

# Create Redshift Connections

## Contents:

- *Pre-requisites*
    - *Permissions*
  - *Limitations*
  - *Create Connection*
    - *Create through application*
    - *Create via API*
  - *Reference*
    - *Connection URL*
    - *Driver Information*
  - *Troubleshooting*
  - *Use*
  - *Testing*
- 

This section provides information on how to enable Redshift connectivity and create one or more connections to Redshift sources.

- Amazon Redshift is a hosted data warehouse available through Amazon Web Services. It is frequently used for hosting of datasets used by downstream analytic tools such as Tableau and Qlik. For more information, see <https://aws.amazon.com/redshift/>.

## Pre-requisites

Before you begin, please verify that your Trifacta® environment meets the following requirements:

**Tip:** If the credentials used to connect to S3 do not provide access to Redshift, you can create an independent IAM role to provide access from Redshift to S3. If this separate role is available, the Redshift connection uses it instead. There may be security considerations.

1. **S3 base storage layer:** Redshift access requires use of S3 as the base storage layer, which must be enabled. See *Set Base Storage Layer*.
2. **Same region:** The Redshift cluster must be in the same region as the default S3 bucket.
3. **Integration:** Your Trifacta instance is connected to a running environment supported by your product edition.
  
4. **Deployment:** Trifacta platform is deployed either on-premises or in EC2.

## Permissions

Access to Redshift requires:

- Each user is able to access S3
- S3 is the base storage layer

If the credentials used to connect to S3 do not provide access to Redshift, you can create an independent IAM role to provide access from Redshift to S3. If this separate role is available, the Redshift connection uses it instead.

**NOTE:** There may be security considerations with using an independent role to govern this capability.

**Steps:**

1. The IAM role must contain the required S3 permissions. See *Required AWS Account Permissions*.
2. The Redshift cluster should be assigned this IAM role. For more information, see <https://docs.aws.amazon.com/redshift/latest/mgmt/authorizing-redshift-service.html>.

**Limitations**

- You can publish any specific job once to Redshift through the export window. See *Publishing Dialog*.
- When publishing to Redshift through the Publishing dialog, output must be in Avro or JSON format. This limitation does not apply to direct writing to Redshift.
- Management of nulls:
  - Nulls are displayed as expected in the Trifacta application.
  - When Redshift jobs are run, the UNLOAD SQL command in Redshift converts all nulls to empty strings. Null values appear as empty strings in generated results, which can be confusing. This is a known issue with Redshift.

**Create Connection**

You can create Redshift connections through the following methods.

**Tip:** SSL connections are recommended. Details are below.

**Create through application**

Any user can create a Redshift connection through the application.

**Steps:**

1. Login to the application.
2. In the menu, click **User menu > Preferences > Connections**.
3. In the Create Connection page, click the Redshift connection card.
4. Specify the properties for your Redshift database connection. The following parameters are specific to Redshift connections:

Property	Description
IAM Role ARN for Redshift-S3 Connectivity	(Optional) You can specify an IAM role ARN that enables role-based connectivity between Redshift and the S3 bucket that is used as intermediate storage during Redshift bulk COPY/UNLOAD operations. Example: <pre>arn:aws:iam::1234567890:role/MyRedshiftRole</pre> For more information, see <i>Configure for EC2 Role-Based Authentication</i> .

For more information, see *Create Connection Window*.

5. Click **Save**.

### Enable SSL connections

To enable SSL connections to Redshift, you must enable them first on your Redshift cluster. For more information, see <https://docs.aws.amazon.com/redshift/latest/mgmt/connecting-ssl-support.html>.

In your connection to Redshift, please add the following string to your Connect String Options:

```
;ssl=true
```

Save your changes.

### Create via API

For more information, see <https://api.trifacta.com/ee/7.6/index.html#operation/createConnection>

## Reference

### Connection URL

The properties that you provide are inserted into the following URL, which connects Trifacta Self-Managed Enterprise Edition to the connection:

```
jdbc:redshift://<host>:<port>/<database><connect-string-options>
```

### Connect string options

The connect string options are optional. If you are passing additional properties and values to complete the connection, the connect string options must be structured in the following manner:

```
;<prop1>=<val1>&<prop2>=<val2>;...
```

where:

- `<prop>` : the name of the property
- `<val>` : the value for the property

### Delimiters:

- `;` : any set of connect string options must begin and end with a semi-colon.
- `;` : all additional property names must be prefixed with a semi-colon.
- `=` : property names and values must be separated with an equal sign (=).

### Driver Information

This connection uses the following driver:

- **Driver name:** `com.amazon.redshift.jdbc41.Driver`
- **Driver version:** `com.amazon.redshift:redshift-jdbc41-no-awssdk:1.2.41.1065`
- **Driver documentation:** <https://docs.aws.amazon.com/redshift/latest/mgmt/configure-jdbc-connection.html>

## Troubleshooting

For more information, see <https://docs.aws.amazon.com/redshift/latest/mgmt/troubleshooting-connections.html>.

## Use

For more information, see *Redshift Browser*.

For more information about interacting with data on Redshift, see *Using Redshift*.

## Testing

Import a dataset from Redshift. Add it to a flow, and specify a publishing action. Run a job.

**NOTE:** When publishing to Redshift through the Publishing dialog, output must be in Avro or JSON format. This limitation does not apply to direct writing to Redshift.