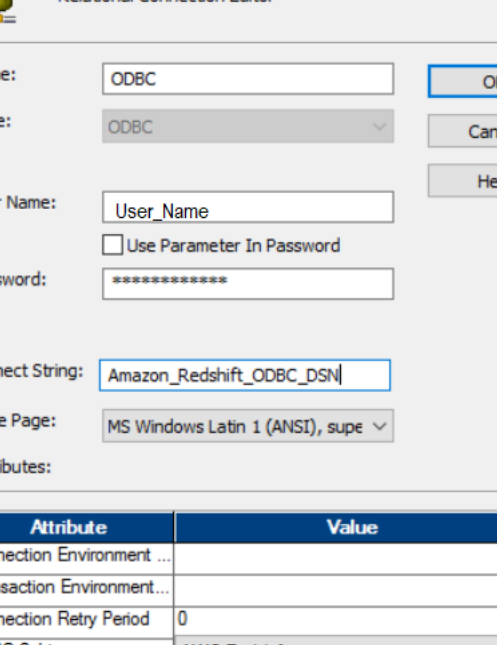
1. In the Workflow Manager, click **Connections**.
2. Select **Relational** from the list.

The **Relational Connection Browser** box appears. The following image shows the **Relational Connection Browser** box:



3. Select **Type** as **ODBC**.
4. Click **New**.

The **Connection Object Definition** box appears. The following image shows the **Connection Object Definition** box:



Connection Object Definition

Relational Connection Editor

Name: ODBC

Type: ODBC

User Name: User_Name

☐ Use Parameter In Password

Password: *****

Connect String: Amazon_Redshift_ODBC_DSN

Code Page: MS Windows Latin 1 (ANSI), supe

Attributes:

Attribute	Value
Connection Environment ...	
Transaction Environment...	
Connection Retry Period	0
ODBC Subtype	AWS Redshift

5. Configure the following relational connection properties:

Relational Connection Property	Description
Name	Enter a name for the connection.
Type	The connection type is set by default. You cannot edit this value.
User Name	Enter the user name to connect to the Amazon Redshift database.
Password	Enter the password to connect to the Amazon Redshift database.
Connect String	Enter the name of the ODBC data source that you created for the Amazon Redshift database.
Code Page	Select the code page that the PowerCenter Integration Service must use to read or write data.
Attributes	Enter the ODBC Subtype attribute value as AWS Redshift .

The Amazon Redshift ODBC connection is created successfully.

Pushdown Optimization Functions

The following table summarizes the availability of pushdown 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, target-side, or full pushdown optimization. Columns marked with an S indicate that the function can be pushed to the Amazon Redshift database by using source-side pushdown optimization. Columns marked with a dash (-) symbol indicate that the function cannot be pushed to the database.

Function	Pushdown	Function	Pushdown	Function	Pushdown
ABORT()	-	INITCAP()	X	REG_MATCH()	-
ABS()	X	INSTR()	X	REG_REPLACE	-
ADD_TO_DATE()	X	IS_DATE()	-	REPLACECHR()	-
AES_DECRYPT()	-	IS_NUMBER()	-	REPLACESTR()	-
AES_ENCRYPT()	-	IS_SPACES()	-	REVERSE()	-
ASCII()	-	ISNULL()	S	ROUND(DATE)	-
AVG()	S	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()	-	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()	-	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

Function	Pushdown	Function	Pushdown	Function	Pushdown
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)	S
GREATEST()	-	PV()	-	TRUNC(NUMBER)	S
IIF()	X	RAND()	-	UPPER()	X
IN()	S	RATE()	-	VARIANCE()	S
INDEXOF()	-	REG_EXTRACT()	-		

The following table lists the pushdown 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, target-side, or full pushdown optimization. Columns marked with an S indicate that the operator can be pushed to the Amazon Redshift database by using source-side pushdown optimization.

Operator	Pushdown
+	X
-	X
*	S
/	X
%	X
	X
>	S
=	S
>=	S
<=	S
!=	S
AND	S
OR	S

Operator	Pushdown
NOT	S
$\wedge =$	S

Rules and Guidelines in Pushdown 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 PowerCenter Integration Service 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.
- For Amazon Redshift, when you define only a string argument for TO_DATE() and TO_CHAR(), the PowerCenter Integration Service considers the default date format present in the session property. The default date format in the session property is: MM/DD/YYYY HH24:MI:SS.US
- 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 PowerCenter Integration Service 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 PowerCenter Integration Service 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
 - HH24
 - MI
 - MM
 - MS
 - SS
 - US

- YYYY
- To push GET_DATE_PART() to Amazon Redshift, you must use the following formats:
 - D
 - DDD
 - HH24
 - MI
 - MM
 - MS
 - SS
 - US
 - YYYY
- You can perform an insert operation when you use tables from an external schema in a mapping enabled with full pushdown optimization.
- PowerCenter with Pushdown Optimization enabled generates an SQL query to create a temporary view. To enable the SQL query to run successfully, include the `WITH NO SCHEMA BINDING` clause in the SQL query. You can set a custom property to include the `WITH NO SCHEMA BINDING` clause in the SQL query. Set the `REDSHIFT_NO_SCHEMA_BINDING` custom property to `Yes` in the **Config Object** tab of the required workflow.

CHAPTER 6

Amazon Redshift Sessions

This chapter includes the following topics:

- [Amazon Redshift Sessions Overview, 30](#)
- [Amazon Redshift Connections, 31](#)
- [Amazon Redshift Source Sessions, 32](#)
- [Amazon Redshift Target Sessions, 40](#)
- [Working with Large Tables, 47](#)
- [Octal values as DELIMITER and QUOTE, 48](#)
- [Success and Error Files, 48](#)

Amazon Redshift Sessions Overview

You must configure an Amazon Redshift connection in the Workflow Manager to read data from or write data to an Amazon Redshift table.

The PowerCenter Integration Service writes the data to a staging directory and then to an Amazon S3 bucket before it writes the data to Amazon Redshift. You must specify the location of the staging directory in the session properties. You must also specify an Amazon S3 bucket name in the session properties. You must have write access to the Amazon S3 bucket.

Note: Informatica recommends to set the DataMovementMode property to Unicode for the PowerCenter Integration Service when running sessions that connect to Amazon Redshift. This will prevent the intermittent data shift or Unicode data corruption issues that are likely to occur during the session run.

Amazon Redshift Connections

Use an Amazon Redshift connection to connect to the Amazon Redshift database. The PowerCenter Integration Service uses the connection when you run an Amazon Redshift session.

PowerExchange for Amazon Redshift Connections

When you configure an Amazon Redshift connection, you define the connection attributes that the PowerCenter Integration Service uses to connect to the Amazon Redshift.

The following table describes the application connection properties:

Property	Description
Name	Name of the Amazon Redshift connection.
Type	The AmazonRSCloudAdapter connection type.
User Name	User name to access the Amazon Redshift database.
Password	Password for the Amazon Redshift database user name.

The following table describes the Amazon Redshift connection attributes:

Property	Description
Schema	Schema name for the Amazon Redshift tables. When you import objects from AmazonRSCloudAdapter in the PowerCenter Designer, the table names are listed in an alphabetical order. Default is public. Note: The public schema might not work for all the Amazon Redshift tables.
AWS Access Key ID	Amazon S3 bucket access key ID.
AWS Secret Access Key	Amazon S3 bucket secret access key ID.
Master Symmetric Key	Optional. 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 specify any of the following values: Customer Generated Customer Master Key Enables client-side or server-side encryption. Default Customer Master Key Enables client-side or server-side encryption. Only the administrator user of the account can use the default customer master key ID to enable client-side encryption.

Property	Description
JDBC URL	Amazon Redshift connection URL.
Number of bytes needed to support multibytes for varchar	Not applicable. This property is not supported as you cannot use the Create Target option from the target session properties on an Amazon Redshift target using the PowerCenter Client.

Configuring an Amazon Redshift Connection

Configure an Amazon Redshift connection in the Workflow Manager to define the connection attributes that the PowerCenter Integration Services uses to connect to the Amazon Redshift database.

1. In the Workflow Manager, click **Connections > Application**.
The **Application Connection Browser** dialog box appears.
2. Click **New**.
The **Select Subtype** dialog box appears.
3. Select **AmazonRSCloudAdapter** and click **OK**.
The **Connection Object Definition** dialog box appears.
4. Enter a name for the Amazon Redshift connection.
5. Enter the application properties for the connection.
6. Enter the Amazon Redshift connection attributes.
7. Click **OK**.

Configuring the Source Qualifier

When you import a source to create a mapping for Amazon Redshift source , you must configure the source qualifier to create the mapping.

1. In the mapping, click **Source Qualifier**.
2. Select the **Configure** tab.
3. Specify the Amazon Redshift connection details.
4. Save the mapping.

Amazon Redshift Source Sessions

Create a mapping with an Amazon Redshift source and a target to read data from Amazon Redshift.

You can encrypt data, specify the location of the staging directory, and securely unload the results of a query to files on Amazon Redshift.

Amazon Redshift Staging Directory for Amazon Redshift Sources

The PowerCenter Integration Service creates a staging file in the directory that you specify in the source session properties.

The PowerCenter Integration Service deletes the staged files from the staging directory before reading data from Amazon Redshift. Specify a staging directory in the session properties with an appropriate amount of disk space for the volume of data that you want to process.

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 on the machine that hosts the PowerCenter Integration Service.

The PowerCenter Integration Service creates subdirectories in the staging directory. The subdirectories use the following naming convention: <staging directory>/infaRedShiftStaging<MMDDHHmmSS>

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 target session 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.

Encryption Type

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 source session properties. 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: PowerExchange for Amazon Redshift for PowerCenter does not support the server-side encryption with the master symmetric key and client-side encryption with the customer master key.

Client-side Encryption

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

To enable client-side encryption, you must provide a master symmetric key, which is a 256-bit AES encryption key in Base64 format in the connection properties. The PowerCenter Integration Service encrypts the data by using the master symmetric key or customer master key.

The PowerCenter Integration Service encrypts the files that are uploaded to Amazon S3 at the client-side. 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. If you provide customer master key ID generated by AWS Key Management Service in the Amazon Redshift connection properties, then the server-side encryption is used for data load. To enable client-side encryption, perform the following tasks:

1. Provide the master symmetric key when you create an Amazon Redshift connection. Ensure that you provide a 256-bit AES encryption key in Base64 format.
2. Update the security policy JAR files to support encryption with maximum security. Update the `local_policy.jar` and the `US_export_policy.jar` files in the following directory: `<Informatica Installation Directory>\java\jre\lib\security`. You can download the JAR files supported by your JAVA environment from the Oracle website.
3. Select **Turn on S3 Client Side Encryption** as the encryption type in the source session properties.

Identity Columns

An identity column contains unique values that are automatically generated.

Rules and Guidelines for Identity Columns

- The data type for an identity column must be either int or bigint.
- When you create a mapping for an insert operation, you must link either all the source and target identity columns or none.
- When you create a mapping for an update, upsert or delete operation, you cannot map the identity columns that are not part of the primary key.
- If an identity column is part of the primary key, you must map the column for update, upsert, and delete operations, or the session fails. However, you cannot set a source value for these columns.
- The `ExplicitID` and `MaxError` count options are removed for the upsert, update, and delete operations.

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 space. The Unload command has the following options and default values:

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

You can also create a property file. 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
```

In the property file, delimit the options by using a new line. For example:

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

It is recommended to use octal representation of non-printable characters as DELIMITER 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 session on PowerCenter Integration Service 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

If you need to extract a large amount of source data, you can partition the sources to improve session performance. Partitioning sources allows the PowerCenter Integration Service to create multiple connections to sources and process partitions of source data concurrently. You can partition sources if the PowerCenter Integration Service can maintain data consistency when it processes the partitioned data.

By default, the Workflow Manager sets the partition type to pass-through for Amazon Redshift tables. In pass-through partitioning, the PowerCenter Integration Service passes all rows at one partition point to the next partition point without redistributing them.

If you create multiple partitions for an Amazon Redshift source session, the PowerCenter Integration Service evaluates the session properties in the following order to run the session:

1. SQL Query
2. INFA ADVANCED FILTER
3. Slices on Amazon Redshift Nodes

Configuring Partitioning

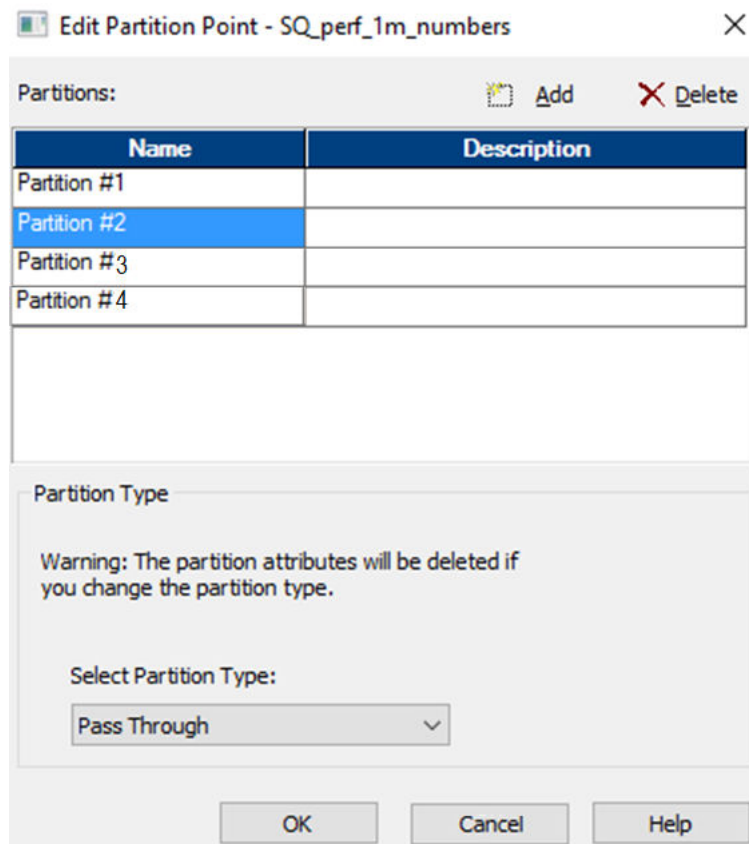
When you create or edit a session, you can add partition to the sources to improve the session performance.

Perform the following steps to configure partitioning:

1. In the Workflow Manager, open the session.
2. Double-click the session.
The **Session Properties** dialog box appears.
3. Click **Mapping**.
The **Mapping** tab appears.
4. Click **Partitions** and select the partition point where you want to add the partition.
5. Click **Edit Partition Point**.
The **Edit Partition Point** dialog box appears.
6. Click **Add** to add a partition.

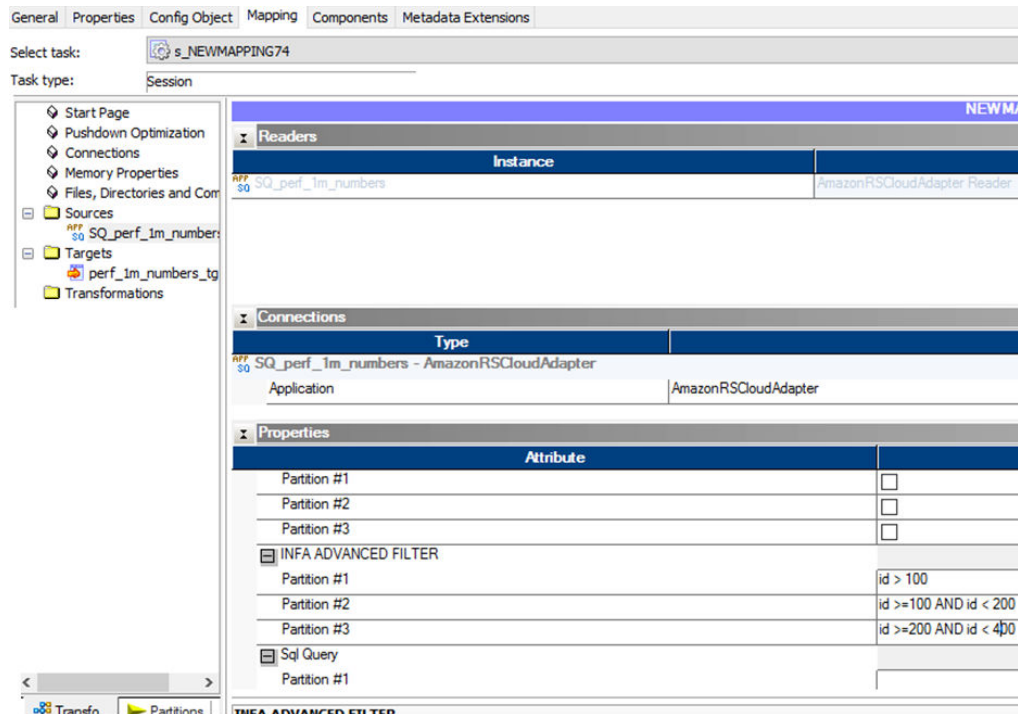
Note: You can add multiple partitions as required.

The following image shows the **Edit Partition Point** dialog box where you can add partitions:



7. Under the **Partition Type** section, select the **Pass Through** partition.
8. Click **OK**.
The **Edit Partition Point** dialog box closes.
9. In the **Properties** section on the **Mapping** tab, navigate to the **INFA ADVANCED FILTER** attribute.
10. Add the filter conditions for each partition to optimize the search.
For example, `id>0 AND id<1000`.

The following image shows the **INFA ADVANCED FILTER** attribute where you can add the filter conditions:



11. Click **OK**.

Amazon Redshift Source Session Configuration

You can configure a session to read data from Amazon Redshift. Define the properties for each source in the session.

The following table describes the source session properties:

Property	Description
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.
Number of Sorted Ports	Number of columns used when sorting rows queried from the source. The PowerCenter Integration Service 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 PowerCenter Integration Service includes a SELECT DISTINCT statement if you choose this option. Amazon Redshift ignores trailing spaces. Therefore, the PowerCenter Integration Service might extract fewer rows than expected.

Property	Description
Source Table Name	You can override the default source table name.
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 staged files before writing the files to Amazon Redshift. Session performance improves when the PowerCenter Integration Service compresses the staged files. Default is selected.
Staging Directory Location	Amazon Redshift staging directory. Specify a directory on the machine that hosts the PowerCenter Integration Service.
UnloadOptions Property File	Path to the property file. Enables you to add options to the unload command to write data from an Amazon Redshift object to an S3 bucket. You can add the following options: <ul style="list-style-type: none"> - DELIMITER - PARALLEL - ESCAPE - AWS_IAM_ROLE 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. Specify a directory on the machine that hosts the PowerCenter Integration Service.
Turn on S3 Client Side Encryption	Indicates that the PowerCenter Integration Service encrypts data before writing the data to Amazon S3 by using a private encryption key.
Encryption Type	Select the source encryption type. You can select from the following encryption types: <ul style="list-style-type: none"> - SSE-S3 - SSE-KMS Default is SSE-S3.
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 8 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.
Infra Advanced Filter	SQL filter command to divide the source database into multiple segments.
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.

Note:

Note: If you add constraints to the table structure in a mapping while the specified constraints are not present in the Amazon Redshift source table, the session fails with an error message.

Amazon Redshift Target Sessions

Create a session and associate it with the mapping that you created to move data to an Amazon Redshift table. Change the connection to an Amazon Redshift connection, and define the session properties to write data to Amazon Redshift.

You can perform insert, update, delete, and upsert operations on an Amazon Redshift target.

If the distribution key column in a target table contains null values and you configure a task with an **Update as Insert** 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 PowerCenter Integration Service creates a staging file in the directory that you specify in the target properties.

The PowerCenter Integration Service deletes the staged files from the staging directory before writing data to Amazon Redshift. Specify a staging directory in the target session properties with an appropriate amount of disk space for the volume of data that you want to process.

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 on the machine that hosts the PowerCenter Integration Service.

The PowerCenter Integration Service creates subdirectories in the staging directory. The subdirectories use the following naming convention: `<staging directory>/infaRedShiftStaging<MMDDHHmmSS>`

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 **Turn on S3 Server Side Encryption** as the encryption type in the target session 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 target session 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

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

To enable client-side encryption, you must provide a master symmetric key, which is a 256-bit AES encryption key in Base64 format in the connection properties. The PowerCenter Integration Service encrypts the data by using the master symmetric key or customer master key.

The PowerCenter Integration Service encrypts the files that are uploaded to Amazon S3 at the client-side. 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. If you provide customer master key ID generated by AWS Key Management Service in the Amazon Redshift connection properties, then the server-side encryption is used for data load. To enable client-side encryption, perform the following tasks:

1. Provide the master symmetric key when you create an Amazon Redshift connection. Ensure that you provide a 256-bit AES encryption key in Base64 format.
2. Update the security policy JAR files to support encryption with maximum security.
Update the `local_policy.jar` and the `US_export_policy.jar` files in the following directory: `<Informatica Installation Directory>\java\jre\lib\security`. You can download the JAR files supported by your JAVA environment from the Oracle website.
3. Select **Turn on S3 Client Side Encryption** as the encryption type in the target session properties.

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.

You can configure vacuum table recovery options in the session properties. You can choose to recover disk space for the entire database or for individual tables in a database. Vacuum databases or tables often to maintain consistent query performance. 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 a session that writes to Amazon Redshift. 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.

Retain staging files

You can retain staging files on Amazon S3 after the PowerCenter Integration Service 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 PowerCenter Integration Service 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 PowerCenter Integration Service 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 PowerCenter Integration Service 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 PowerCenter Integration Service 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 space. The Copy command has the following options and default values:

```
DELIMITER=\036 ACCEPTINVCHARS=? QUOTE=\037 COMPUPDATE=OFF AWS_IAM_ROLE=arn:aws:iam::<account ID>:role/<role-name>
```

You can also create a property file. 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
```

In the property file, delimit the options by using a new line. For example:

```
DELIMITER=\036
ACCEPTINVCHARS=?
QUOTE=\037
COMPUPDATE=OFF
AWS_IAM_ROLE=arn:aws:iam::<account ID>:role/<role-name>
```

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

Copy command options

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

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

DELIMITER

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

ACCEPTINVCHARS

Loads data into VARCHAR columns even if the data contains UTF-8 characters that are not valid. When you specify ACCEPTINVCHARS, the PowerCenter Integration Service 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 PowerCenter Integration Service 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 session on PowerCenter Integration Service 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.

Amazon Redshift Target Session Configuration

You can configure a session to write data to Amazon Redshift. Define the properties for each target instance in the session.

The following table describes the target session 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 staged files before writing the files to Amazon Redshift. Session performance improves when the PowerCenter Integration Service compresses the staged files. Default is selected.
Staging Directory Location	Amazon Redshift staging directory. Specify a directory on the machine that hosts the PowerCenter Integration Service.
Batch Size	Minimum number of rows in a batch. Enter a number greater than 0. Default is 1000. Note: Configure the commit interval in the Session > Properties when you configure the batch size. Specify the same value for commit interval that you specified for the batch size.
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 PowerCenter Integration Service writes the entire batch to the error file. Default is 1.
Truncate Target Table Before Data Load	Truncates an Amazon Redshift target table 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 PowerCenter Integration Service to make the staged files consistent with the list of files available on Amazon S3. Default is 5.

Property	Description
CopyOptions Property File	<p>Name of the property file.</p> <p>Enables you to add additional options to the copy command for writing data from an Amazon S3 source to an Amazon Redshift target when the default delimiter comma (,) or double-quote (") is used in the data.</p> <p>You can add the following options:</p> <ul style="list-style-type: none"> - DELIMITER - ACCEPTINVCHARS - QUOTE - COMPUPDATE - AWS_IAM_ROLE <p>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.</p> <p>Specify a directory on the machine that hosts the PowerCenter Integration Service.</p>
Turn on S3 Server Side Encryption	<p>Indicates that Amazon S3 encrypts data during upload and decrypts data at the time of access.</p>
Turn on S3 Client Side Encryption	<p>Indicates that the PowerCenter Integration Service encrypts data by using a private encryption key.</p> <p>If you enable both server side and client side encryption, the PowerCenter Integration Service ignores the server side encryption.</p>
Vacuum Target Table	<p>Recovers disk space and sorts rows in a specified table or all tables in the database.</p> <p>You can select the following recovery options:</p> <p>None</p> <p>Does not sort rows or recover disk space.</p> <p>Full</p> <p>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.</p> <p>Sort Only</p> <p>Sorts the specified table or all tables in the database without recovering space freed by deleted rows.</p> <p>Delete Only</p> <p>Recovers disk space occupied by rows marked for deletion by previous update and delete operations, and compresses the table to free up used space.</p> <p>Reindex</p> <p>Analyzes the distribution of the values in the interleaved sort key columns to configure the entire Vacuum table operations for a better performance.</p> <p>Default is None.</p>
Analyze Target Table	<p>Improve the efficiency of the write operations.</p> <p>The query planner on Amazon Redshift updates the statistical metadata to build and choose optimal plans to improve the efficiency of queries.</p>
Prefix for Retaining Staging Files on S3	<p>Retains staging files on Amazon S3.</p> <p>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, <code>backup_dir/backup_file</code> or <code>backup_file</code>.</p>

Property	Description
INSERT	If enabled, the PowerCenter Integration Service inserts all rows flagged for insert. If disabled, the PowerCenter Integration Service rejects the rows flagged for insert. By default, the insert operation is enabled.
DELETE	If enabled, the PowerCenter Integration Service deletes all rows flagged for delete. If disabled, the PowerCenter Integration Service rejects all rows flagged for delete. To perform a delete operation, you must map the primary key column and at least one column other than primary key column.
UPDATE	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 session target attributes: Update as Update The PowerCenter Integration Service updates all rows as updates. Update as Insert The PowerCenter Integration Service updates existing rows and inserts other rows as if marked for insert. Note: In the session properties, if you set Update as the value of the Treat source as rows property and select the Update as Insert option in the target session properties, the mapping runs successfully. However, the PowerCenter Integration Service rejects the data.
Success File Directory	Directory for the Amazon Redshift success file. Specify a directory on the machine that hosts the PowerCenter Integration Service.
Error File Directory	Directory for the Amazon Redshift error file. Specify a directory on the machine that hosts the PowerCenter Integration Service.
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	Calculates the number of the staging files per batch. If you do not provide the number of files, PowerExchange for Amazon Redshift calculates the number of the staging files.

Configuring HTTP Proxy Options

If your organization uses a proxy server to access the internet, you can configure the HTTP proxy server authentication settings at run time to read data from or write data to Amazon Redshift.

Note: PowerExchange for Amazon Redshift does not support proxy connection in a PowerCenter Designer.

Configuring HTTP Proxy Options at Run-Time

If your organization uses a proxy server to access the internet, you can configure the proxy server authentication settings in the Administrator Console to read data from or write data to Amazon Redshift.

Perform the following steps to configure the HTTP proxy options at run-time:

- Ensure that you enable the proxy server settings from your web browser.
- In the Administrator Console, navigate to the PowerCenter Integration Service for which you want to set the proxy server settings.
- Add the following custom properties as JVMOptions:

Property	Description
-Dhttps.proxyHost=	Name of the HTTP proxy server.
-Dhttps.proxyPort=	Port number of the HTTP proxy server.
-Dhttps.proxyUser=	Authenticated user name for the HTTP proxy server. This is required if the proxy server requires authentication.
-Dhttps.proxyPassword=	Password for the authenticated user. This is required if the proxy server requires authentication. Note: The password is in plain text and not encrypted.

For example:

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

```
JVMOption2=-Dhttps.proxyPort=8081
```

```
JVMOption3=-Dhttps.proxyUser=adminuser
```

```
JVMOption4=-Dhttps.proxyPassword=password
```

Working with Large Tables

You can upload or download a large object as a set of multiple independent parts.

PowerExchange for Amazon Redshift uses the AWS TransferManager API to upload a large object in multiple parts to Amazon S3. While downloading a large object, the PowerCenter Integration Service 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 the **Part Size** and **TransferManager Thread Pool Size** options in the target session 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 PowerCenter Integration Service generates success and error files after you run a session. Success and error files are .csv files that contain row-level details.

The PowerCenter Integration Service does not overwrite success or error files. You can manually delete the files that you no longer need.

Success Files

The PowerCenter Integration Service 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 PowerCenter Integration Service writes to the Amazon S3 bucket and then to the Amazon Redshift target.

If you want the PowerCenter Integration Service to generate a success file, specify a directory for the success file in the session properties. Specify a directory on the machine that hosts the PowerCenter Integration Service.

Rules and Guidelines for Success Files

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

- If you do not specify a directory for the success file as a session property, the PowerCenter Integration Service writes an empty file to the following directory: `$PMTargetFileDir`.
- The PowerCenter Integration Service generates the success file with the following naming convention:
`<sessionName>_<timestamp>_success.csv`.
- If you do not map a primary key column in a mapping, the insert, update, upsert, and delete operations fail.

Error Files

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 PowerCenter Integration Service does not write data to the Amazon Redshift target.

If you want the PowerCenter Integration Service to generate an error file, specify a directory for the error file in the session properties. You must not specify a network path or shared path.

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

- If you do not specify a directory for the error file as a session property, the PowerCenter Integration Service writes a blank file to the following directory: `$PMBadFileDir`.
- The PowerCenter Integration Service generates errors file with the following naming convention:
`<sessionName>_<timestamp>_error.csv`.

Sample Error File

If a target table has the fields `f_integer`, `f_char`, and `f_varchar`, and if a row is rejected, the PowerCenter Integration Service generates an error file in the following format:

Errors Details	f_integer	f_char	f_varchar
"Query Start Time: 2014-03-24 11:41:30.629 Offending File: INSERT_bdt_with_composite_key.batch_0.csv.0.gz Line Number: 4 Column Name: f_char Column Type: char Offending Value:Furniture Values Intl LLC_upd_upd ERROR Reason: Multibyte character not supported for CHAR (Hint: try using VARCHAR). Invalid char: c3 a6"	"3"	""æ^'Furniture Values Intl LLC_upd_upd""	""001E000000SI3jIAT""
"Query Start Time: 2014-03-24 11:42:00.763 Offending File: INSERT_bdt_with_composite_key.batch_8.csv.0.gz Line Number: 80 Column Name: f_char Column Type: char Offending Value:Heitkamp Inc_upd_upd ERROR Reason: Multibyte character not supported for CHAR (Hint: try using VARCHAR). Invalid char: c3 a6"	"9999"	""æ^'Heitkamp Inc_upd_upd""	""001E000000SHd7ZIAT""

APPENDIX A

Amazon Redshift Data Type Reference

This appendix includes the following topics:

- [Data Type Reference Overview, 50](#)
- [Amazon Redshift and Transformation Data Types, 50](#)

Data Type Reference Overview

When you run the session to read data from or write data to Amazon Redshift, the PowerCenter Integration Service converts the transformation data types to comparable native Amazon Redshift data types.

Amazon Redshift and Transformation Data Types

The Amazon Redshift data types are the names and the aliases represent how the PowerCenter Integration Service stores the data types.

For example, SMALLINT is the Amazon Redshift data type name. The data type is stored as a 2-byte integer. Here, SMALLINT is the Amazon Redshift data type name and INT2 is the Amazon Redshift data type alias.

The following table compares the Amazon Redshift data types and the transformation data types:

Amazon Redshift Data Type	Amazon Redshift Data Type Aliases	Description	Transformation Data Type
SMALLINT	INT2	Signed two-byte integer.	Small Integer
INTEGER	INT, INT4	Signed four-byte integer.	Integer
BIGINT	INT8	Signed eight-byte integer.	Bigint
DECIMAL	NUMERIC	Exact numeric of selectable precision.	Decimal

Amazon Redshift Data Type	Amazon Redshift Data Type Aliases	Description	Transformation Data Type
REAL	FLOAT4	Single precision floating-point number.	Double
DOUBLE PRECISION	FLOAT8, FLOAT	Double precision floating-point number.	Double
BOOLEAN	BOOL	Logical Boolean (true/false).	Small Integer
CHAR	CHARACTER, NCHAR, BPCHAR	Fixed-length character string.	String
VARCHAR	CHARACTER VARYING, NVARCHAR, TEXT	Variable-length character string with a user-defined limit.	String
DATE	NA	Calendar date (year, month, day).	Timestamp
TIMESTAMP	TIMESTAMP WITHOUT TIME ZONE	Date and time (without time zone).	Timestamp

APPENDIX B

Troubleshooting

This appendix includes the following topics:

- [Troubleshooting Overview, 52](#)
- [Troubleshooting for PowerExchange for Amazon Redshift, 52](#)

Troubleshooting Overview

Use the following sections to troubleshoot errors in PowerExchange for Amazon Redshift.

Troubleshooting for PowerExchange for Amazon Redshift

How to configure performance tuning and sizing guidelines for PowerExchange for Amazon Redshift for PowerCenter?

For more information about performance tuning and sizing guidelines, see <https://docs.informatica.com/data-integration/powerexchange-adapters-for-powercenter/h2l/1110-performance-tuning-and-sizing-guidelines-for-powerexchange-/abstract.html>

How to configure pushdown optimization for PowerExchange for Amazon Redshift task using an ODBC connection?

For information about configuring pushdown optimization using an ODBC connection, see <https://docs.informatica.com/data-integration/powerexchange-adapters-for-powercenter/h2l/1117-configuring-pushdown-optimization-for-powerexchange-for-ama/abstract.html>

How to retain the null values when I create a mapping to read data from a source using PowerExchange for Amazon Redshift for PowerCenter?

For information about retaining the null values when you create a mapping to read data from a source, see <https://kb.informatica.com/howto/6/Pages/20/515191.aspx?myk=515191>

How to optimize the performance when I perform an upsert operation to write large amount of data to an Amazon Redshift target using PowerExchange for Amazon Redshift for PowerCenter?

For information about optimizing the performance when you perform an upsert operation to write large amount of data to an Amazon Redshift target, see <https://kb.informatica.com/howto/6/Pages/20/516302.aspx?myk=516302>

INDEX

A

- administration
 - IAM authentication [11](#)
 - minimal Amazon IAM policy [12](#)
- Amazon Redshift
 - connection [31](#)
 - sessions overview [30](#)
 - source session properties [38](#)
 - target session properties [44](#)
- Amazon Redshift data types
 - comparing with transformation data types [50](#)
 - overview [50](#)
- Amazon Redshift mapping example [18](#)
- Amazon Redshift mappings
 - overview [17](#)
- Amazon Redshift source
 - server-side encryption [33](#)
 - staging directory [33](#)
- Amazon Redshift target
 - sessions [40](#)
 - staging directory [40](#)
- Amazon Redshift targets
 - data encryption [40](#)

C

- Configure
 - Amazon Redshift connection [32](#)
- Configuring Amazon Redshift
 - ODBC connections [19](#)
 - ODBC connections on Windows [20](#)
- Configuring Amazon Redshift ODBC connection
 - Linux [23](#)
- configuring partitioning [36](#)
- Creating an Amazon Redshift ODBC connection [24](#)

E

- encryption type [33](#)
- error file [49](#)

I

- IAM authentication
 - administration [11](#)

J

- Java heap memory [13](#)

L

- lookup transformation [16](#)

O

- octal values
 - DELIMITER [48](#)
 - QUOTE [48](#)

P

- partitioning
 - Amazon Redshift, sessions [36](#)
- plug-ins
 - registering PowerExchange for Redshift [13](#)
- prerequisites [10](#)
- pushdown optimization
 - functions [26](#)
 - overview [19](#)
- pushdown optimization function
 - rules and guidelines [28](#)

R

- rules and guidelines
 - success files
 - error files [48](#)

S

- sessions
 - Amazon Redshift, partitioning [36](#)
- source details [14](#)
- success file [48](#)

T

- troubleshooting
 - PowerExchange for Amazon Redshift for PowerCenter [52](#)
- troubleshooting overview [52](#)