Quickstart: Simplify Data Integration for the Modern Data Stack with Databricks & Fivetran

Summary	Simplify Data Integration for the Modern Data Stack with Databricks & Fivetran
ID	ingest-into-databricks-lakehouse-using-fivetran
URL	Ingest-into-databricks-lakehouse-using-fivetran go/fivetranquickstart
Category	Web
Status	Complete
Feedback Link	
Author	Prasad Kona and Yevindra De Silva

Simplify Data Integration for the Modern Data Stack with Databricks & Fivetran

Hands-on Workshop

Author	Prasad Kona & Yevindra De Silva
Version	1.0.5
Date	07/22/2022
Change History	New introductory workshop for Ingesting data into Databricks Lakehouse using Fivetran

Overview

Ingest data into the Databricks lakehouse using Fivetran

Fivetran provides fully managed connectors that help create pipelines that automatically extract data from various sources. These pipelines continuously ingest and update data in the Databricks Lakehouse and enable organizations to focus on their use cases around data engineering, data science, machine learning and Business Intelligence..

Agenda:

- Introductions and overview of Databricks and Fivetran
- Access Fivetran from Partner Connect
- Build a data ingestion pipeline using Fivetran & Databricks
- Explore ingested data in Databricks SQL database and tables
- Create & execute queries and visualizations using Databricks SQL
- Create and share dashboards

What you'll learn

- How to ingest data into Databricks lakehouse.
- How to extract data using any of 150+ sources using Fivetran's fully managed connectors and ingest into Databricks lakehouse.

What you'll need

- A Databricks Account. Please sign up here if you are not already a Databricks user.

What you'll Build

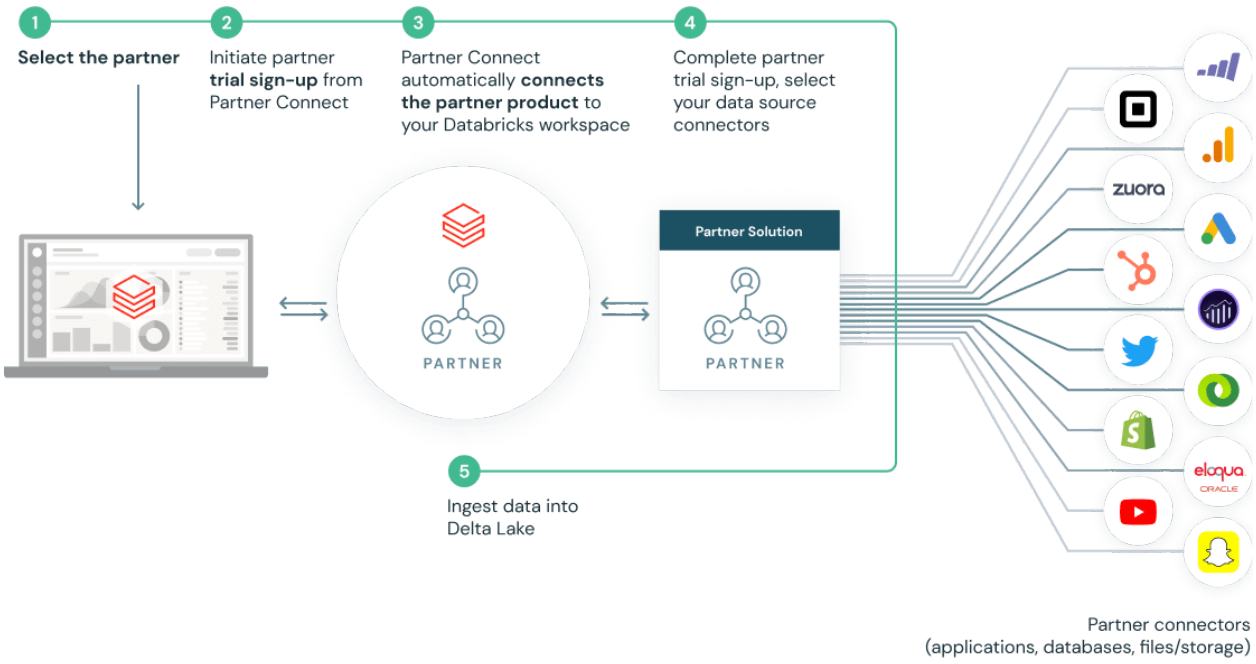
- You will use Fivetran's fully managed connectors to extract data from Salesforce and ingest into the Databricks Lakehouse.

Introduction to Databricks Partner Connect

Partner Connect makes it easy for you to discover data, analytics and AI tools directly within the Databricks platform — and quickly integrate the tools you already use today. With Partner Connect, you can simplify tool integration to just a few clicks and rapidly expand the capabilities of your lakehouse. It is a one-stop portal for validated partner solutions.

Partner Connect simplifies your integrations by automatically configuring resources — including clusters, tokens and connection files ; to connect with partner solutions

Fivetran is available on Databricks Partner Connect



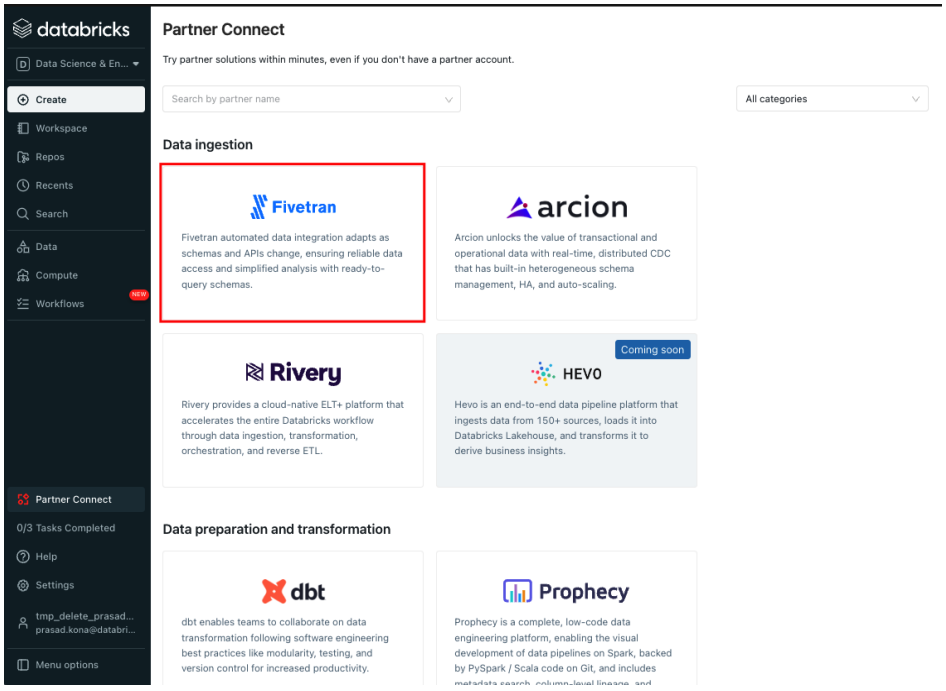
Introduction to Fivetran

Fivetran provides fully managed connectors that help create pipelines that automatically extract data from various sources. These pipelines continuously ingest and update data in the Databricks Lakehouse and enable organizations to focus on their use cases around data engineering, data science, machine learning and Business Intelligence..

Pre-Lab: Initiate Fivetran Trial using Databricks Partner Connect (Done in advance)

In this lab, you will use Databricks Partner connect to initiate a Fivetran trail.

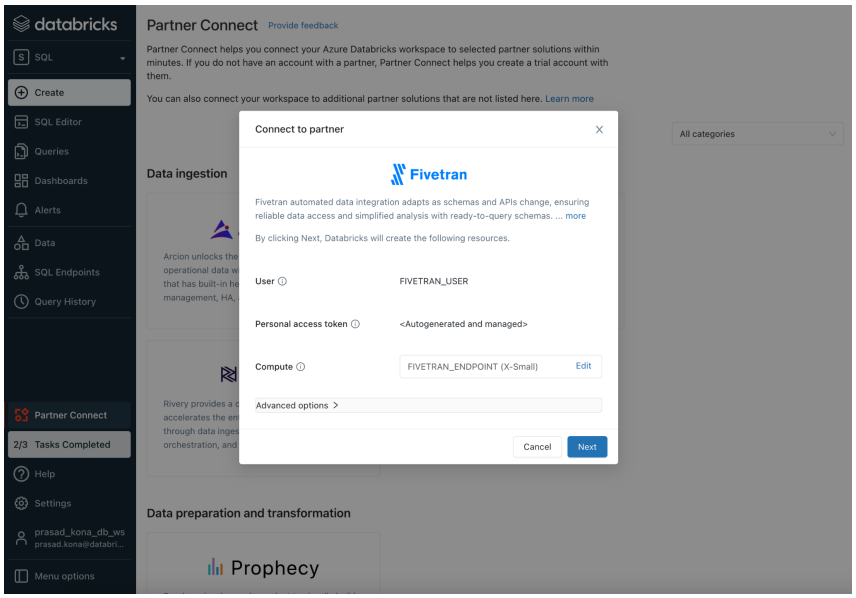
1. Click to open the Databricks Partner Connect UI
2. Click to initiate the trial with Fivetran. Notice there is no Tick mark ( ) next to Fivetran)
3. If you notice the Tick mark ( ) next to “Fivetran”. This means that the connection to Fivetran has already been established
    a. Clicking on the “Fivetran” tile would open a new tab to login to Fivetran and take you to the Add Source Connector page on Fivetran. Go to “Configuring the Fivetran source connector for Salesforce”
4. Click on Fivetran tile on Partner Connect UI



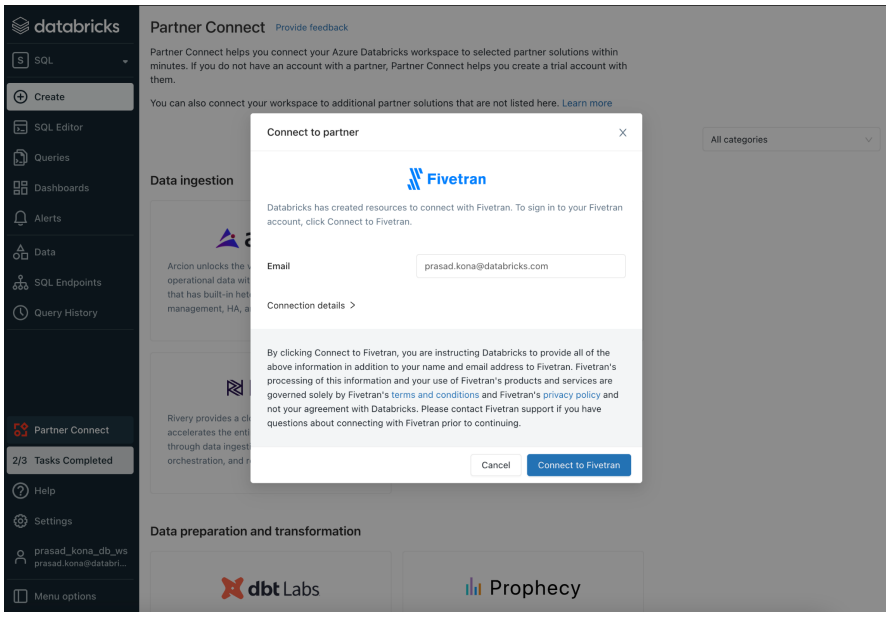
5. The Connect to partner dialog displays around Fivetran and details around resources that are going to be pre configured.
6. As part of the partner connect integration

a. Databricks creates a preconfigured SQL endpoint, service user and access credentials
b. Fivetran creates a pre-configured Databricks destination using the information Databricks provided

7. On the Connect to partner dialog click on the Next button.



8. The username to be shared with Fivetran is displayed on the screen.
9. Click the button with the label “Connect to Fivetran” or “Sign in”.

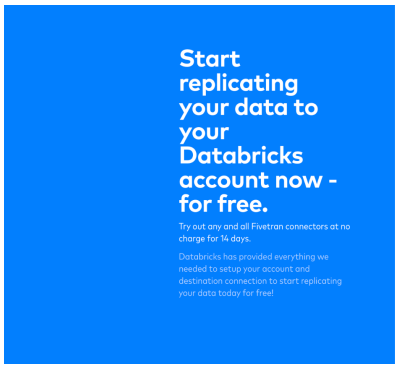


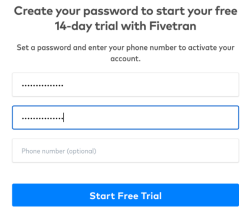
10. This opens a new tab on your browser, which takes you to the final step to initiate Fivetran trial

11. Complete the on-screen instructions in Fivetran to create your 14-day trial Fivetran account, or to sign in to your existing Fivetran account.

12. Do one of the following:

    - If Fivetran just created your 14-day trial Fivetran account, continue with Use a new 14-day trial Fivetran account.
    - If you signed in to your existing Fivetran account, skip ahead to Use an existing Fivetran account.

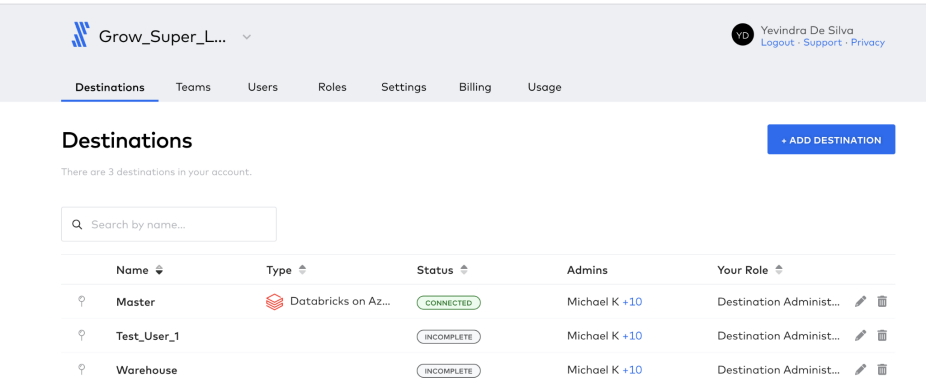




Login to Fivetran

Access Fivetran by navigating to the Manage Account page in Fivetran

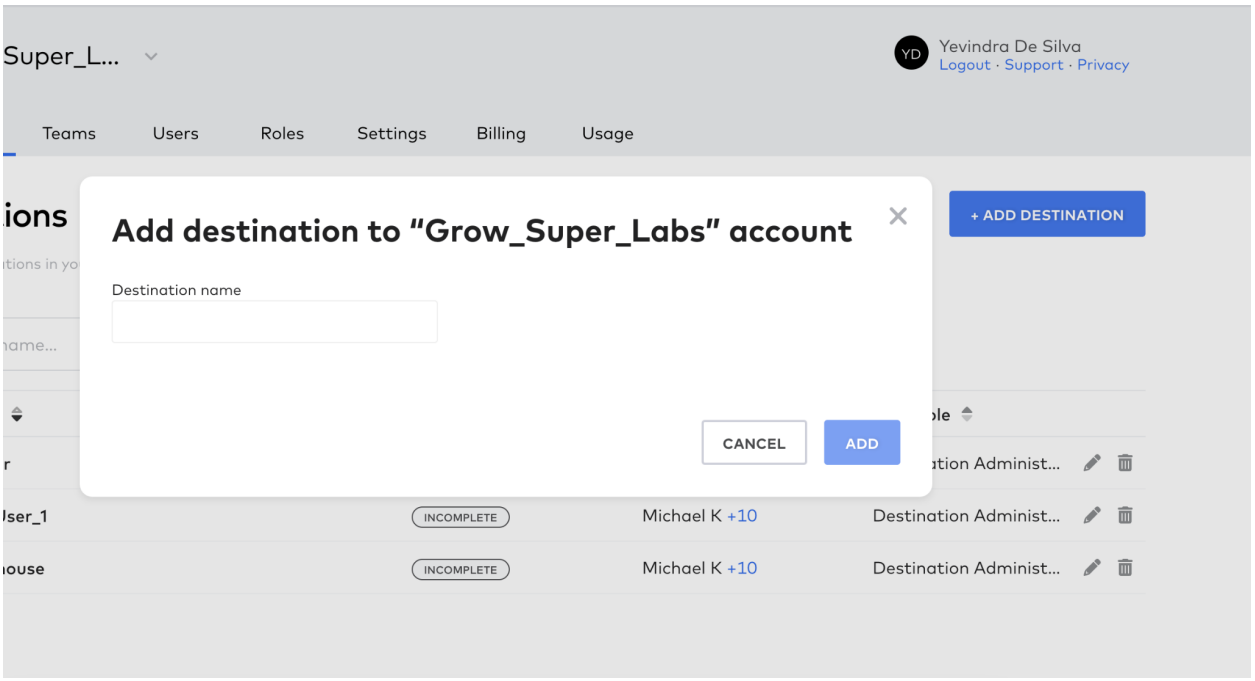
1. Use the credentials you've created to login to Fivetran



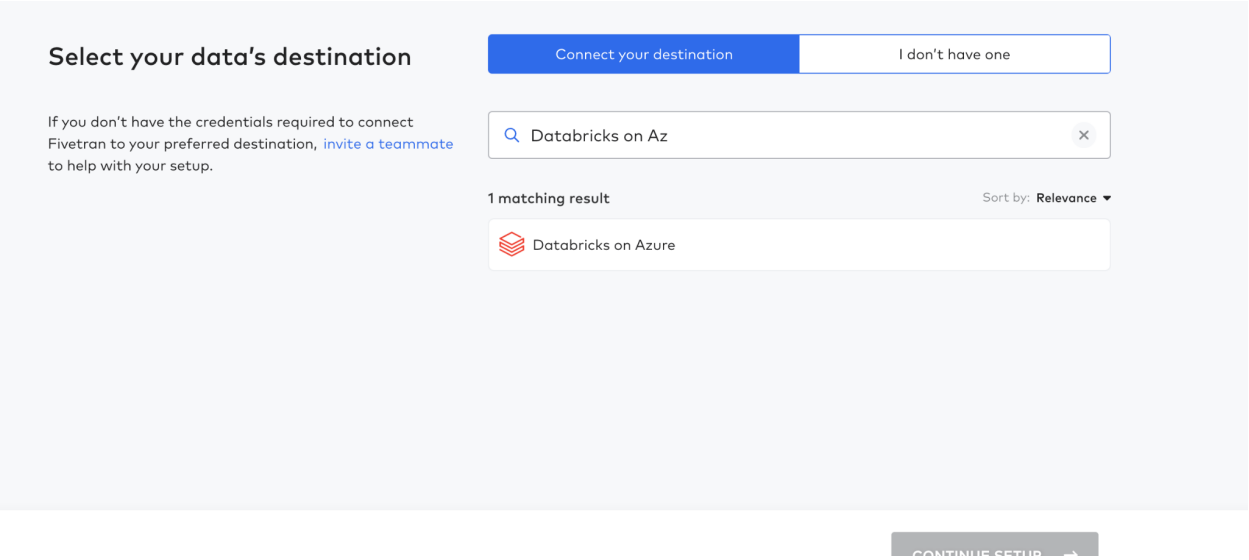
Lab 1: Configuring a Databricks Destination

In this lab, you will configure a new Destination for you to bring in data

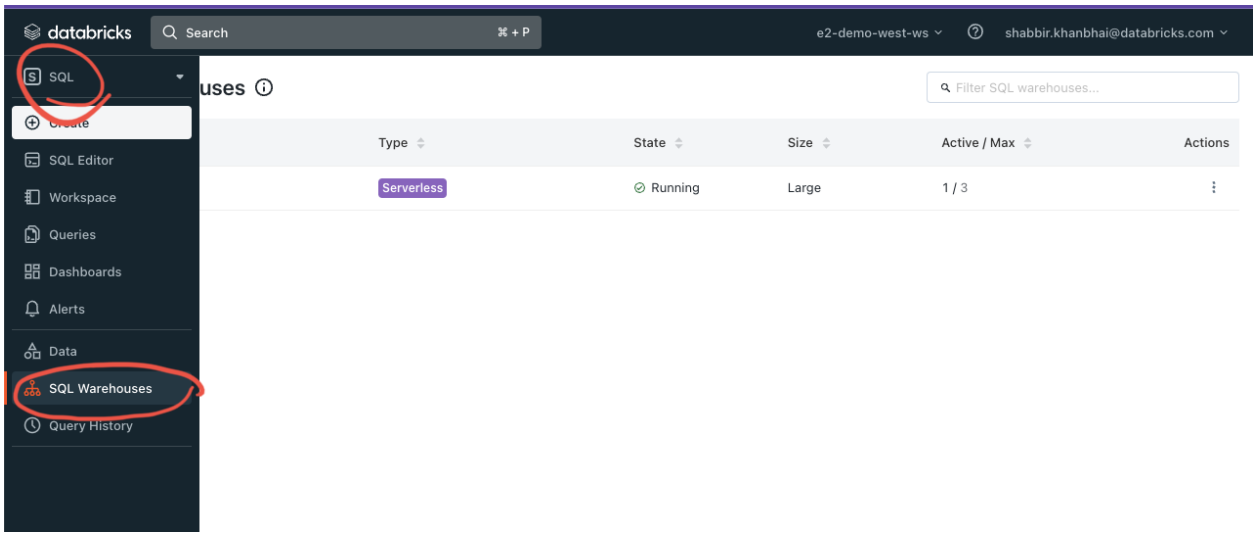
1. Click Add Destination



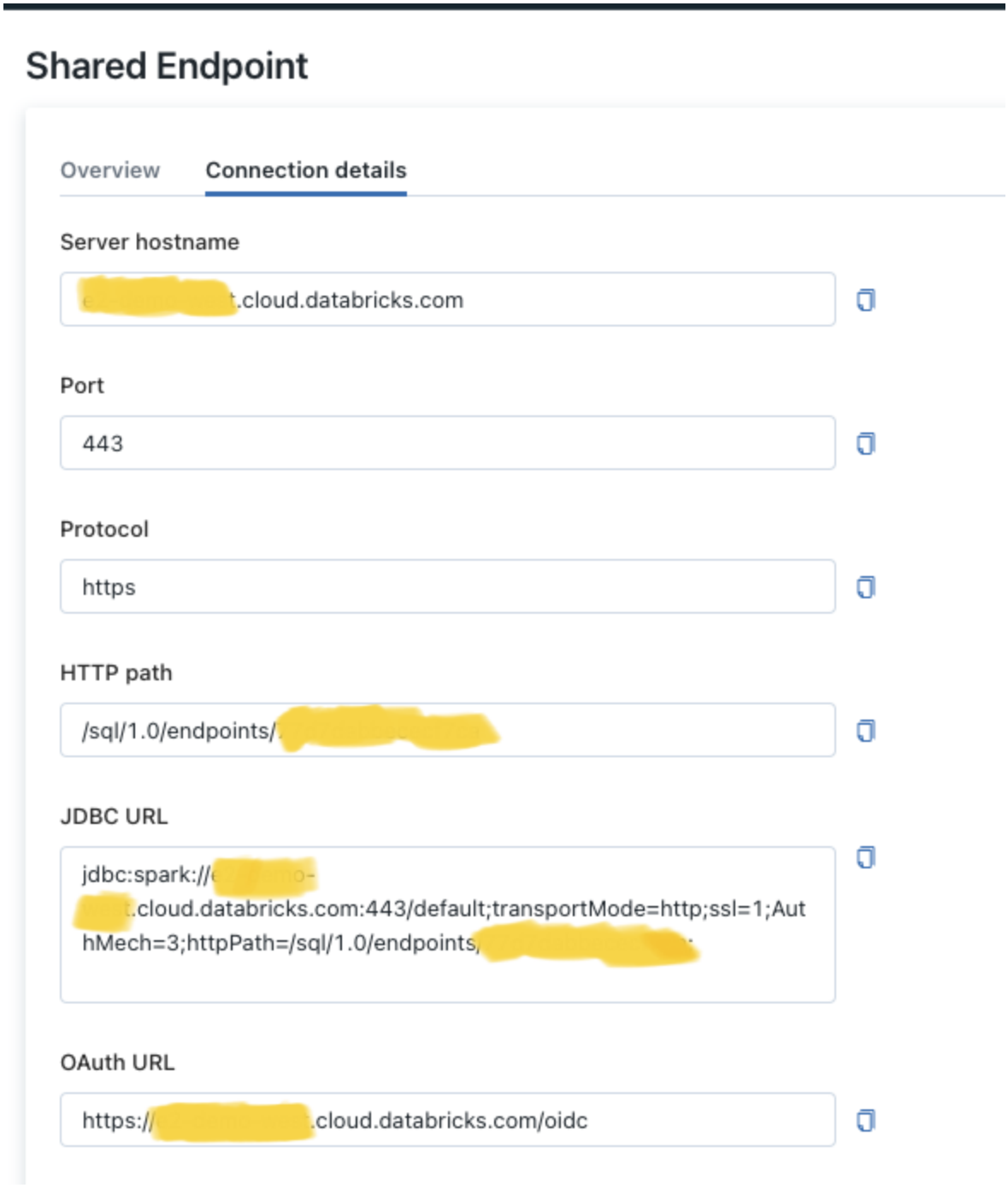
2. Follow the following instructions to setup a Databricks Destination



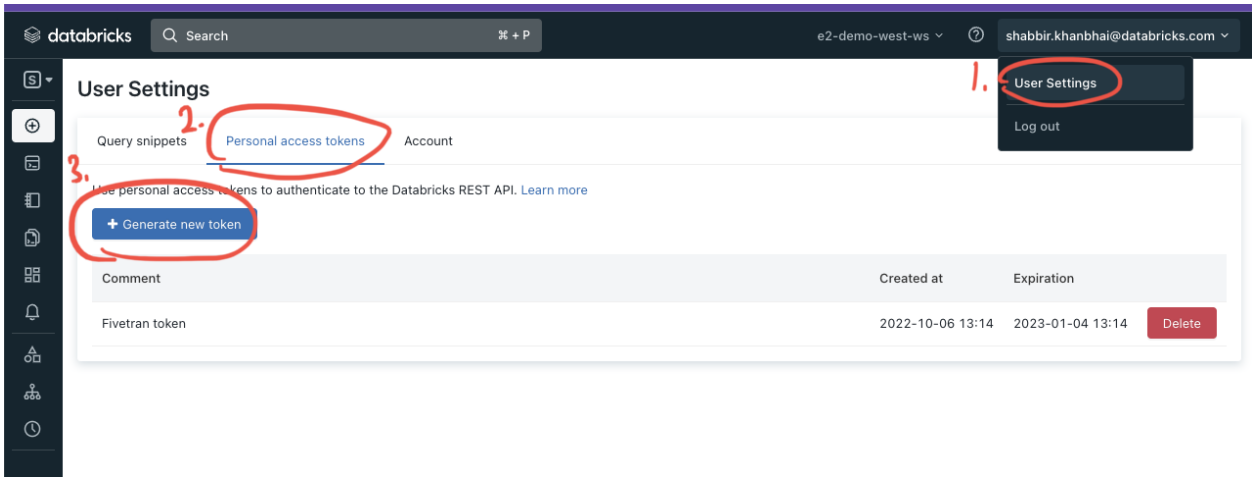
- Leverage existing SQL Warehouse details provided (or create your own!) (SQL Persona -> SQL Warehouses)



- Navigate to the warehouse's "Connection details" and get the details from there
    - You will need to copy the "Server hostname", "Port" and "HTTP path" of the SQL Warehouse


Shared Endpoint
Overview
Connection details
Server hostname
.cloud.databricks.com
Port
443
Protocol
https
HTTP path
/sql/1.0/endpoints/
JDBC URL
jdbc:spark://
.cloud.databricks.com:443/default;transportMode=http;ssl=1;Aut
hMech=3;httpPath=/sql/1.0/endpoints,
OAuth URL
https://
.cloud.databricks.com/oidc

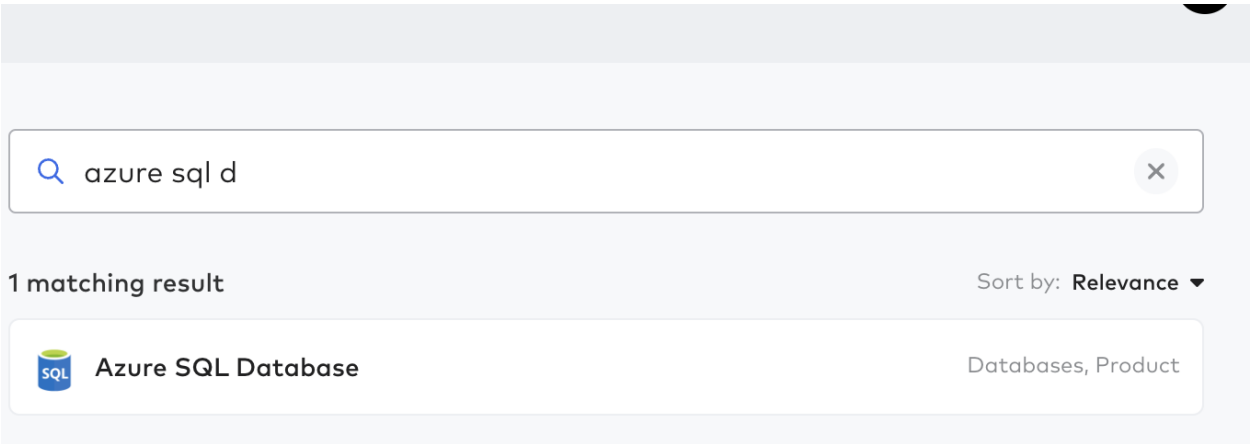
- You will need to create a new personal access token (Settings -> User Settings)



Lab 2: Create a Database Connector

In this lab, you will configure a Databases Connector.
We have preconfigured an Azure SQL Database which you will now connect to using the below details.

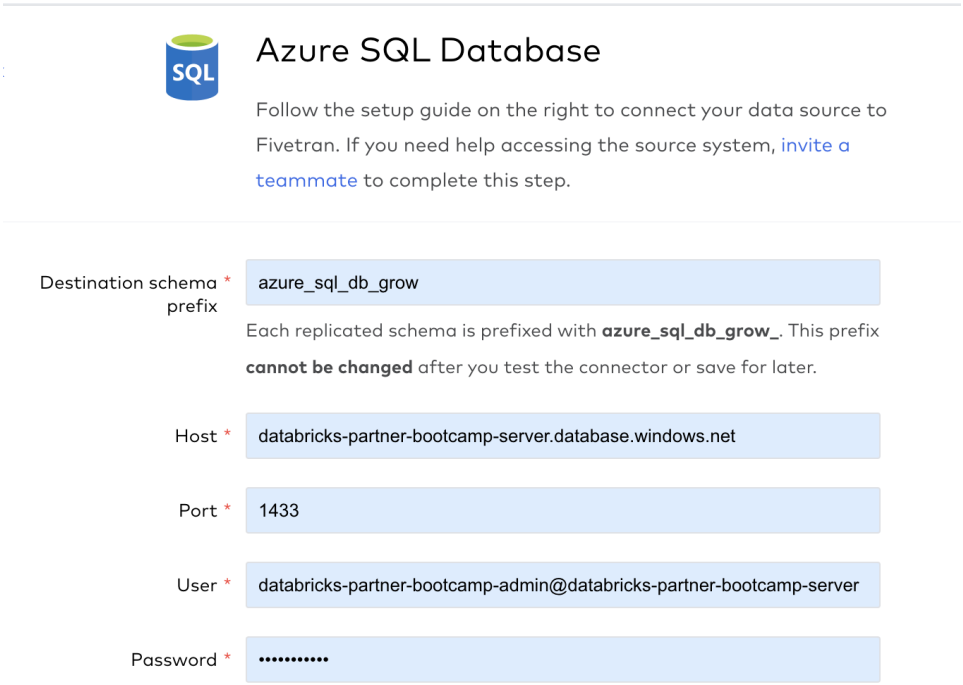
1. Add a connector and navigate to Azure SQL Database Connector



2. To authenticate to the preconfigured database, we will need to insert the following admin details and then click “Save & Test”

Destination schema prefix	<your_name_here>
Host	databricks-partner-bootcamp-server.database.windows.net

Database	databricks-partner-bootcamp-db
Port	1433
User	databricks-partner-bootcamp-admin@databricks-partner-bootcamp-server
Password	<see slides>



3. Select all the tables and follow the wizard to setup the connector
4. Read through the Setup guide to understand what normally needs to be setup on the database prior to setting up connector