Portfolio Project:

Creating and Using Dataflows (Gen2) in Microsoft Fabric

Project Overview

This project focused on creating and utilizing Dataflows (Gen2) in Microsoft Fabric, which serve as a powerful tool for connecting to various data sources and performing transformations in Power Query Online. The objective was to effectively utilize Dataflows within Data Pipelines, facilitating the ingestion of data into a lakehouse or defining datasets for Power BI reports.

Objectives:

1. **Create a Workspace:**
   - Set up a workspace.
2. **Create a Lakehouse:**
   - Establish a data lakehouse for data ingestion.
3. **Create a Dataflow (Gen2):**
   - Define a dataflow to encapsulate the ETL process for data ingestion.
4. **Add Data Destination for Dataflow:**
   - Configure the data destination to the lakehouse.
5. **Add Dataflow to a Pipeline:**
   - Include the dataflow as an activity in a data pipeline.

Experience

Create a Workspace

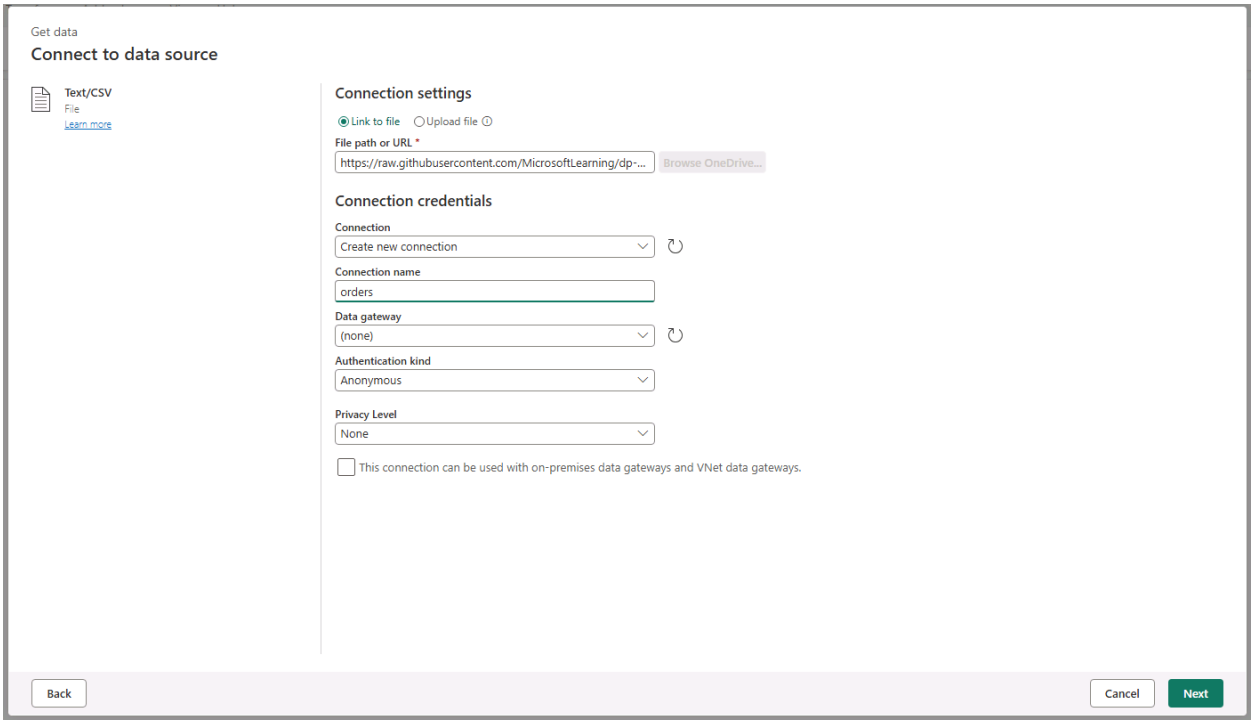
- Navigated to the Microsoft Fabric home page and signed in with credentials.
- Selected Workspaces from the menu bar and created a new workspace.

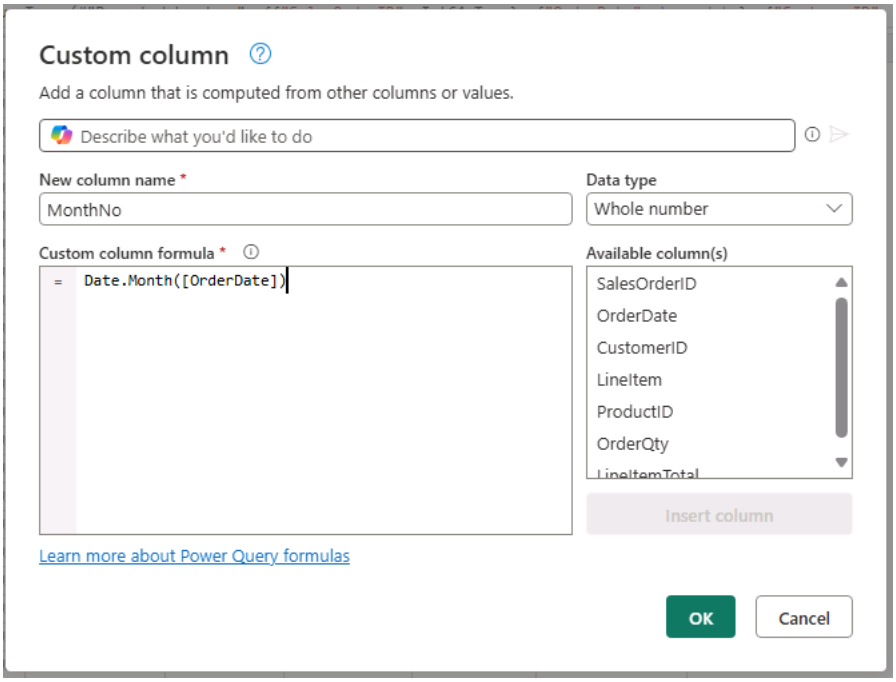
Create a Lakehouse

- In the workspace, selected Create and chose Lakehouse under the Data Engineering section, assigning it a unique name.

Create a Dataflow (Gen2)

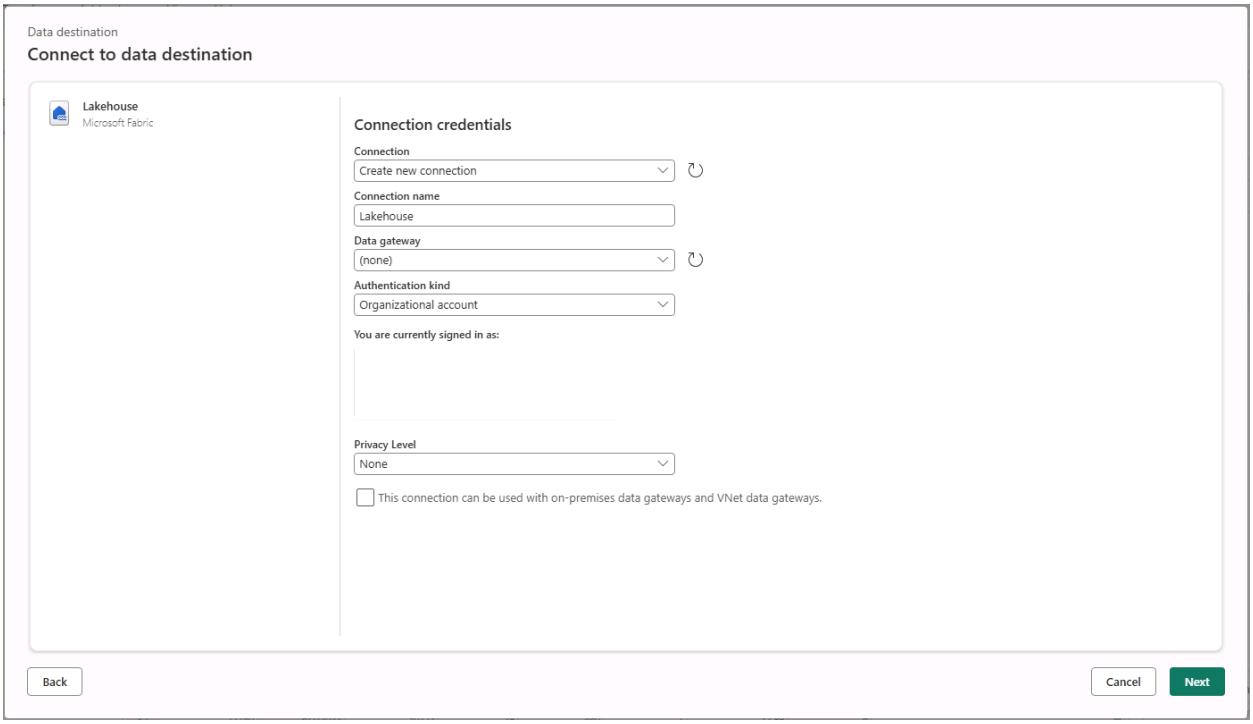
- In the lakehouse, selected Get data > New Dataflow Gen2.
- Chose Import from a Text/CSV file and set the file path.
- Created a custom column named MonthNo using the formula Date.Month([OrderDate]).

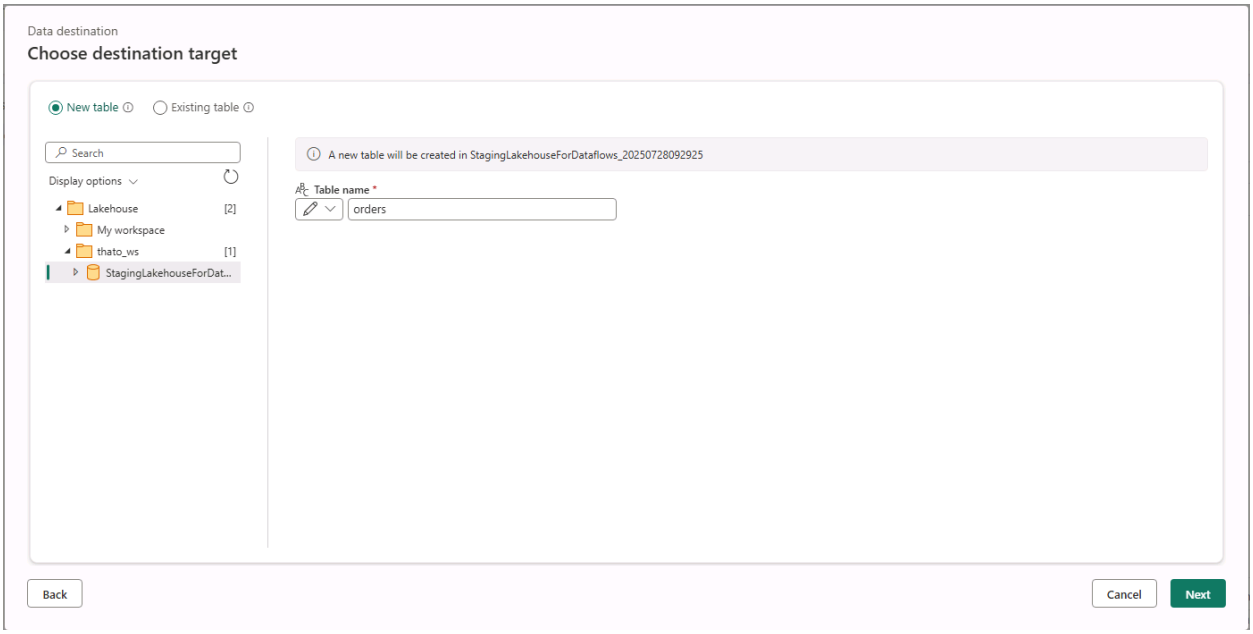

Get data
Connect to data source
Text/CSV
File
Learn more
Connection settings
Link to file
Upload file
File path or URL *
https://raw.githubusercontent.com/MicrosoftLearning/dp-...
Browse OneDrive...
Connection credentials
Connection
Create new connection
Connection name
orders
Data gateway
(none)
Authentication kind
Anonymous
Privacy Level
None
This connection can be used with on-premises data gateways and VNet data gateways.
Back
Cancel
Next

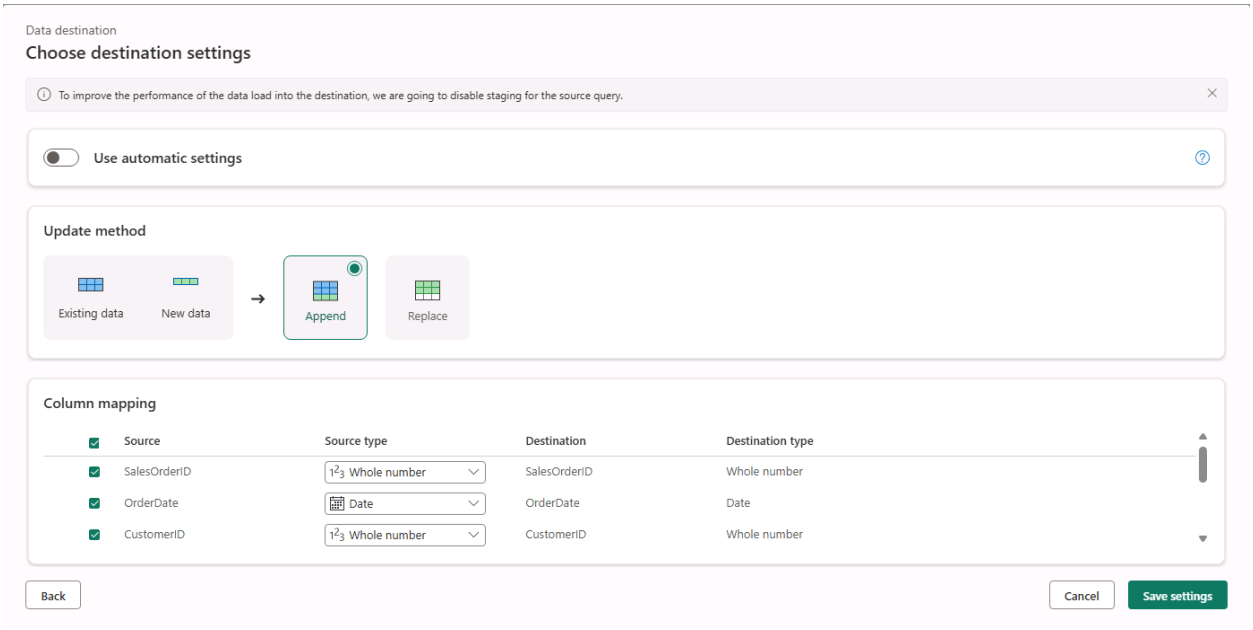

Custom column
Add a column that is computed from other columns or values.
Describe what you'd like to do
New column name *
MonthNo
Data type
Whole number
Custom column formula *
= Date.Month([OrderDate])
Available column(s)
SalesOrderID
OrderDate
CustomerID
LineItem
ProductID
OrderQty
Insert column
Learn more about Power Query formulas
OK
Cancel

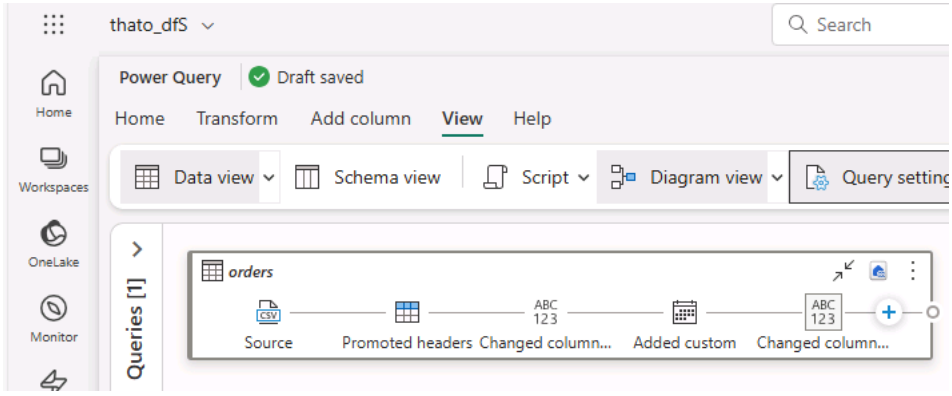
Add Data Destination for Dataflow

- Selected Lakehouse as the data destination and signed in with a Power BI organizational account.
- Specified the lakehouse and created a new table named orders, then saved the settings.









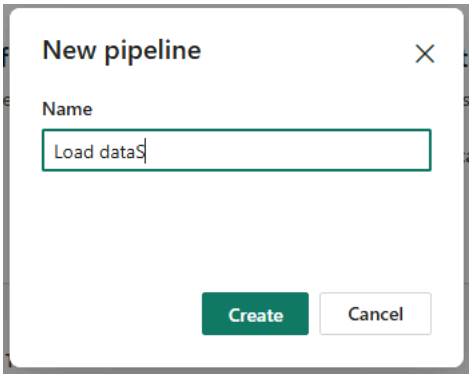

```
Table.TransformColumnTypes(#"Added custom", {{"OrderDate", type date}, {"MonthNo", Int64.Type}})
```

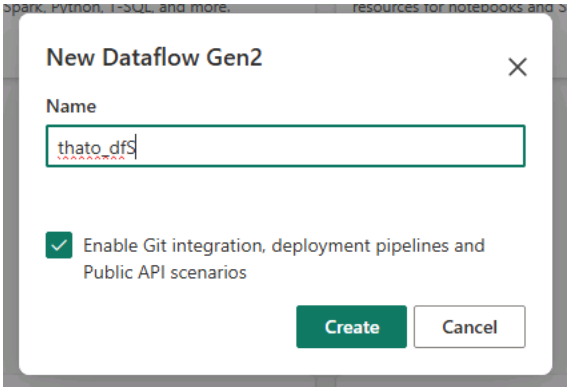
	SalesOrderID	OrderDate	CustomerID	LineItem	ProductID	OrderQty	LineItemTotal	MonthNo
1	71774	6/1/2022	29847	1	836	1	356.9	6
2	71774	6/1/2022	29847	2	822	1	356.9	6
3	71776	6/1/2022	30072	1	907	1	63.9	6
4	71780	6/1/2022	30113	1	905	4	873.82	6
5	71780	6/1/2022	30113	2	983	2	923.39	6
6	71780	6/1/2022	30113	3	988	6	406.79	6
7	71780	6/1/2022	30113	4	748	2	1637.4	6
8	71780	6/1/2022	30113	5	990	1	323.99	6
9	71780	6/1/2022	30113	6	926	1	149.87	6
10	71780	6/1/2022	30113	7	743	1	809.76	6
11	71780	6/1/2022	30113	8	782	4	5507.98	6
12	71780	6/1/2022	30113	9	918	2	316.86	6

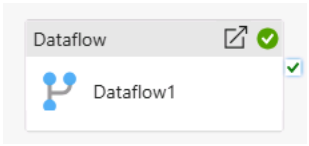
86 s) Columns: 8 Rows: 99+ Add default destination...

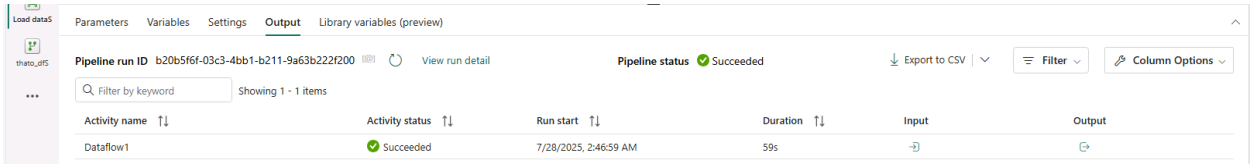
Add Dataflow to a Pipeline

- Created a new data pipeline named Load data and added a Dataflow activity.
- Selected the previously created and saved the pipeline.
- Ran the pipeline and waited for it to complete.









Results

- ✓ A Microsoft Fabric workspace and lakehouse were successfully created.
- ✓ A Dataflow (Gen2) was defined to ingest data from a CSV file, including the addition of a custom column.
- ✓ The data destination was configured to the lakehouse, resulting in the creation of a new table.
- ✓ The dataflow was included in a pipeline and executed, leading to the successful ingestion of data into the lakehouse.

Conclusion

This project provided a practical introduction to creating and using Dataflows (Gen2) in Microsoft Fabric, encompassing workspace setup, lakehouse creation, and data ingestion processes. Valuable insights were gained into the capabilities of Microsoft Fabric in managing data workflows and integrating data transformation within a modern data architecture.

Resources

Source file:

https://raw.githubusercontent.com/MicrosoftLearning/dp-data/main/orders.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/)