Portfolio Project:

Creating and Using Pipelines in Microsoft Fabric

Project Overview

This project focused on ingesting data with a pipeline in Microsoft Fabric, which serves as a common analytical data store for cloud-scale analytics solutions. By utilizing Microsoft Fabric, it is possible to implement extract, transform, and load (ETL) or extract, load, and transform (ELT) solutions through the creation of pipelines.

Objectives:

1. **Create a Workspace:**
   - Set up a workspace with Fabric.
2. **Create a Lakehouse:**
   - Establish a data lakehouse for data ingestion.
3. **Create a Pipeline:**
   - Define a pipeline to copy data from an external source into the lakehouse.
4. **Create a Notebook:**
   - Implement a notebook to transform the ingested data and load it into a table.
5. **Modify the Pipeline:**
   - Incorporate the notebook into the pipeline for a reusable ETL process.

Experience

Create a Workspace

- Navigated to the Microsoft Fabric home page and signed in with credentials.
- Selected Workspaces from the menu bar and create a new workspace.

Create a Lakehouse

- In the workspace, selected Create and chose Lakehouse under the Data Engineering section, assigning it a unique name.
- Waited for the new lakehouse to be created.

Create a Pipeline

- On the Home page for the lakehouse, selected Get data and then New data pipeline, naming it Ingest Sales Data.
- Used the Copy Data wizard to set up the connection to the data source, entering the URL for the sales data and configured the necessary settings.
- Specified the data destination options and reviewed the copy summary before saving and running the pipeline.

New pipeline

×

Name

Ingest Sales Data

[Create] [Cancel]

Copy data into Lakehouse «

● **Choose data source**
Select a connector. Then enter the connection information.

○ Connect to data source

○ Choose data destination

○ Connect to data destination

○ Review + save

Connect to data source

×

Http
Other

Connection settings

Url *

https://raw.githubusercontent.com/MicrosoftLearning/dp-...

Connection credentials

Connection

Create new connection ⌄   ↻

Connection name

sales

Data gateway

(none) ⌄   ↻

Authentication kind

Anonymous ⌄

Privacy Level

None ⌄

☐ This connection can be used with on-premises data gateways and VNet data gateways.

[Back] [Next]

Copy data into Lakehouse «

✓ Choose data source

● **Connect to data source**
Select, preview, and choose the data.

○ Choose data destination

○ Connect to data destination

○ Review + save

Connect to data source

×

Connection
sales User1-53337658

Base URL
https://raw.githubusercontent.com/Microsof...

Relative URL ⓘ

Request method ⓘ

GET ⌄

Additional headers ⓘ

☐ Binary copy ⓘ

Request timeout ⓘ

Max concurrent connections ⓘ

[Back] [Next]

Copy data into Lakehouse

«

- ✓ Choose data source
- ● **Connect to data source**
  Select, preview, and choose the data.
- ○ Choose data destination
- ○ Connect to data destination
- ○ Review + save

Connect to data source   ✕

File format

| DelimitedText ⌄ | [Detect text format]   [🔗 Preview data]

Column delimiter

| Comma (,) ⌄ |

Row delimiter

| Line feed (\n) ⌄ |

☑ First row as header ⓘ

Compression type

| No compression ⌄ |

[Back] [Next]

Copy data into Lakehouse

«

- ✓ Choose data source
- ✓ Connect to data source
- ● **Connect to data destination**
  Select and map to folder path or table.
- ○ Review + save

Connect to data destination   ✕

Connection	thato_ls
Root folder	○ Tables ● Files

Folder path

If the identity you use to access the data store only has permission to subdirectory instead of the entire account, specify the path to browse.

| new_data | 📁 Browse

File name

| sales.csv |

Copy behavior ⓘ

| Select... ⌄ |

[Back] [Next]

Create a Notebook

- Selected New notebook from the Open notebook menu on the Home page for the lakehouse.
- Replaced the default code in the notebook with a variable declaration and added code to read the sales data, transform it, and save it as a table.
- Ran all cells in the notebook to execute the transformations.
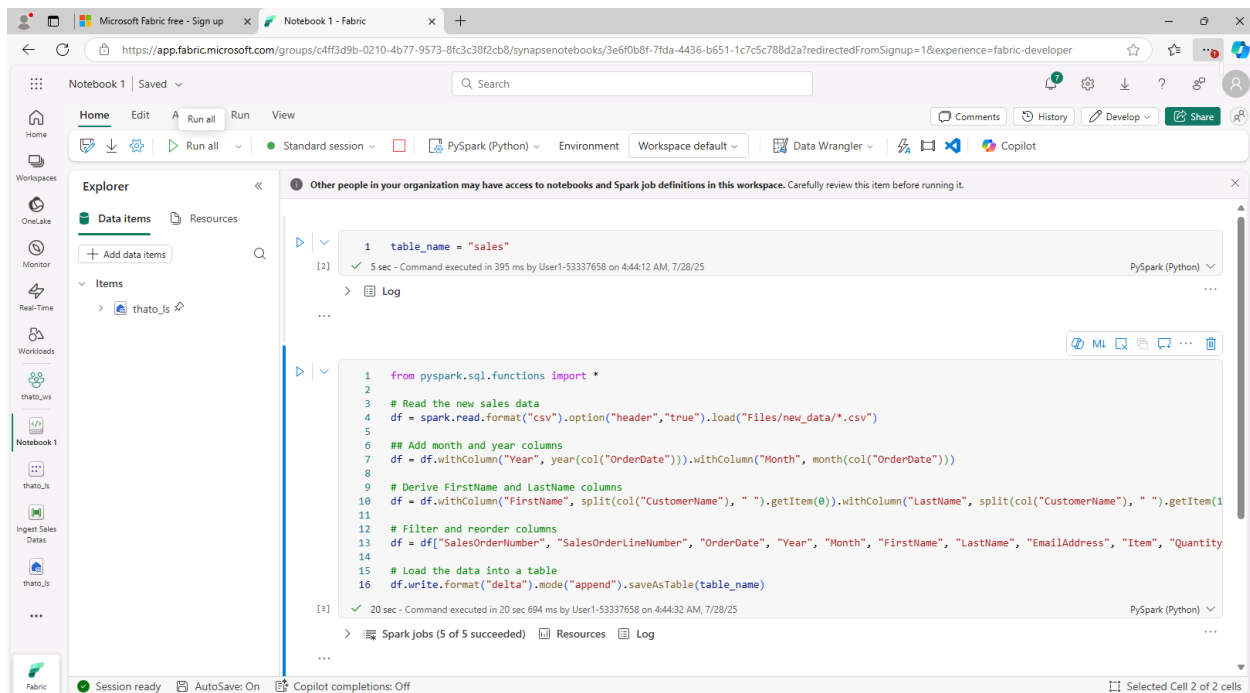
Modify the Pipeline

- Selected the Ingest Sales Data pipeline and add a Delete data activity to remove old files before copying new data.
- Incorporated the Load Sales notebook into the pipeline and set the necessary parameters.
- Saved and ran the modified pipeline, ensuring all activities complete successfully.
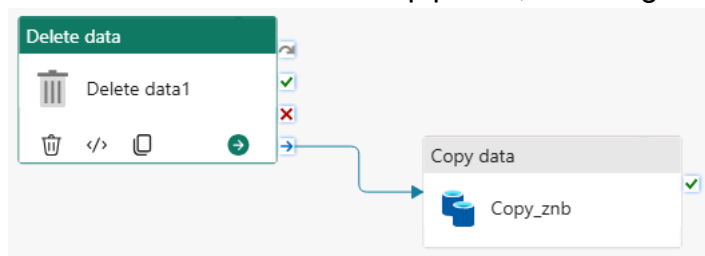
General   **Source**   Logging settings [1]

Name *   Delete old files   Learn more ⬈

Description

Activity state ⓘ   ⦿ Activated   ◯ Deactivated

Timeout ⓘ   0.12:00:00

Retry ⓘ   0

› Advanced

General   **Source**   Logging settings [1]

Connection *   🔷 thato_ls   ⌄   ↻ Refresh   ⬈ Open

File path type   ◯ File path   ⦿ Wildcard file path   ◯ List of files ⓘ

Folder path   Files /   new_data   📁 Browse   ⌄

Wildcard file name   *.csv

Recursively ⓘ   ✅

› Advanced

General   Source   **Logging settings**

Enable logging ⓘ   ☐
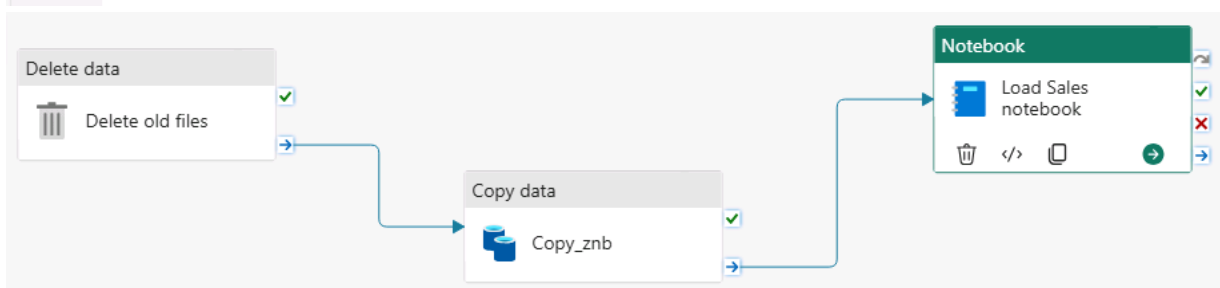
General   Settings

Name *   Load Sales notebook   Learn more ⬈

Description

Activity state ⓘ   ⦿ Activated   ◯ Deactivated

Timeout ⓘ   0.12:00:00

Retry ⓘ   0

› Advanced

Results

- ✓ I successfully created a Microsoft Fabric workspace and lakehouse.
- ✓ I defined a pipeline to ingest data from an external source and created a notebook to transform the data and load it into a table.
- ✓ I modified the pipeline to include the notebook, creating a reusable ETL process.
- ✓ The pipeline ran successfully, and I verified that the new_sales table was created in the lakehouse.

Conclusion

This project provided a comprehensive introduction to creating and using pipelines in Microsoft Fabric, covering workspace setup, lakehouse creation, data ingestion, and transformation processes. I gained valuable insights into the capabilities of Microsoft Fabric for managing data workflows and implementing ETL solutions.

Resources

Source file:

https://raw.githubusercontent.com/MicrosoftLearning/dp-data/main/sales.csv

GitHub: https://github.com/ThatoMTNG/Microsoft-Fabric-Analytics-Engineer-DP-600-

Mentions

Project Author: Thato Metsing (https://www.linkedin.com/in/thatometsing/)

Project Mentor: Maureen Direro (https://www.linkedin.com/in/maureen-direro-46a6b220/)