databricks

Using Databricks with Toad Data Point

Souvik Pratiher, CSE, Databricks
Email: souvik.pratiher@databricks.com
&
Vish Kamat, Sr. SA, Databricks
Email: vishwanath.kamat@databricks.com

databricks

Toad Data Point with Databricks

Toad Data Point is a single, easy-to-use tool that enables users to access all data sources. This solution is a cross-platform self-service data-integration tool. Toad Data Point simplifies data access, data preparation, and provisioning. Users can quickly and easily access, integrate, prepare and provision data. Toad Data Point supports Databricks as well as other popular tools

This article describes how to use your local development machine to install, configure, and use Toad Data Point to work with databases in Databricks.

Requirements

Before you setup and configure Toad Data Point, your local development machine must meet the following requirements:

- A Windows 64-bit operating system
- Download the Databricks ODBC Driver onto your local development machine
- Install the ODBC Driver by extracting the ***.msi*** file from the downloaded archive and installing it.
- You must have the following ***two*** permissions in case of
    - *Databricks Clusters*:
        - **Can Attach To** permission to connect to the running cluster.
        - **Can Restart** permission to automatically trigger the cluster to start if its state is terminated when connecting.
    - *SQL Endpoints*:
        - You need **Can Use** permission. The Databricks SQL endpoint will be automatically started if it was stopped.
- You must have a personal access token with the above mentioned permissions

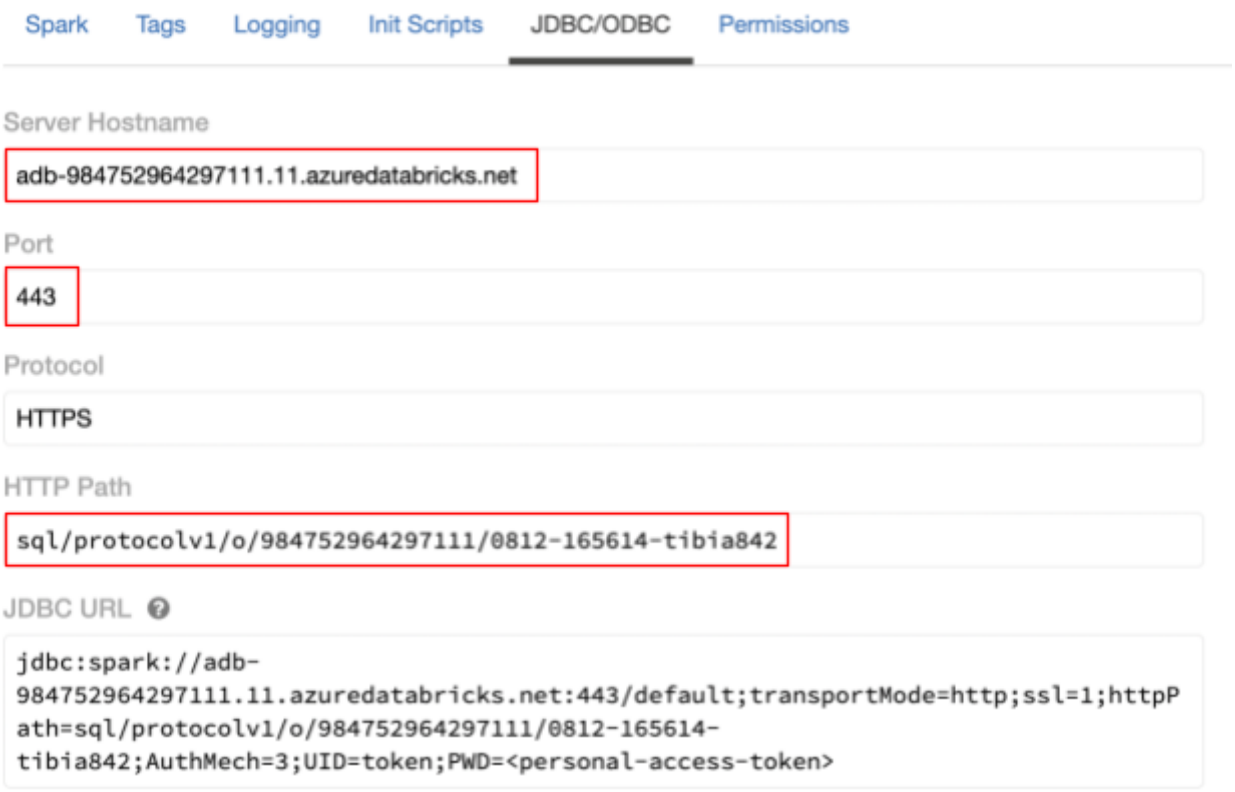
Step 1: Gather the required information for configuration

Follow the below steps to gather the connection configuration. The steps vary slightly based on which kind of environment we are gathering from (Databricks DS&E Workspace or DBSQL). Steps for both kinds of setup have been elaborated in the next page.

- For Databricks Clusters:
    a. Click **Compute** in the sidebar.
    b. Choose a cluster to connect to.
    c. Navigate to **Advanced Options**.
    d. Click on the **JDBC/ODBC** tab.
    e. Copy the connection details.

- For SQL Endpoints:
    a. Click **SQL Endpoints** in the sidebar.
    b. Choose an endpoint to connect to.
    c. Navigate to the Connection Details tab.
    d. Copy the connection details.

Below is the pictorial clue (for Databricks Clusters) on what are we talking about here:



The steps are the same for SQL Endpoints, the same details (server hostname, port, http path) are available there also.

databricks

Step 2: Configuring the ODBC driver

As part of the pre-requisites mentioned above after the *.msi* file has been installed and a personal access token has been acquired with the required permission(s), we have to now configure the DSN (Data Source Name). Follow the steps below:

- From the Start menu, search for **ODBC Data Sources (64-bit)** to launch the **ODBC Data Source Administrator**.
- Navigate to the **Drivers** tab to verify that the driver (Simba Spark ODBC Driver) is installed.
- Go to the **User DSN** or **System DSN** tab and in one of them **Simba Spark ODBC Driver** should show up.
- Select it and click on **Configure.** Before configuration make sure you have the following values in handy (server hostname, port, http path).

  Example configuration values:

  Host(s): adb-8159029426904320.0.azuredatabricks.net

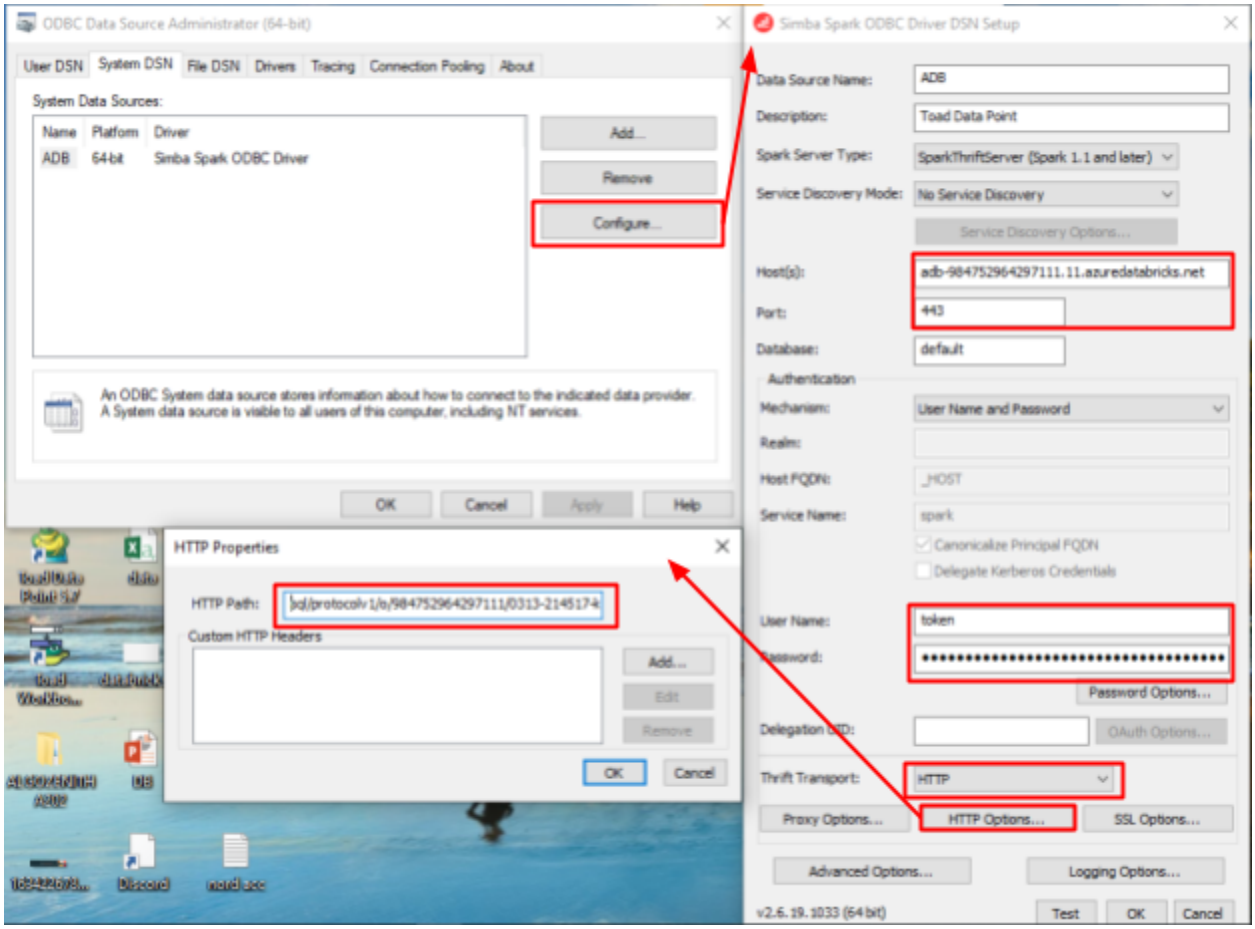
  Port: 443

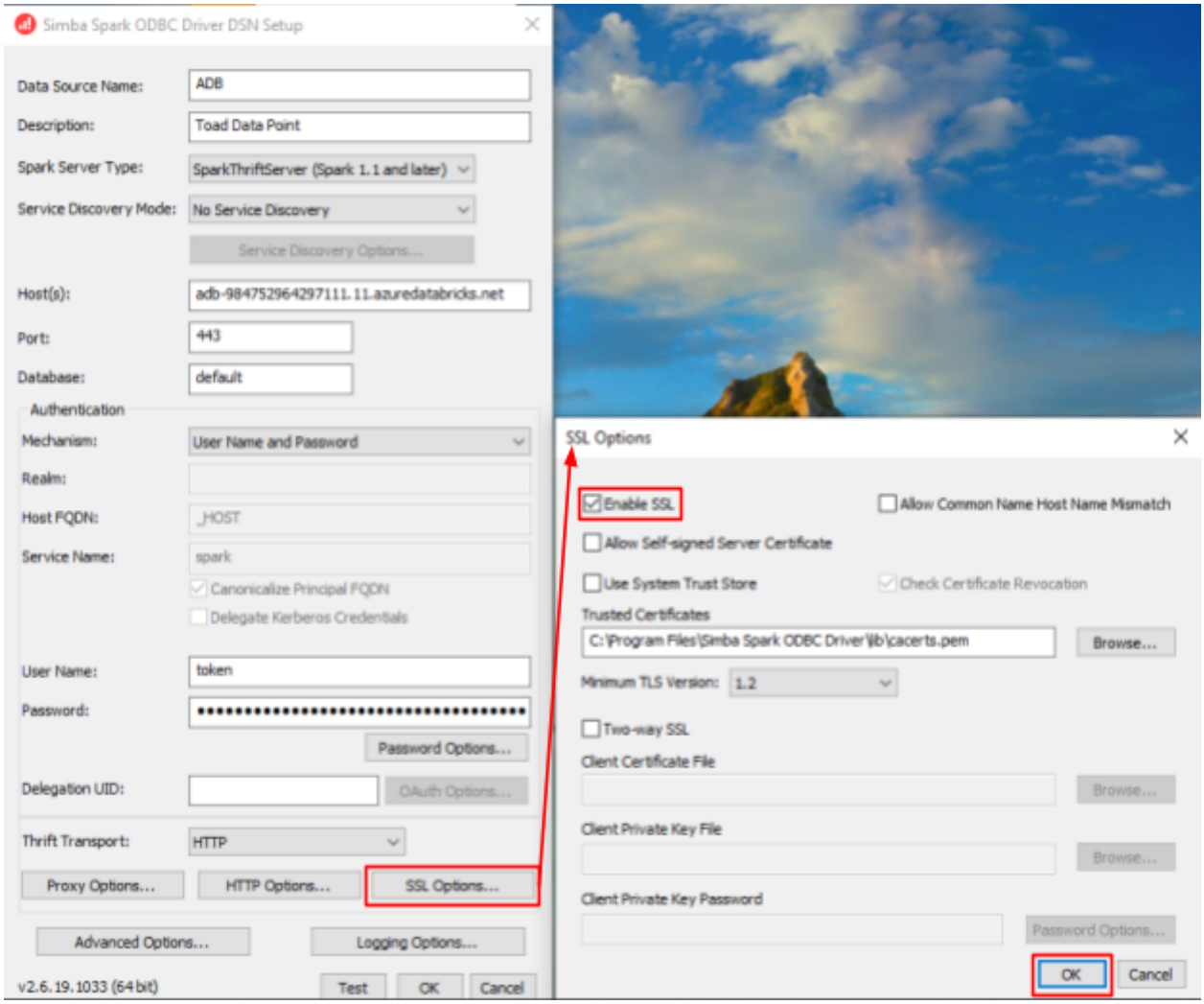
  User Name: token

  Password: <personal-access-token>

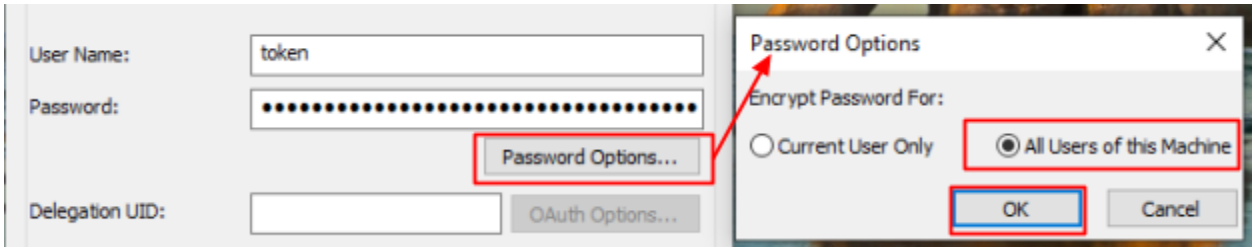
  HTTP Path: sql/protocolv1/o/8159029426904320/0925-105713-bite618

- In the configuration window populate the information as shown below:

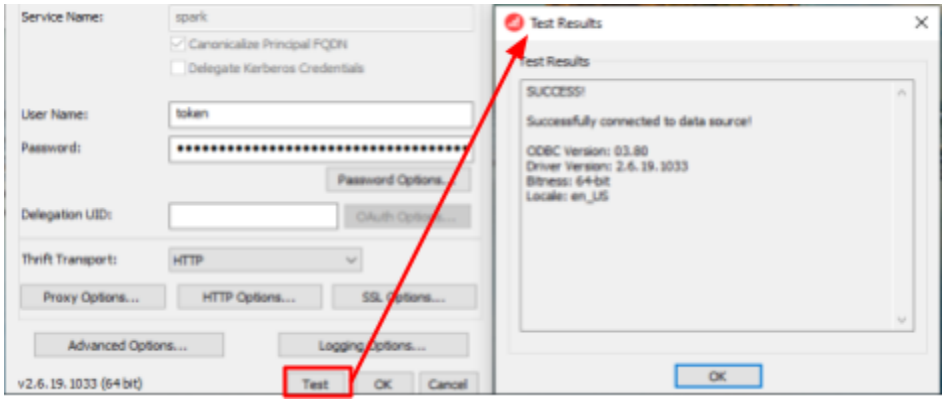




Now the ***username will be "token" without quotes*** and ***password will be the personal-access-token that you have***. And it needs to be saved with encryption. So do the following also:



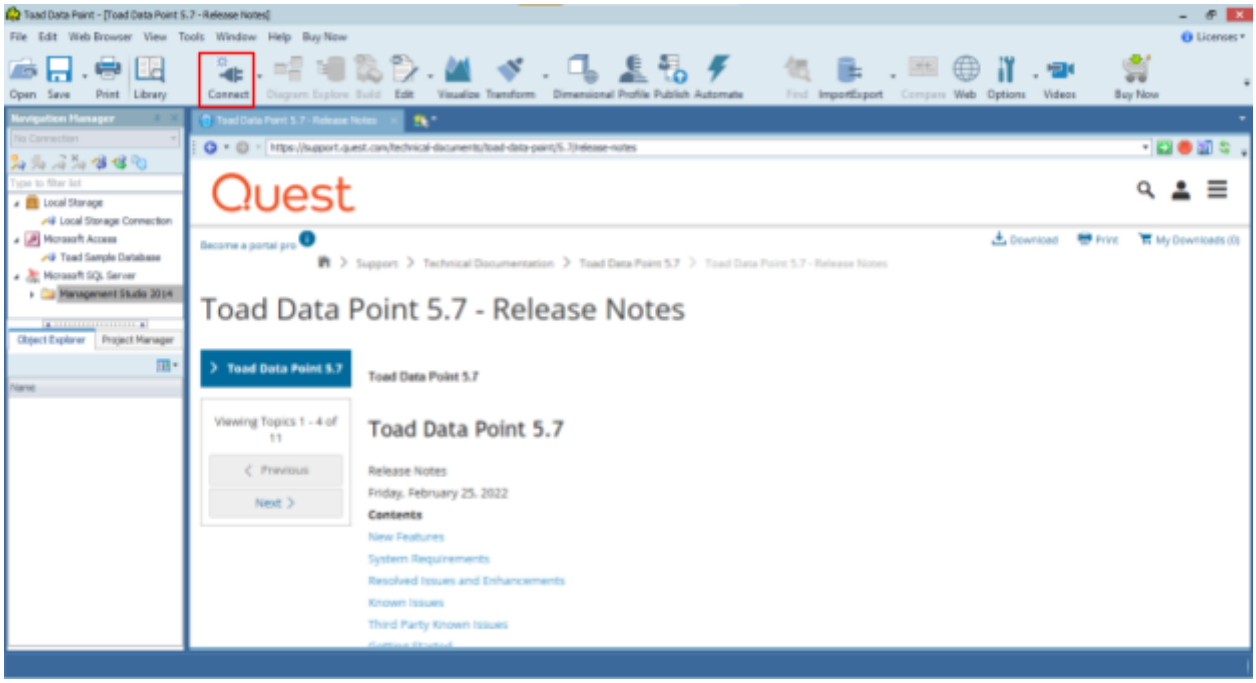
Finally do the ***test connection*** before ***clicking on OK to complete*** the configuration



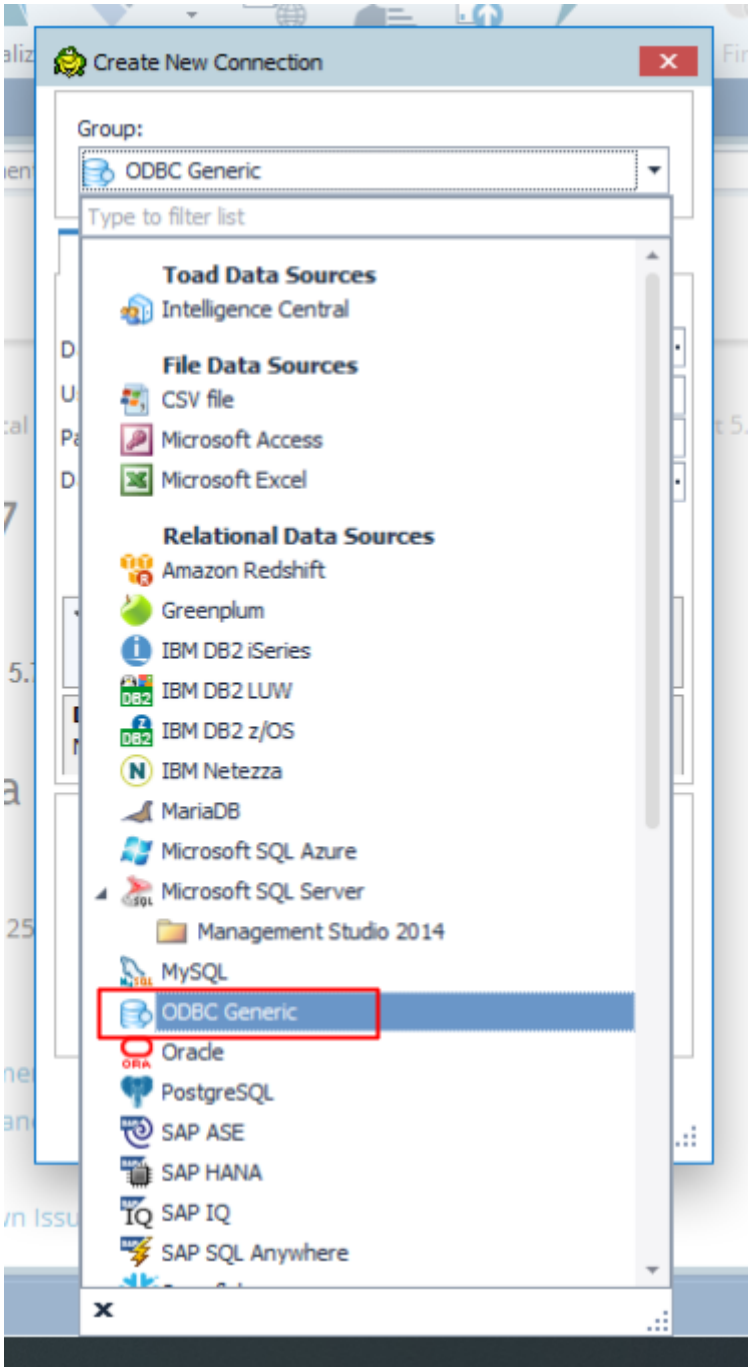
Step 3: Using the newly configured DSN in Toad

After configuration of the DSN as per Step 2. We are now ready to use it to access our Databricks Cluster in Toad Data Point. Only thing that is left now is to use the source and setup the Toad connection from the Toad UI. Follow the below mentioned steps for it:

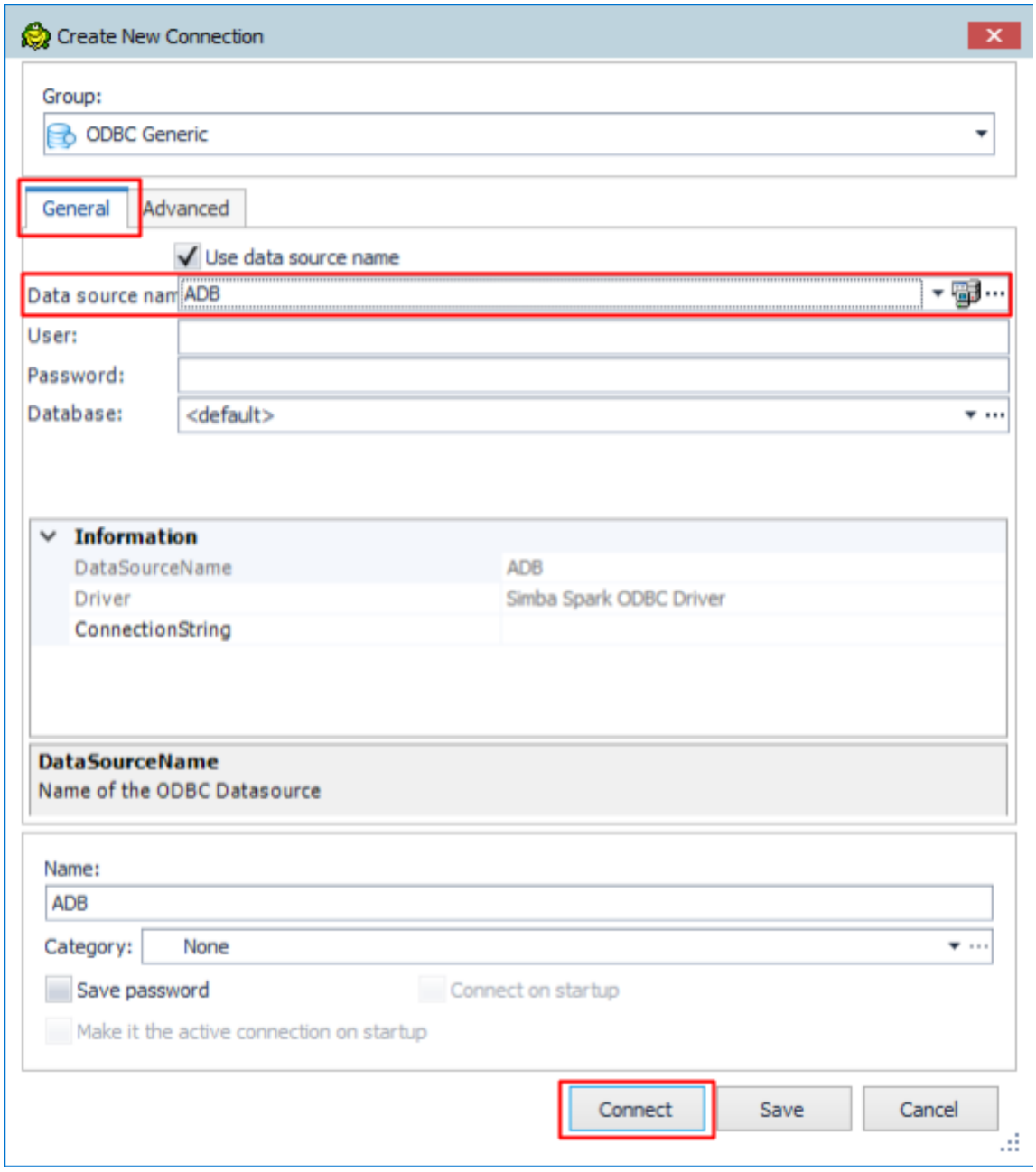
- Click on Connect button in the Toad UI.



- From the **Create New Connection UI**, select **ODBC Generic** under **Relational Data Sources.**



- After selecting the **ODBC Generic Source**, in the **window under the general tab** select the correct **Data Source Name**(the one configured as per Step 2) and hit **Connect**.



- And Voila! You are now connected to your Databricks Cluster or SQL Endpoint from Toad.