Microsoft Fabric in a Day Lab Manual – Lab 4

Prepared by: Will Crayger

# Data Engineering – Data Transformation and Engineering in Fabric

## Introduction:

In this lab, you’re going to create a **Silver** Lakehouse and begin loading data from the **Bronze** Lakehouse that was created in **Lab 1**. As you proceed through the lab, please note that you’re working in a case-sensitive environment.

## Part 1: Creating the Silver Lakehouse

**Creating the Silver Lakehouse:** The purpose of creating a Silver Lakehouse is to simulate the **medallion development pattern**. The Silver Lakehouse will provide a separation from the Bronze layer, allowing you to start cleansing and shaping the data.

To begin, select **New** **item** from the Workspace home page. Scroll down and select **Lakehouse** from the item list. From the tiles at the top of the page, select **Lakehouse**. Name your lakehouse **silver\_lakehouse\_<your\_initials>** and click **Create**.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Creating a Shortcut to Azure Data Lake Gen 2:** To access semi-curated invoice data, you’ll need to create a shortcut to the Azure Data Lake Gen 2.

Use the following connection details provided in the Shortcut Connection Details file provided as part of the coursework.

After entering these details, select the **invoices** folder and click **Next**. Once everything is set, click **Create** to finalize the shortcut setup.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

## Part 2: Creating a New Notebook

**Mounting the Notebook:** From your **Silver Lakehouse**, click **Open Notebook** and select **New Notebook**. Rename the notebook to **notebook\_write\_to\_silver**.

Next, click **Lakehouses** from the **Explorer** blade. Click the **+ Lakehouse** button, check the **Existing Lakehouse** radial button, and click **Add**. Select the **Bronze Lakehouse** and click **Add** again.

The tables from the **Bronze Lakehouse** will now be visible in the notebook, ready for transformation.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Writing Data to Silver:** Click and drag **application\_people** from the **Explorer** to the first cell in the notebook. Remove the **LIMIT 1000** clause from the cell and press **Shift + Enter** to execute the cell. This action will start a Spark session using the default **single\_node** Spark Pool that was created earlier.

Once the cell executes, a preview of the results will be displayed. To filter the dataframe and include only individuals identified as a salesperson, use the following command:

**df\_salesperson = df.filter('IsSalesperson = true')**

**display(df\_salesperson)**

Next, write the new dataframe to the **Silver Lakehouse** using the following command then switch to the **Silver Lakehouse** from the **Notebook Explorer** to confirm that the table was successfully written.

**df\_salesperson.write.format('delta').mode('overwrite').saveAsTable('silver\_lakehouse\_wtc.salesperson')**

A screen shot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

## Part 3: Importing a Notebook

**Importing a Notebook:** In this section, the **Notebook** will serve as the primary documentation source, complete with comments for visibility and clarity, along with code examples that can be referenced to complete each exercise. The notebook used for these exercises is the **notebook\_spark\_engineering.ipynb** file, which was provided as part of the course material.

To set up the environment and launch the notebook, follow these steps:

Navigate to the **Data Engineering** landing page of **Fabric**. Click **Import Notebook** from the actions across the top. Then click **Upload**, navigate to the location where the course files were saved, select the **notebook\_spark\_engineering.ipynb**, and click **Open**. After this, return to the workspace home page, and you should see the notebook listed as one of the workspace artifacts.

A screen shot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Attach the Notebook to Bronze and Silver Lakehouses:** Navigate back to your workspace and click the notebook to open it. In the **Add Lakehouse** blade, click **Add**. Select **Existing Lakehouse** and click **Add** again. Choose the **Bronze** **Lakehouse** created in **Lab 1** as well as the **Silver Lakehouse** from **Part 1** of this lab, then click **Add**.

The remainder of this section will be completed using the provided notebook. The notebook itself will serve as the documentation to complete part 3.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

## Part 3: Data Engineering via Lakehouse SQL Endpoint

**Using the Lakehouse SQL Endpoint:** We can also leverage the **Lakehouse SQL Analytics Endpoint** to transform and prepare our data. In this section, you’ll use cross-database joins to query data from the **Bronze Lakehouse** and create a view in the **Silver Lakehouse**.

To begin, navigate to the **Silver Lakehouse SQL Analytics Endpoint** from the workspace landing page. Click the **+ Warehouses** button at the top of the **Explorer**. Then, check the box for your **Lab 1 Lakehouse** and click **Connect**. You should now see the schema from the **Lab 1 Lakehouse** available in the Explorer.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

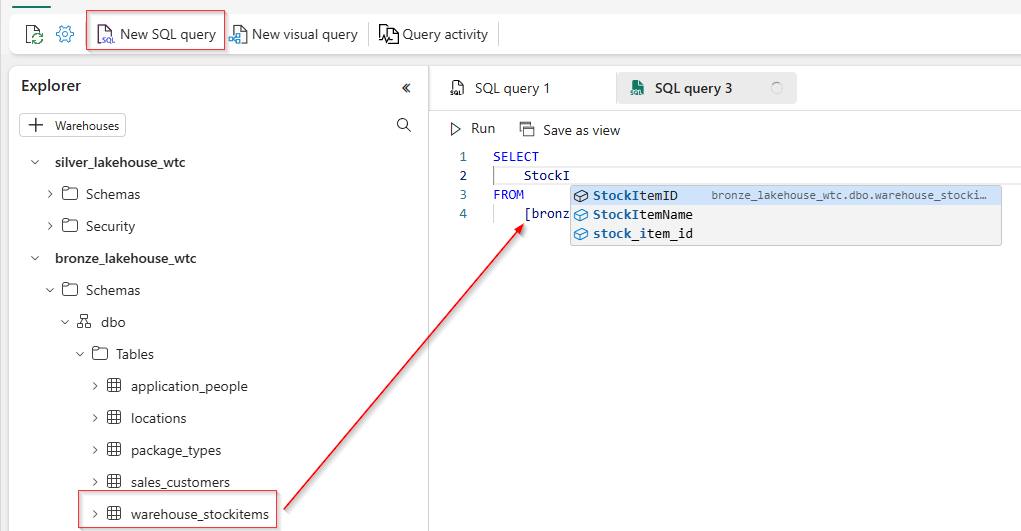
Description automatically generated

A screenshot of a computer

Description automatically generated

Next, select **New SQL Query** from the action ribbon at the top. Begin building your query by writing a **SELECT FROM** statement and dragging **warehouse\_stockitems** from the **Bronze Lakehouse** schema into the query window.

Update the **SELECT** statement to include only the following columns: **StockItemID**, **StockItemName**, **Brand,** and **SearchDetails**. You can drag these columns to the query window or type them in directly. As you begin typing, you’ll notice **Intellisense** working to identify the schema and help you with the query writing.



Once your query is complete, it should look similar to the image provided below with aliasing. Execute the query by clicking **Run** and review the output.

With the query complete, copy the query from the query window and click **Save as view**. Name the view **stock\_items** and click **OK**. **Note, the code must be highlighted for the Save as view option to work.**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Create another **View** for **customers** using the following **CustomerID**, **CustomerName**, **CityName**, **StateProvinceCode**, and **PhoneNumber**.

A screenshot of a computer

Description automatically generated

You have successfully completed **Part 3** of this lab.