A close-up of a person smiling

Description automatically generated

# Lab 02: Data warehouse Persona

### 

### Introduction

The primary objective of this lab is to show activities that are usually performed by Data Warehouse engineers. You will create tables, load data into the tables, query tables from a Warehouse and a Lakehouse, use visual query, do cross workspace and cross database query.

### Objectives

After completing this lab, you will be able to:

1. Use T-SQL to create and query tables from a Warehouse and a Lakehouse
2. Use visual query to join 2 tables
3. Perform cross workspace and cross database query

**Estimated time to complete this lab**

60 minutes

Contents

[Lab 02: Data warehouse Persona 1](#_Toc150964604)

[Introduction 1](#_Toc150964605)

[Objectives 1](#_Toc150964606)

[Task 1: Create a lake house 3](#_Toc150964607)

[Task 2: Create a table using T-SQL 4](#_Toc150964608)

[Task 3: Load data in the lake house 5](#_Toc150964609)

[Task 4: Load data into a table using Data Flow gen 2 7](#_Toc150964610)

[Task 6: Visual Query 10](#_Toc150964611)

[Task 7: Cross-workspace, cross database query 14](#_Toc150964612)

**Lab Prerequisites**

* Workspace: Fabric, Power Premium or Fabric trial
* Individual license: Power Pro or Premium Per User account
* Fabric Warehouse. You should have already created a warehouse in the previous lab. If you don’t have one already, please follow the instructions [here](https://learn.microsoft.com/en-us/fabric/data-warehouse/create-warehouse) to create a warehouse.

**Information provided by your training provider**

* Trial tenant (if applicable): login & password, workspace to use for the lab.

### Task 1: Create a lake house

This lake house will be used to show cross-query between lake house and data warehouse later.

Note: make sure to create the lake house in the same workspace as the data warehouse.

1. Go to the **Data Engineering** **experience**, create a new lake house and give it a name

A screenshot of a computer

Description automatically generated

### Task 2: Create a table using T-SQL

1. Go to the **Data Warehouse** **experience** and open your data warehouse.
2. Under **Queries**, create a new SQL query and paste the SQL queries from file **Tables.sql** (located in the folder Labs\scripts).

A screenshot of a computer

Description automatically generated

1. Execute the first 2 T-SQL statements, until line 34, on the data warehouse.

The query **creates a table** **and loads data** from a public dataset. This takes 4-5 minutes to run.

Note: Don’t close this SQL query. You will come back to it in task 7.

### Task 3: Load data in the lake house

1. Go to the **Data Engineering experience**, open the lake house and upload a local CSV file named **dimLocation**, located in the folder Labs\data

|  |  |
| --- | --- |
| A screenshot of a computer  Description automatically generated | A screenshot of a computer  Description automatically generated |

1. Once the file is uploaded, go to **Files**, right click on the file, and load the data into a new Delta table – where it will be made query-able in the lakehouse. Name the table as **dimLocation\_TaxiZone**.

A screenshot of a computer

Description automatically generated

1. Within the lake house, switch over to the **SQL analytics endpoint**:

A screenshot of a computer

Description automatically generated

1. The table which was just loaded can be queried with SQL– you can simply run **SELECT \* FROM table**. We will see later how this table can be queried from the data warehouse.

A screenshot of a computer

Description automatically generated

1. You should see a new query window with the T-SQL query and the result set:

A screenshot of a computer

Description automatically generated

### Task 4: Load data into a table using Data Flow gen 2

1. **Go to the Data Warehouse experience** and open the DW youcreated earlier, click on creating a **new Data Flow Gen2** and click on Text/CSV source.

A screenshot of a computer

Description automatically generated

1. Specify the URL <https://data.cityofnewyork.us/api/views/755u-8jsi/rows.csv> and click **Next**, then click **Create**.

A screenshot of a computer

Description automatically generated

1. Various data transformations are available which are similar to Power Query wizard experience.

***Note***: you will go into more details of how to use **Data Flow Gen2** in the next lab.

1. You can see quickly the capabilities of **Data Flow Gen2** by removing all the other columns except keeping only the 3 columns **Zone**, **LocationID** and **Borough**.

Select the column you want to remove, then click **Remove columns** > **Remove columns.**

A screenshot of a computer

Description automatically generated

1. You can hover over the i button and see that the file will be loaded to your DW.

A screenshot of a computer

Description automatically generated

1. Click on the **Settings** gear (right of i button), save the table as **DimLocation** and click Next.

A screenshot of a computer

Description automatically generated

1. Finally, you can **Publish** the transformations.
2. Wait for the publishing to be completed and once done, open the data warehouse, expand the table and observe that the table loaded from the Data Flow Gen2 gets created.

Note: You can open the **Monitoring Hub** to see the status of the data flow execution.

A screenshot of a computer

Description automatically generated

### Task 6: Visual Query

For students who are familiar with Power Query (Dataflow/PBI), this is a great feature that you would enjoy in warehouse experience.

The tables created can now be **dragged over** on a canvas to create a transformation which can be translated into a SQL query output too – this is Visual Query feature that allows developers to transform tables using GUI.

In this part of the lab, we will show how to perform a lookup to **join** the **DimLocation** table with **dimLocation\_TaxiZone**.

To recap, you have 2 tables:

* In the lake house: CSV file dimLocation is loaded and saved as Delta, **dimLocation\_TaxiZone**
* In the data warehouse: CSV file <https://data.cityofnewyork.us/api/views/755u-8jsi/rows.csv> is loaded and saved as table **DimLocation**

1. In the **Data warehouse experience**, navigate to your DW, click on the **Visual Query** tab at the top.

A screenshot of a computer

Description automatically generated

1. **Drag and drop** on the canvas the table which was just created - **DimLocation**.

A screenshot of a computer

Description automatically generated

1. Click on **+ Warehouses**, add the lake house (through SQL endpoint) which was created earlier.

|  |  |
| --- | --- |
| A screenshot of a computer  Description automatically generated | A screenshot of a data hub  Description automatically generated |

1. In the Explorer window you will see the lake house. **Drag and drop** the Delta table, **dimLocation\_TaxiZone,** which was created earlier

A screenshot of a computer

Description automatically generated

1. Under **Combine**, click **Merge queries**.

A screenshot of a computer

Description automatically generated

1. Select the join key columns **LocationID** between the two tables.

A screenshot of a computer

Description automatically generated

1. Click on **View SQL** button to show how the wizard translated the queries into SQL.

A screenshot of a computer

Description automatically generated

A computer screen shot

Description automatically generated

1. Click on **Visualize results**, to demonstrate Visualize Query feature that can create a visual out of the query which translates the transformation.

A screenshot of a computer

Description automatically generated

1. You can **configure the visual** as in the picture. This shows how quick it is to explore data as it is still being transformed.

A screenshot of a graph

Description automatically generated

1. Finally click **Save** in the top left corner which allows you to create Power BI dataset out of this transformed query.

### Task 7: Cross-workspace, cross database query

1. Go to the SQL query you created at task 2 (the one containing the queries from **Tables.sql**)
2. Run the code between the lines 39-45.

The block that demonstrates how a powerful insight can be enabled by allowing data query from different workspaces altogether – of course enabling cross-database query.

The lakehouse can also be part of another workspace altogether – this is a great use case for data mesh architecture.

This is the end of the lab. Congratulations for finishing the lab!