

SQL database in Microsoft Fabric

Series Scenario – Conference session application

You need:

- to store and serve data for the conference webpage.
- to be able to run analysis over the data.
- an application that allows attendees to easily search for and find sessions of interest to them.
- to follow modern best practices for application lifecycle.
- to be able to monitor and troubleshoot database and query performance.

Course Overview and agenda



Episode 1: Introduction and Overview; Getting started



Episode 2: Dataflows, Notebooks, Reports



Episode 3: GenAI and vector databases



Episode 4: GraphQL and application development



Episode 5: Application lifecycle management



Episode 6: Performance Dashboard, Recap

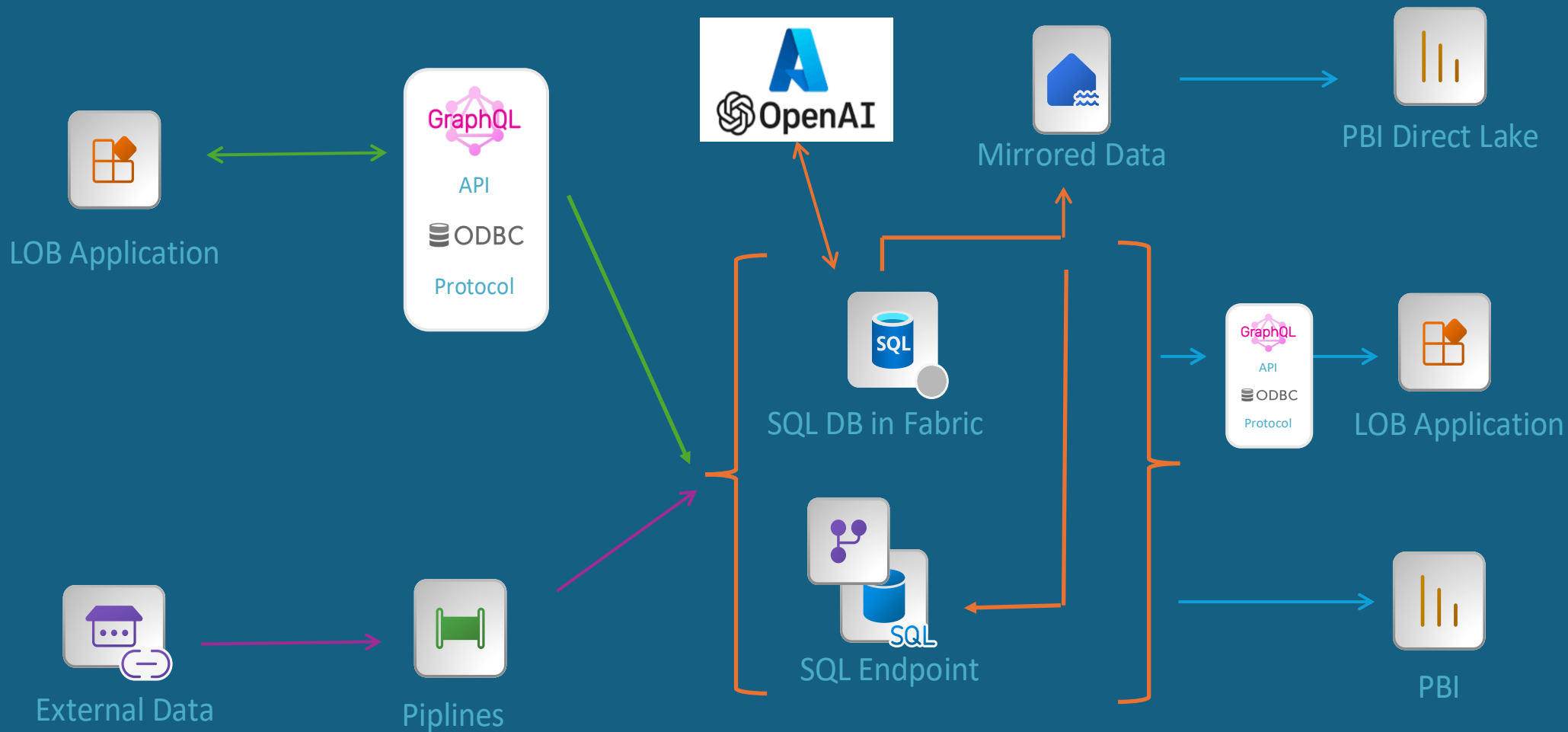
DATA SOURCE

INGESTION

AI

STORE

EXPOSE



Episode 2: Ingest and work with data, build reports

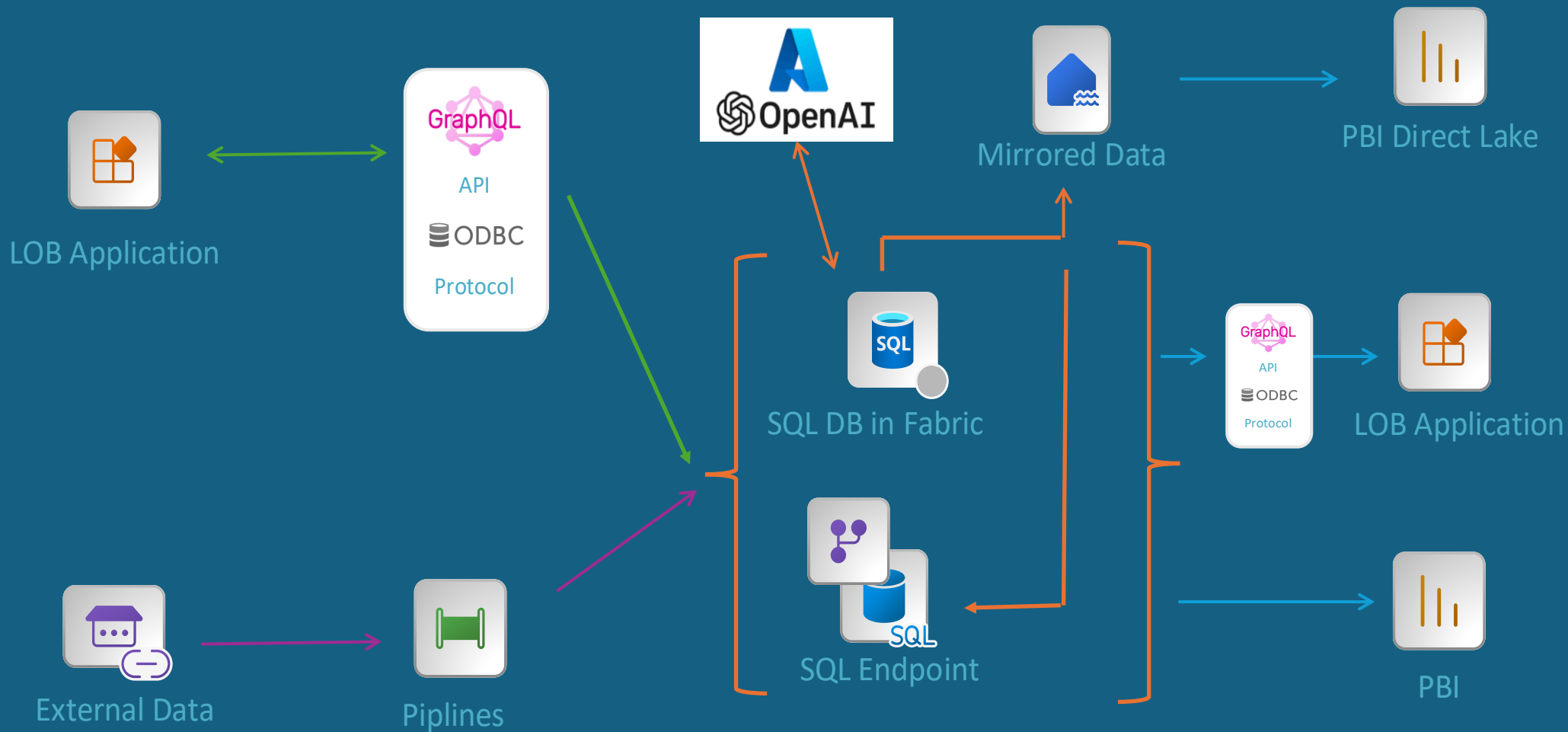
DATA SOURCE

INGESTION

AI

STORE

EXPOSE



Data Pipelines

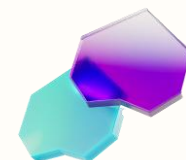
Data Pipelines enable powerful workflow capabilities at cloud-scale. Visually build complex workflows, move PB-size data, and defining sophisticated control flow pipelines.

Automate ETL processes and build workflows that can invoke any Fabric item tasks at scale.

Additionally, control flow capabilities are built into pipelines so you can visually design workflows that include loops and conditional logic

The screenshot displays the Microsoft Fabric Data Pipeline interface. The top navigation bar includes 'Home', 'Activities', 'Run', and 'View'. The main canvas shows a workflow for 'Contoso Daily Sales' with the following steps: 'Look up' (LookupTableList), 'For Each' (Iterate folder), 'Dataflow' (Dedup data and validate schema), 'Switch' (Validate schema), and 'Teams' (Send teams notification). The 'For Each' loop contains 'Past Due Orders' and 'Copy to datalake'. The 'Dataflow' step contains 'Load Synapse datamart' and 'Copy1'. The 'Switch' step contains 'Log error rows'. The 'Teams' step contains 'Send teams notification'. The right sidebar shows the 'Trigger Pipeline' and 'Monitor' sections. The 'Monitor' section shows the source as 'AzureEvents-eventstream' and the trigger as 'BlobFileCreated'. The 'Save' section shows 'My Workspace / My Reflex' as the save location. The bottom section shows the 'Output' tab with a table of pipeline run details.

Name	Activity status	Start time	Duration	Input	Output
LookupTableList	Succeeded	4:30 AM, 7/19/23	30 sec	-	-
Iterate folder	Succeeded	4:30 AM, 7/19/23	30 sec	-	-
Past Due Orders	Succeeded	4:30 AM, 7/19/23	30 sec	-	-
Copy to datalake	Succeeded	4:30 AM, 7/19/23	30 sec	-	-
Dedup data and validate schema	Succeeded	4:30 AM, 7/19/23	30 sec	-	-
Validate schema	Succeeded	4:30 AM, 7/19/23	30 sec	-	-
Load Synapse datamart	Succeeded	4:30 AM, 7/19/23	30 sec	-	-

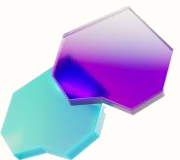
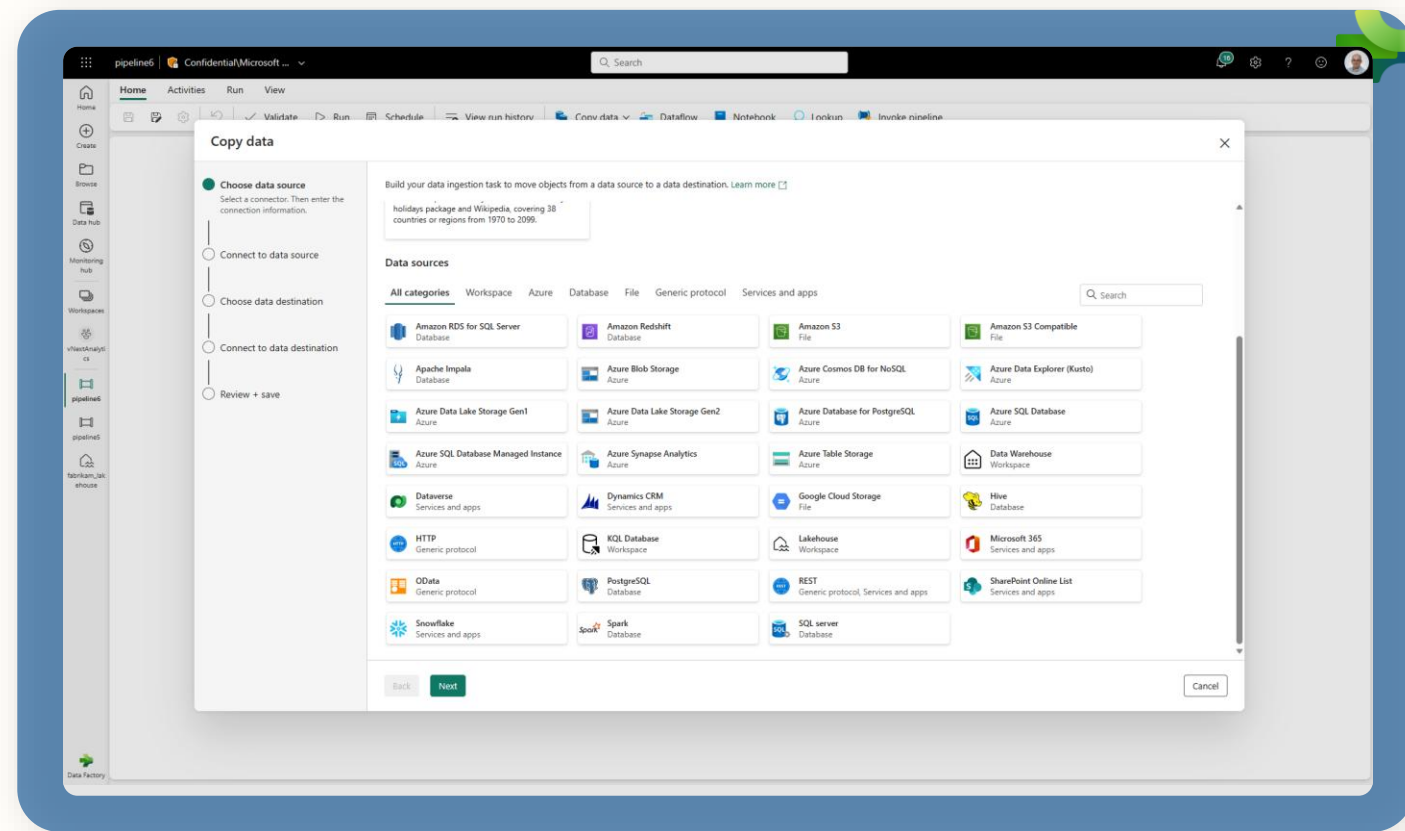


Data Pipelines | Connectors

Connectors provide a low-code interface for ingesting data from a variety of data sources

Connectors:

- Warehouse Connector; connect to existing Azure
- Lakehouse connector
- 100+ connectors in the copy activity
- Access to on-premises data
- Access protected data inside of a VNET



What pipelines look like

The screenshot displays the Microsoft Power BI Data Engineering Pipelines interface. The top navigation bar includes the title "Ingest Sales Data - Data engine...", a search bar, and user profile information. The main workspace shows a pipeline with three steps: "Delete data" (Delete old files), "Copy data" (Copy_o6d), and "Notebook" (Load Sales notebook). Each step has a green checkmark indicating success. Below the pipeline, the "Output" tab is selected, showing a table of pipeline run results.

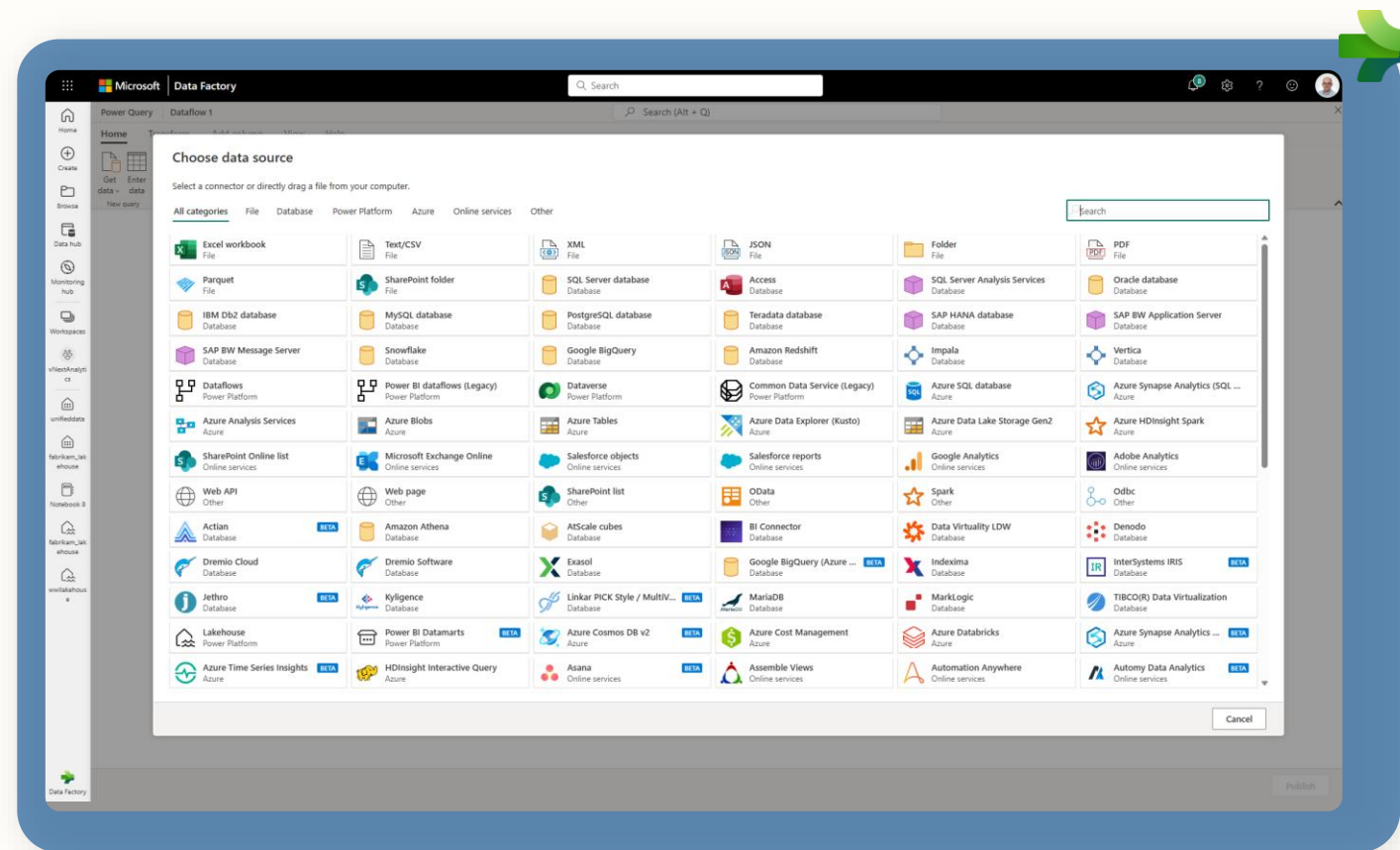
Pipeline run ID: **f8d1e987-8f79-4151-9903-186e36c14628**

Name	Type	Run start	Duration	Status	Run ID
Load Sales notebook	Notebook	2/17/2023, 2:54:37 PM	00:00:56	✓ Succeeded	d55a802a-e027-41a8-bf25
Copy_o6d	Copy data	2/17/2023, 2:54:24 PM	00:00:12	✓ Succeeded	bbf7b717-7f05-4f73-9629
Delete old files	Delete data	2/17/2023, 2:54:18 PM	00:00:05	✓ Succeeded	de025dd3-98c9-4a16-80c1

Dataflow Gen2

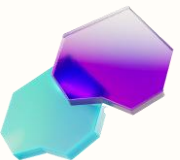
Dataflow provides a low-code interface (Power Query) for ingesting data from hundreds of data sources

Dataflows quickly and easily transform and unify data across multiple sources, establish a more collaborative analytics approach, and promote more informed, agile decision making.



Key Capabilities:

- Accelerate data transformation with code-free data flows
- Scale out using Fabric compute and Data Factory fast copy
- Load results of data transformations into multiple destinations (Azure SQL Databases, Lakehouse, etc.)



Dataflow Gen2 | Output Options

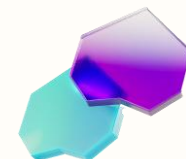
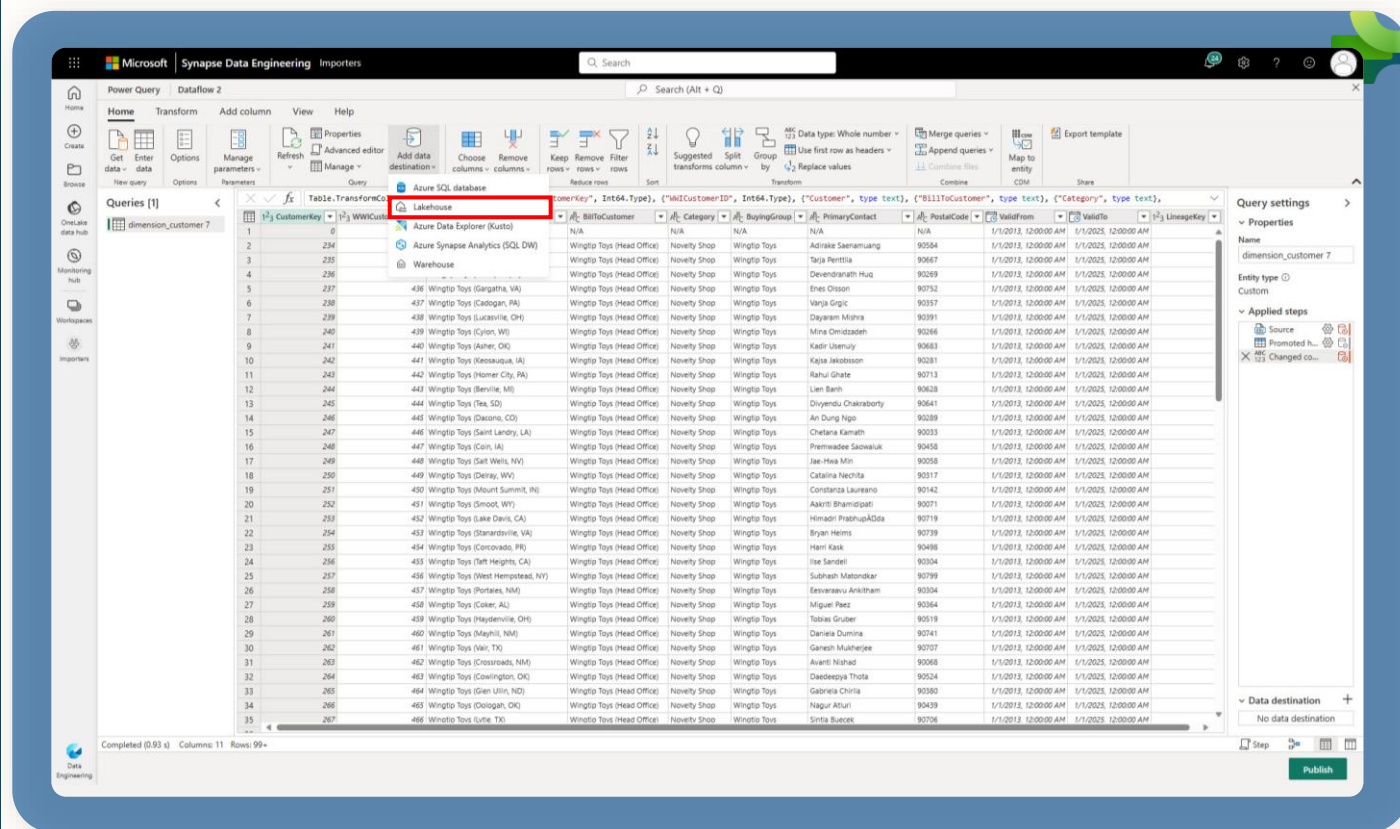
SQL database in Fabric:

Simply write into a Fabric SQL DB from a Dataflow

Users select the SQL Database output destination from the list and configure the connection.

Others:

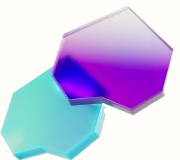
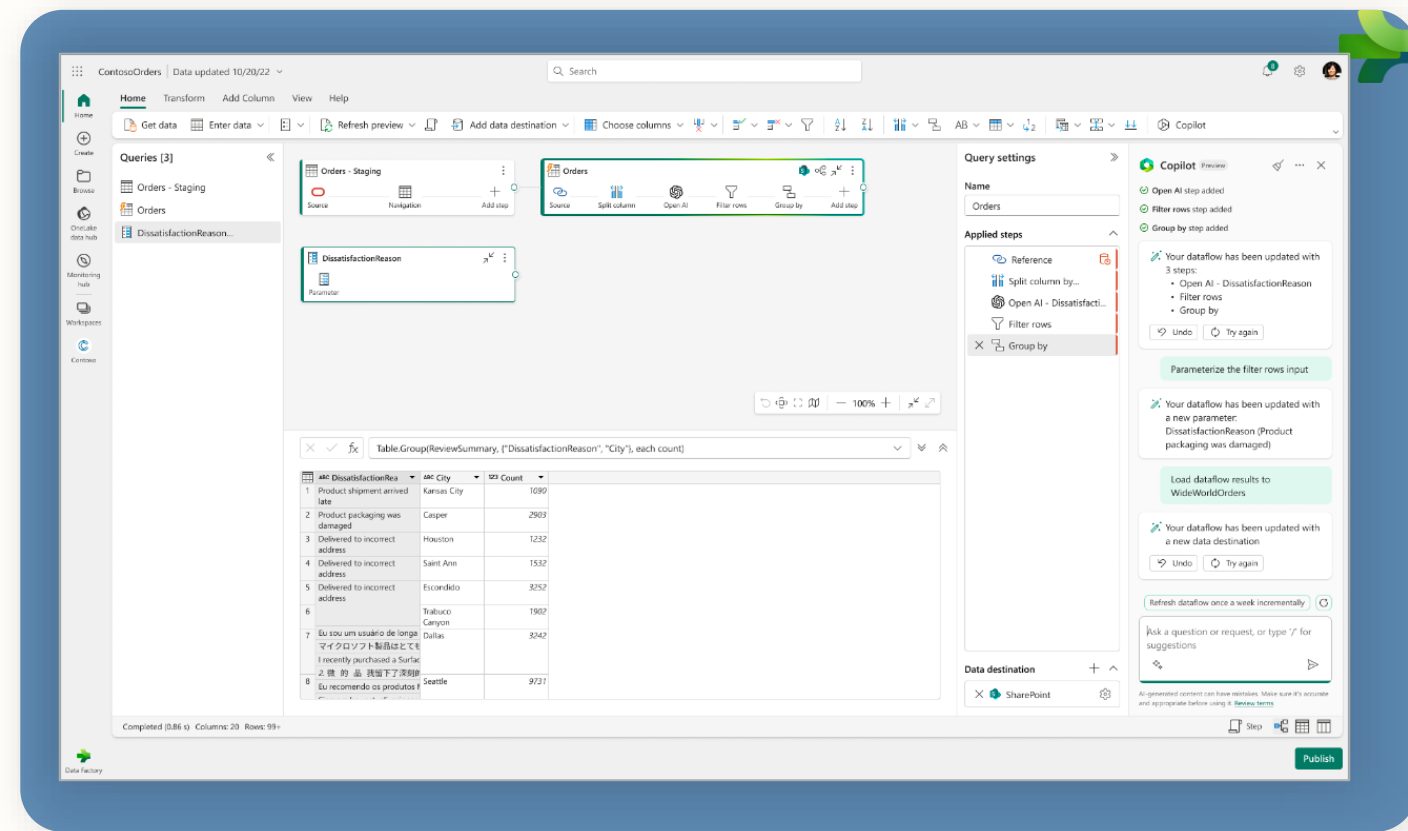
- Lakehouse
- Warehouse
- Azure SQL Database
- Kusto
- SQL DW



Data Factory Copilot

Allow AI Copilot to help design and build you Data Factory in Microsoft Fabric!

Generative AI Copilot for Data Factory enables natural language interfaces for constructing your Data Factory items



Loading Data with Dataflow Gen 2



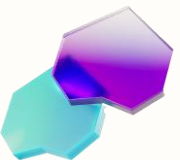
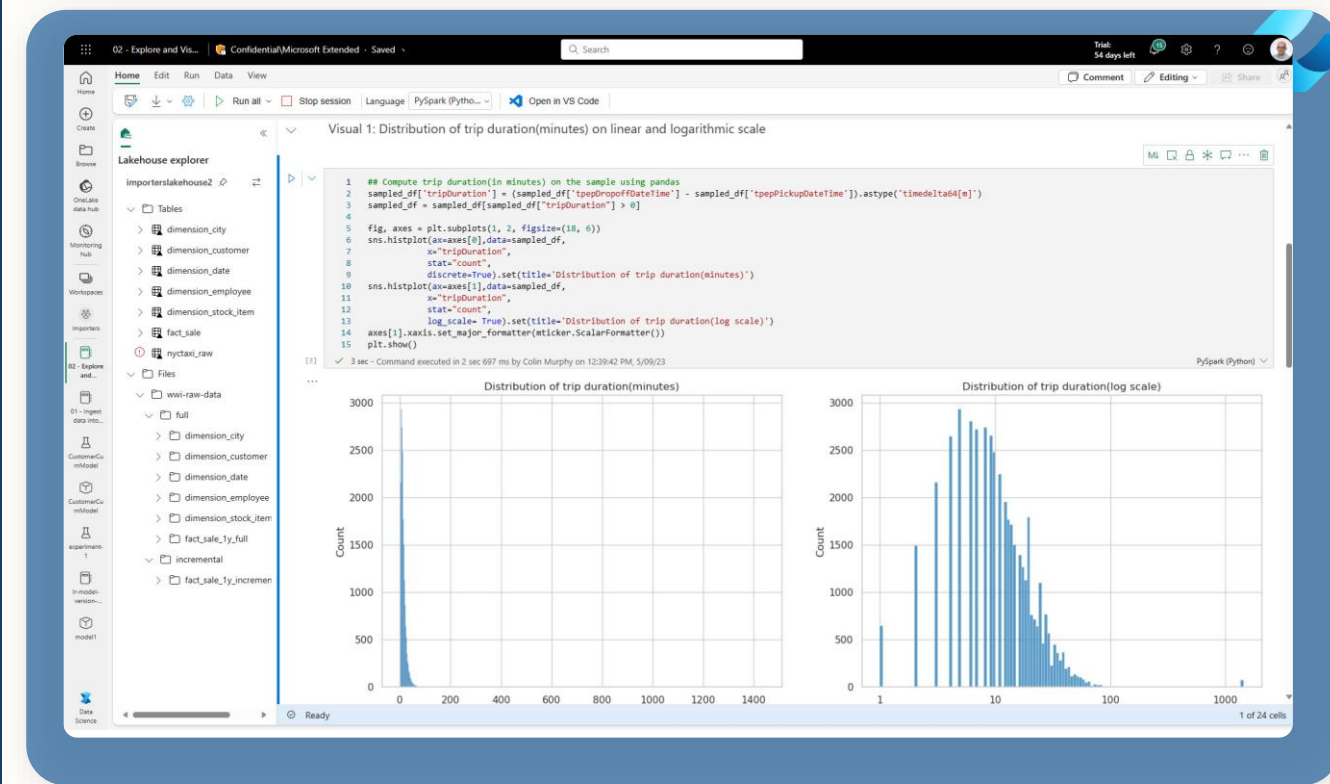
Notebooks

Immersive authoring experience for data developers

Rich notebook capabilities including native Lakehouse integration, real-time collaboration with commenting support, auto-save support, lightweight scheduling and pipeline integration

Key Capabilities:

- Manage your Python and R libraries through in-line installs using commands like %pip install
- Advanced notebook development support with ability to reference notebooks in notebooks, and snapshots for easy troubleshooting
- In context monitoring complete with real time advice and error analysis
- Streamline data prep without giving up the power of reproducibility of Python



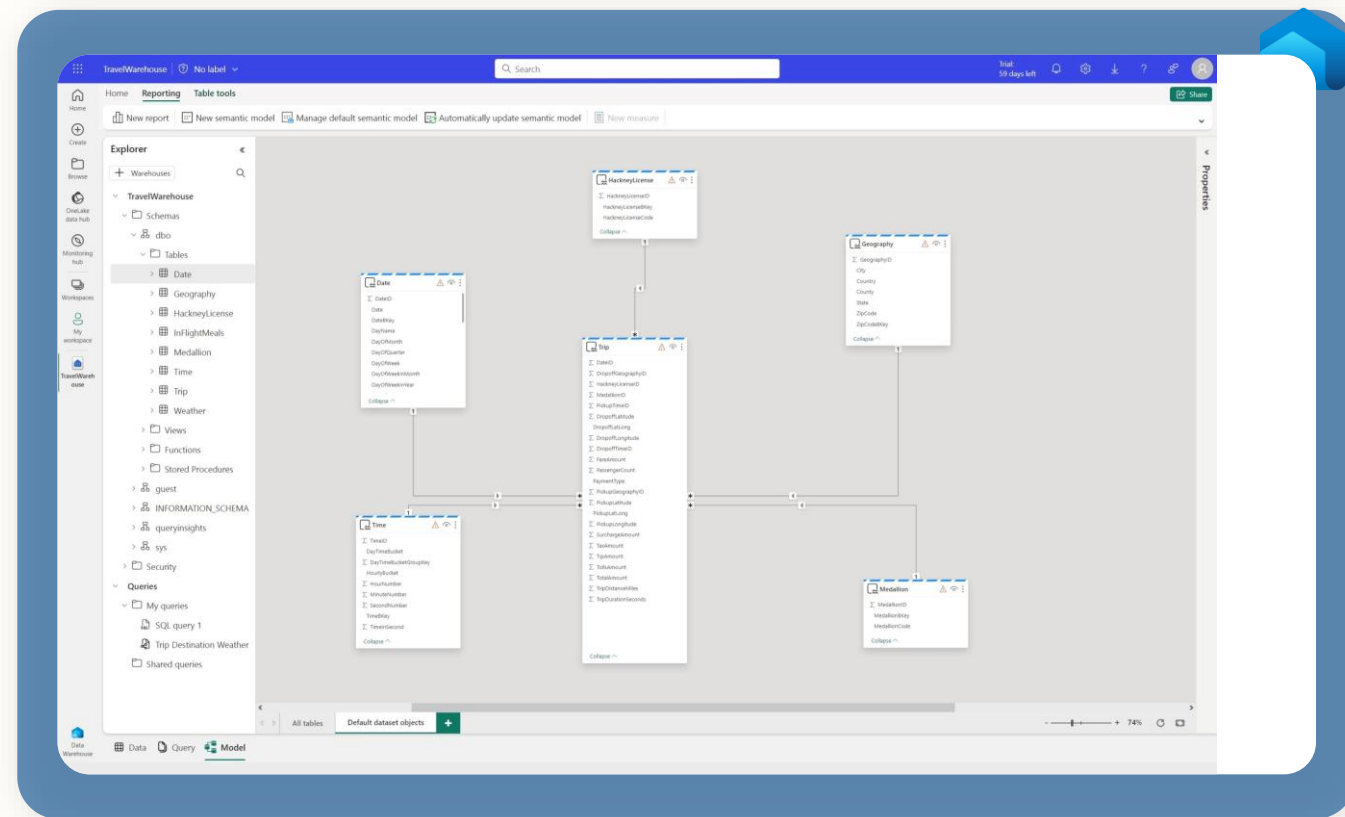
Create a
Notebook



Fully integrated Power BI semantic layer

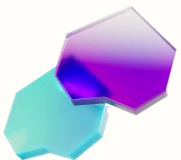
Reduce integration and gain insight
from your data in seconds

Built-in Power BI enables everyone
to visualize their data in seconds.



Key Capabilities:

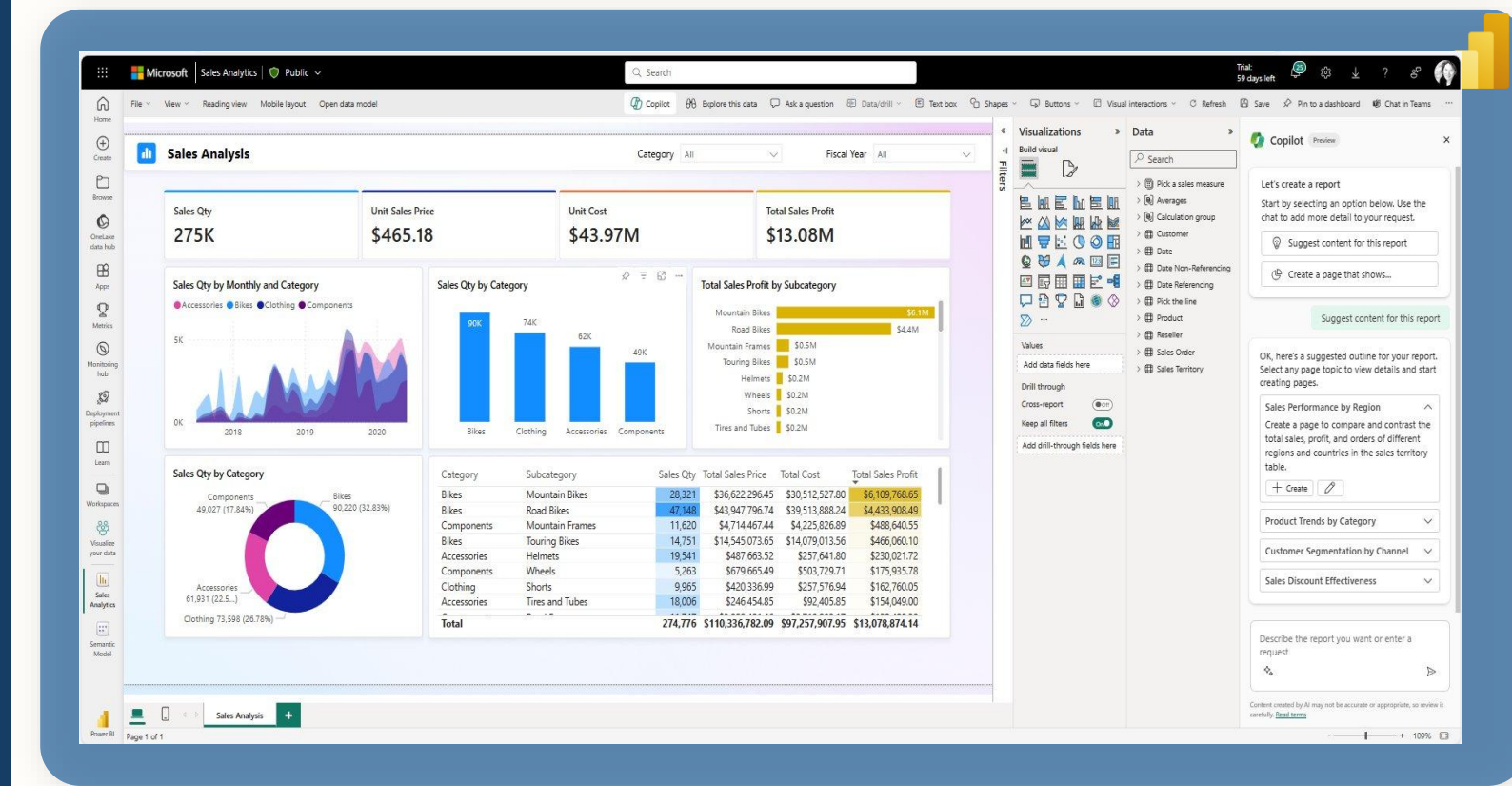
- Auto-generated semantic models always in sync
- Default dataset in Direct Lake mode but automatically switches to Direct Query or Import mode as security or performance needs change
- Flexibility to add/remove tables to dataset
- Full web authoring experience for creating measures



Power BI

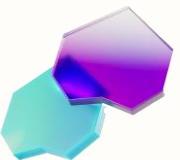
Auto create Power BI reports
from your semantic model

Blazing fast performance with
Direct Lake



Key capabilities:

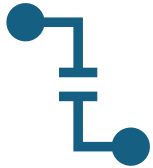
- Create an interactive report to discover and share business insights
- Use Copilot to help create, understand, and summarize reports
- Share interactively with Teams and PowerPoint
- View on phone or tablet with mobile-ready layouts for every report
- Explore data and find quick insights



Build and share a
report



Review



Imported data using dataflows

Simple and straightforward data import

...but it didn't have to be simple!



Created a notebook

Connected to analytics endpoint

Data was automatically and seamlessly mirrored to analytics endpoint

Saved a few views



Created a report over our data

Semantic model was fully integrated

Views we saved from our notebook were automatically visible and usable for reports.

End Episode 2

Ingest Data – See it in action with Dataflows

Use Notebooks to query analytics endpoint

Create a report over data