

Amazon Redshift Import Integration

Learn more about [Amazon Redshift Export Integration](#).

Open You can connect Amazon Redshift to import data into Treasure Data.

This topic contains:

- [Prerequisites](#)
- [Use the TD Console to Create Your Connection](#)
 - [Create a New Connection](#)
 - [Transfer Your Redshift Data Account Data to Treasure Data](#)
 - [Connection](#)
 - [Data Settings](#)
 - [Data Preview](#)
 - [Data Placement](#)

Prerequisites

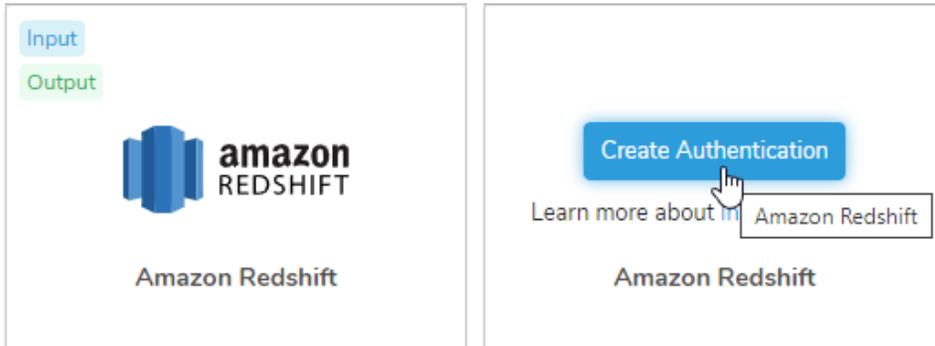
- Redshift instance created.
- Treasure Data TD Toolbelt installed.
- Basic knowledge of Amazon Redshift.
- Basic knowledge of Treasure Data, including the [TD Toolbelt](#).

Use the TD Console to Create Your Connection

Create a New Connection

When you configure a data connection, you provide authentication to access the integration. In Treasure Data, you configure the authentication and then specify the source information.

1. Open TD Console.
2. Navigate to **Integrations Hub > Catalog**
3. Click the search icon on the far-right of the Catalog screen, and enter **Amazon Redshift**.
4. Hover over the Amazon Redshift connector and select **Create Authentication**.



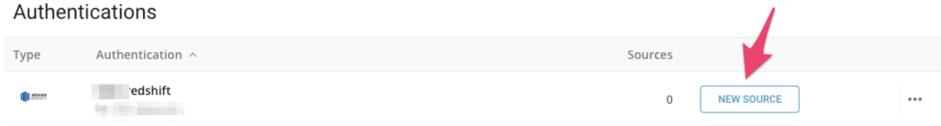
5. Specify the required credentials.
 - Host:** The host information of the source database, such as an IP address.
 - Port:** The connection port on the source instance. The PostgreSQL default is 5432.
 - User:** Username to connect to the source database.
 - Password:** Password to connect to the source database.
 - Use SSL:** Check this box to connect using SSL
 - JDBC Connection options:** Any special JDBC connections required by the source database (optional).
 - Region:** The AWS regions in which your Redshift instance is hosted.
 - Socket connection timeout:** Timeout (in seconds) for socket connection (default is 300).
 - Network timeout:** Timeout (in seconds) for network socket operations. 0 means no timeout.
 - Rows per batch:** Number of rows to fetch one time.
6. Select **Continue** after entering the required connection details.
7. Name the connection so you can find it later should you need to modify any of the connection details.
8. Optionally, select **Share with others**, if you would like to share this connection with other users in your organization.
9. Select **Done**.

If the connection is a success, then the connection appears in your list of authentications.

Transfer Your Redshift Data Account Data to Treasure Data

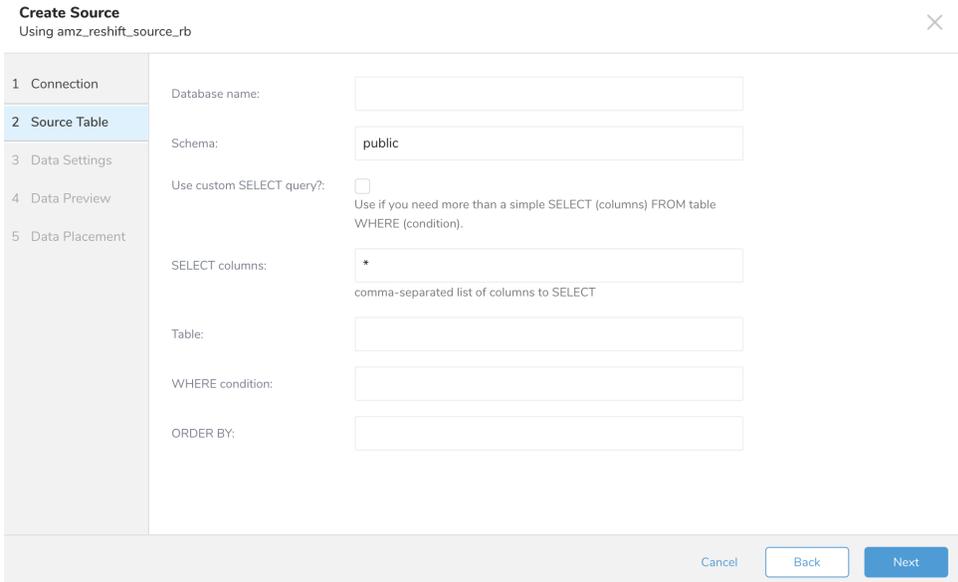
After creating the authenticated connection, you are automatically taken to Authentications.

1. Search for the connection you created.
2. Select **New Source**.



Connection

1. Type a name for your **Source** in the Data Transfer field.
2. Click **Next**.



3. Specify the details of the database and table that you want to ingest data from.
Database name: The name of the database you are transferring data from. (Ex. `your_database_name`)
Use custom SELECT query?: Use if you need more than a simple `SELECT (columns) FROM table WHERE (condition)`.
Schema: The schema to transfer data from.
SELECT columns: If there are only specific columns you would like to pull data from, list them here. Otherwise all columns are transferred.
Table: The table from which you would like to import the data.
WHERE condition: If you need additional specificity on the data retrieved from the table you can specify it here as part of `WHERE` clause.
ORDER BY: Specify if you need the records ordered by a particular field.

Data Settings

1. Select **Next**.
The Data Settings page opens.

Create Source

Using amz_redshift_auth2_rb

1 Connection

2 Source Table

3 Data Settings

4 Data Preview

5 Data Placement

Optionally, you can modify data settings and then see your changes in Data Preview. [Skip This Step](#)

Incremental?:
When run repeatedly, attempt to only import new data since the last import

Default timezone:

After SELECT:
This SQL will be executed after the SELECT query in the same transaction.

▼ COLUMN OPTIONS

Advanced options for individual columns.

2. Optionally, edit the data settings or skip this page of the dialog.
Incremental: When you want to repeatedly run this transfer, use the checkbox to import data only since the last time the import was run.
Default timezone: The timezone to be used when doing the import. Default is UTC.
After SELECT: This SQL is executed after the SELECT query in the same transaction.
Column Options: Select this option to modify the type of column before importing it.
3. Select **Next**.

Data Preview

You can see a [preview](#) of your data before running the import by selecting Generate Preview.

Data shown in the data preview is approximated from your source. It is not the actual data that is imported.

1. Click **Next**.
Data preview is optional and you can safely skip to the next page of the dialog if you want.
2. To preview your data, select **Generate Preview**. Optionally, click **Next**.
3. Verify that the data looks approximately like you expect it to.

Create Source
Using onetrust_demo

1 Connection

2 Source Table

3 Data Settings

4 Data Preview

5 Data Placement

The preview shows a subset of data from the source based on the data settings. Refer to [help document](#) to learn more about preview data.

	Ab_id	Ab_language	Ab_identifier	last_updated_date	Ab_link_token	
1	f7abf910-b5da-47c2-bbee-37f4c86...	NULL	Quan3	2020-09-25 22:42:59...	NULL	0
2	9022117f-cf3c-418c-b527-a8bd9a9...	NULL	Quan2	2020-08-05 03:48:19...	NULL	0
3	a432b52f-3d93-483b-b65f-3c7530...	NULL	Quan4	2020-08-05 03:48:19...	NULL	0
4	233ec0c2-70ab-4de4-ac48-a4a048f...	NULL	Quan5	2020-08-05 03:48:19...	NULL	0
5	f78be70b-8b5d-404e-b663-b606a2...	NULL	Quan1	2020-08-05 03:48:19...	NULL	0
6	db5d8f89-c264-4d82-a246-5939e5...	NULL	example@otrprivacy.com	2020-08-06 17:51:12...	NULL	[[T...
7	5ef9542c-315d-4b56-ad1c-c63ad0...	NULL	Michael.White@gmail.com	2020-09-09 20:01:45...	NULL	0
8	3f1dfcb9-1904-4517-9087-0cc45f0...	NULL	Robert.Brown@gmail.com	2020-09-09 20:01:45...	NULL	0
9	4a3a88dd-11a3-4c8b-a1d9-d7043f...	NULL	Mary.Anderson@mail.com	2020-09-09 20:01:46...	NULL	0
10	4fd8983a-9e49-46dc-9519-1cf9dea...	NULL	Elizabeth.Scott@gmail.com	2020-09-09 20:01:47...	NULL	0
11	33342e5d-4c95-4cfe-a622-4e91dc5...	NULL	David.Miller@aol.com	2020-09-09 20:01:47...	NULL	0
12	f54b0d7c-df75-4bf3-934a-dc19a96...	NULL	Robert.Anderson@att.com	2020-09-10 04:57:16...	NULL	0
13	43bfe156-dfba-43b8-964d-1b2a4ae...	NULL	Elizabeth.Miller@google.com	2020-09-10 04:57:16...	NULL	0

8 columns

4. Select **Next**.

Data Placement

For data placement, select the target database and table where you want your data placed and indicate how often the import should run.

1. Select **Next**. Under Storage you will create a new or select an existing database and create a new or select an existing table for where you want to place the imported data.

The screenshot shows a configuration window for data placement. On the left, a sidebar lists steps: 1 Connection, 2 Source Table, 3 Data Settings, 4 Data Preview, and 5 Data Placement (highlighted). The main area is titled 'STORAGE' and contains the following settings:

- Database:** chung_default_db
- Table:** sftp_v2_devproxy
- Method:**
 - Append: Add records into existing table.
 - Always Replace: Always clear the destination table before adding records.
 - Replace on new data: When there is new data, delete existing data, and insert new data.
- Timestamp-based Partition Key:** time
- Data Storage Timezone:** UTC (default)

Below the STORAGE section is the 'SCHEDULE' section:

- Repeat:**
 - Off
 - On
- Scheduling Timezone:** Asia/Saigon

2. Select a **Database** > **Select an existing** or **Create New Database**.
3. Optionally, type a database name.
4. Select a **Table**> **Select an existing** or **Create New Table**.
5. Optionally, type a table name.
6. Choose the method for importing the data.
 - **Append** (default)-Data import results are appended to the table. If the table does not exist, it will be created.
 - **Always Replace**-Replaces the entire content of an existing table with the result output of the query. If the table does not exist, a new table is created.
 - **Replace on New Data**-Only replace the entire content of an existing table with the result output when there is new data.
7. Select the **Timestamp-based Partition Key** column. If you want to set a different partition key seed than the default key, you can specify the long or timestamp column as the partitioning time. As a default time column, it uses upload_time with the add_time filter.
8. Select the **Timezone** for your data storage.
9. Under **Schedule**, you can choose when and how often you want to run this query.
 - Run once:
 - a. Select **Off**.
 - b. Select **Scheduling Timezone**.
 - c. Select **Create & Run Now**.
 - Repeat the query:
 - a. Select **On**.
 - b. Select the **Schedule**. The UI provides these four options: *@hourly*, *@daily* and *@monthly* or custom *cron*.
 - c. You can also select **Delay Transfer** and add a delay of execution time.
 - d. Select **Scheduling Timezone**.
 - e. Select **Create & Run Now**.

After your transfer has run, you can see the results of your transfer in **Data Workbench** > **Databases**.