Enable Snowflake Connections

Contents:

This section describes how to integrate the Trifacta® platform with Snowflake databases.

- Snowflake provides a cloud-database datawarehouse designed for big data processing and analytics. For more information, see *https://www.snowflake.com*.

- For more information on supported versions, see *Connection Types*.

Limitations

> **NOTE:** This integration is supported only for deployments of Trifacta in customer-managed AWS infrastructures. These deployments must use S3 as the base storage layer. For more information, see *Supported Deployment Scenarios for AWS*.

- SSO connections are not supported.

Pre-requisites

- If you do not provide a stage database, then the Trifacta platform must create one for you in the default database. In this default database, you must include a schema named `PUBLIC`. For more information, please see the Snowflake documentation.

Enable

When relational connections are enabled, this connection type is automatically available. For more information, see *Enable Relational Connections*.

Configure

To create a Snowflake connection, you must enable the following feature. The job manifest feature enables the creation of a manifest file to track the set of temporary files written to S3 before publication to Snowflake.

> **NOTE:** To guarantee consistency you can enable consistent view on the EMR cluster. This step must be done when the cluster is created. See *Configure for EMR*.

Steps:

1. You can apply this change through the *Admin Settings Page* (recommended) or

```
trifacta-conf.json
```
. For more information, see *Platform Configuration Methods*.
2. Locate the following parameter and set it to `true`:

```
"feature.enableJobOutputManifest": true,
```

3. Save your changes and restart the platform.

Create Stage

In Snowflake terminology, a **stage** is a database object that points to an external location on S3. It must be an external stage containing access credentials.

- If a stage is used, it is typically the default bucket used on S3 for storage.

> **Tip:** You can specify a separate database to use for your stage.

> **NOTE:** For read-only connections to Snowflake, you must specify a Database for Stage. The connecting user must have write access to this database.

- If a stage is not specified, a temporary stage is created using the current user's AWS credentials.

> **NOTE:** Without a defined stage, you must have write permissions to the database from which you import. This database is used to create the temporary stage.

For more information on stages, see *https://docs.snowflake.net/manuals/sql-reference/sql/create-stage.html*.

In the Trifacta platform, the stage location is specified as part of creating the Snowflake connection.

Create Snowflake Connection

For more information, see *Create Snowflake Connections*.

Testing

Steps:

1. After you create your connection, load a small dataset based on a table in the connected Snowflake database.

> **NOTE:** For Snowflake connections, you must have write access to the database from which you are importing.

   See *Import Data Page*.
2. Perform a few simple transformations to the data. Run the job. See *Transformer Page*.
3. Verify the results.

For more information, see *Verify Operations*.