Create Redshift Connections

Contents:

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:

> **NOTE:** In the Admin Settings page are some deprecated parameters pertaining to Redshift. Please ignore these parameters and their settings. They do not apply to this release.

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.

Limitations

1. 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.
2.
   You can publish any specific job once to Redshift through the export window. See *Publishing Dialog*.
3.
4. Management of nulls:
    a. Nulls are displayed as expected in the Trifacta application.
    b. 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 **Settings menu > 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: `arn:aws:iam::1234567890:role/MyRedshiftRole` For more information, see *Configure for EC2 Role-Based Authentication*.

   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 *API Connections Create v4*.

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.

For more information, see *Verify Operations*.

After you have run your job, you can publish the results to Redshift through the Job Details page. See *Publishing Dialog*.