A close-up of a person smiling

Description automatically generated

# Lab 03: Implementing Data Pipelines with Fabric Data Factory

### 

### Introduction

In this lab, you will ingest parquet containing sales data directly to the Managed Lakehouse table, using a Data Pipeline with Copy activity, to append this dataset to the existing Sales table.

### Objectives

After completing this lab, you will be better able to:

1. Create a data pipeline
2. Execute and monitor the pipeline
3. Create a basic Semantic Model
4. Auto-create a Power BI report to visualize data

**Estimated time to complete this lab**

60 minutes

Contents

[Lab 03: Implementing Data Pipelines with Fabric Data Factory 1](#_Toc926584786)

[Introduction 1](#_Toc852597412)

[Objectives 1](#_Toc1983618386)

[Task 1: Create a Data pipeline 3](#_Toc887396040)

[Task 2: Execute and monitor the pipeline 8](#_Toc1213107869)

[Task 3: Create the Semantic Model 11](#_Toc1544668303)

[Task 4: Create the Report 15](#_Toc2105073071)

**Lab Prerequisites**

* Workspace: Fabric, Power Premium or Fabric trial
* Individual license: Power Pro or Premium Per User account

**Information provided by your training provider**

* Trial tenant (if applicable): login & password, workspace to use for the lab.
* Azure Data Lake Gen2 (containing data sources): account name & shared access signature.

### Task 1: Create a Data pipeline

In this task, you will create the Copy Activity that copies the parquet files from a private storage account to the Sales table from the Managed Lakehouse.

* On the MS Fabric menu, navigate to Data Factory and click on the Data pipeline. Provide the name Pipeline Lab 03

A white background with black text

Description automatically generated

* From the top menu, click on the Copy data and select Add to canvas:

A screenshot of a computer

Description automatically generated

* Click on the Copy data that you just added to the Canvas:

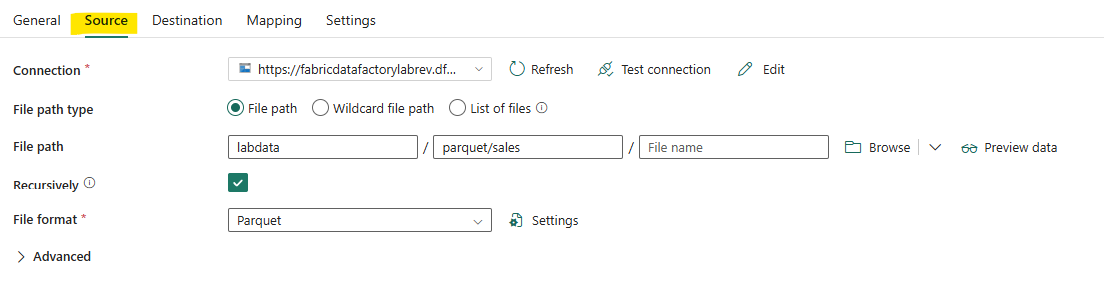
Go to the **General** Tab and under the Name type: Copy Parquet to Lakehouse. The other settings under this tab will remain unchanged.

A screenshot of a computer

Description automatically generated

* Click on **Source** Tab and select the following:
* Connection: Choose the name you provided in the first lab when you connected to the Private Storage account and test the connection to be successful
* File path type: File path
* File path: click on browse and navigate to the container labdata/parquet/sales
* **Recursively**: checked ( this will ensure that all the files from the subfolders labdata/parquet/sales will be read)
* File format: Parquet
* All the other settings from this tab will remain unchanged.

Make sure that your Source tab looks like this at the end:



* Click on the Destination tab:
* Connection: Reuse the connection for the Contoso Lakehouse
* Root folder: Tables
* Table name: Sales
* Expand Advanced
  + Table action: Append

At the end the Destination tab should look like this:

A screenshot of a computer

Description automatically generated

* On the **mapping** tab, click on Import Schemas to see the mapping

A screenshot of a computer

Description automatically generated

* The Settings tab remains unchanged:

A screenshot of a computer

Description automatically generated

* Once you finish building the Pipeline, from the top tab, click on the Validate to make sure that there are no errors and then finally on the Save button.

A screenshot of a computer

Description automatically generated

### Task 2: Execute and monitor the pipeline

In this task, you will start by trigger the pipeline

* From the top menu click on the Run button.

A screenshot of a computer

Description automatically generated

Then at the button of the page click on the Output and then on the Pipeline:

A screenshot of a computer

Description automatically generated

You can explore the copy data details to understand how much data you have transferred to the Lakehouse table.

A screenshot of a computer

Description automatically generated

As it can be observed, at the time we read the data from ADLS Gen 2 we had 2.689GB with 36 files, however, upon migration to the Lakehouse, the dataset transformed into a more compact form, occupying 183.042MB distributed across 20 files

This exercise illustrates the efficiency of transforming data from ADLS Gen 2 into a Delta table within the Lakehouse environment. By leveraging Delta tables, data compression is improved, and smaller files are consolidated into larger ones, optimizing data storage and access.

Come back to the SQL Analytics Endpoint of your Lakehouse, and run this basic SQL query to count the number of rows.

* You can realize that the Sales now contains more than 150M of rows.

A screenshot of a computer

Description automatically generated

* + You can also use the previously created SQL View to control sales per store.

A white rectangular object with a white border

Description automatically generated with medium confidence

### Task 3: Create the Semantic Model

In this task, you will create a simple Semantic Model from the Lakehouse, based on the Direct Lakehouse Mode.

* From the Lab Workspace, select the Contoso Lakehouse
* In the Lakehouse,viewer, click on the New semantic model button

A screenshot of a computer

Description automatically generated

* On the New semantic model UI, type SalesAnalysis as semantic model name, and select the following tables:
  + Sales
  + Calendar
  + Customers
  + Geography
  + Products
  + Stores
  + Promotions
  + Click on **Confirm** to initiate the Dataset creation

A screenshot of a computer

Description automatically generated

* + The objective is now to design the semantic model with the following relationships:

A screenshot of a computer

Description automatically generated

* + Click on the Manage relationships button to create validate the requiered relationships between tables.

A screenshot of a computer

Description automatically generated

* + You will now implement 2 very simple Measures (business calculations)
  + Create the measures Quantity using the following DAX definition:

. Quantity=SUM(Sales[SalesQuantity])

A screenshot of a computer

Description automatically generated

* + Create the measures NbBuyers using the following DAX definition:

NbBuyers=DISTINCTCOUNT(Sales[CustomerKey])

A screenshot of a computer

Description automatically generated

The semantic model is now ready to support report visualizations.

### Task 4: Create the Report

In this task, you will create a simple report to visualize data from the dataset using the Auto-create report function, then you will customize it.

* + From the Lab workspace, select the SalesAnalysis semantic model, click on the dotes and select the option Auto-create report

A screenshot of a computer

Description automatically generated

* + After few seconds, a basic report is created

A screenshot of a computer

Description automatically generated

* + Then you can personalize each visual by changing the Visualization type and the columns and measures displayed on the visual axis.

A screenshot of a computer

Description automatically generated

* + The screenshot hereafter displays the result using a donut chart with Sales quantity per Product Brands names.

A screenshot of a graph

Description automatically generated

* + Finally, save your report

A screenshot of a survey

Description automatically generated