



Getting Data into OneLake



Martin Catherall

Data Analytics Consultant

 @MartyCatherall

www.MartinCatherall.com

- Data Architect

- SQL Server
 - Since 2002
- Microsoft Azure
 - Since 2015
- Microsoft Fabric
 - Since 2022



Agenda and content

- Brief Intro
 - Data in One-Lake
- Demos - A number of ways to get data into Fabric
 - Slide supplied (and in case the internet lets me down ☺)
- Summary

Slides



<https://tinyurl.com/3r6jjbkh>

Data Integration or Data Engineering?

Data Integration

Getting data **into** Fabric from various sources (the "Extract" and initial "Load")

- Batch Ingestion

- Real-Time Ingestion

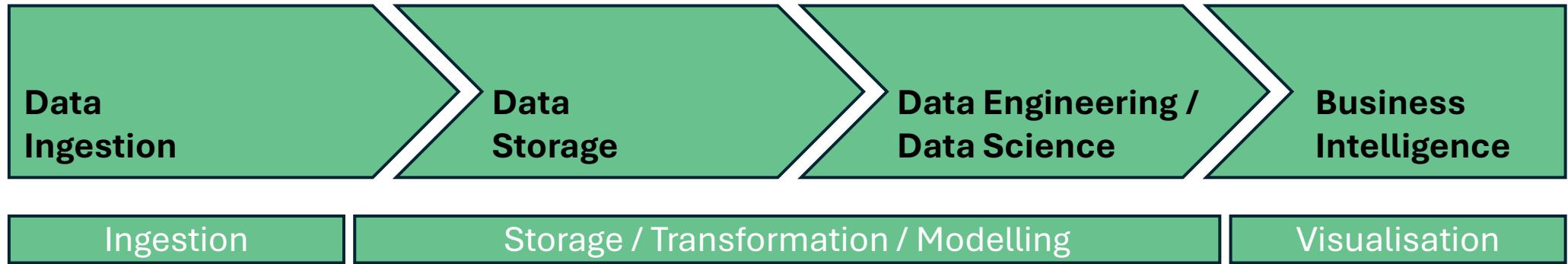
- Hybrid Ingestion

Data Engineering

What you **do with** the data once it's there (transformation, optimization, pipeline orchestration, data modeling)

The Data Journey.

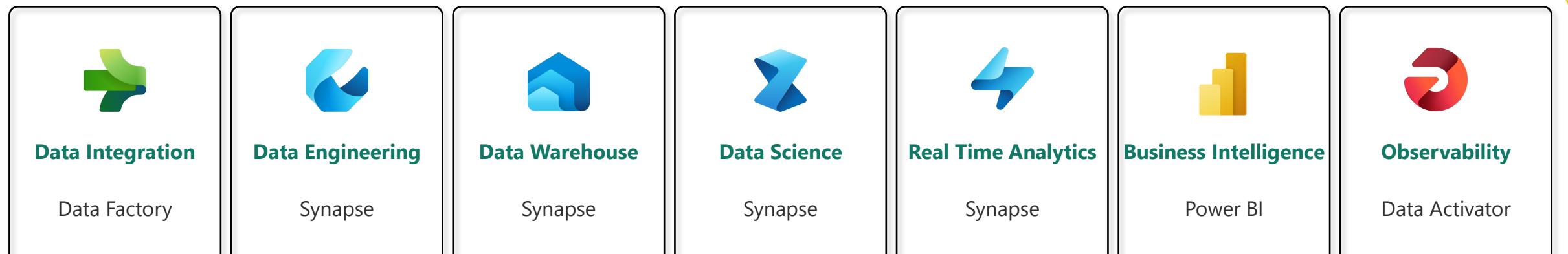
From Source to Visualisation.



- Excel (Power Pivot)
- Power BI
- Azure (Synapse)
- Microsoft Fabric

Microsoft Fabric does it all—in a unified solution

An end-to-end analytics platform that brings together all the data and analytics tools that organizations need to go from the data lake to the business user



UNIFIED

SaaS product experience

Security and governance

Compute and storage

Business model

One Lake - Acquiring Data

- | | | | |
|---|--------------------|----|---------------|
| 1 | Manually | 9 | SQL on Fabric |
| 2 | Data Integration | 10 | Warehouse |
| | - Notebooks | | COPY |
| 3 | Data Engineering | | OPENROWSET |
| | - (Data) Pipelines | | T-SQL (CTAS) |
| | - Dataflow Gen 2 | 11 | Mount ADF |
| 4 | Mirroring | | |
| 5 | Shortcuts | | |
| 6 | Semantic Model | | |
| 7 | Programmatically | | |
| 8 | Real Time | | |

One Lake - Acquiring Data

We'll need a place to store the data

Lakehouse

Warehouse

EventHouse

SQL

So, where do we put it? What do we need?

One Lake - Acquiring Data

Workspace, capacity and somewhere to put the data

Workspace settings

License info

Choose a license for this workspace.

Fabric capacity

License capacity
acapacity
SKU: F2, Region: Australia East

Semantic model storage format
Small semantic model storage format

Connection link

Use this link to connect third-party software to the workspace. Copy the link and add it to your third-party software.

powerbi://api.powerbi.com/v1.0/myorg/DataPlat01

General

Azure connections

System storage

Git integration

OneLake

Workspace identity

Network security

Monitoring

Power BI

Delegated Settings

Data Engineering/Science

Data Factory

Home > **acapacity** ...

Search ...

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Resource visualizer

> Scale

Pause Refresh Move Delete

Resource group (move) : Capacities

Status : Active

Location : Australia East

Subscription (move) : Microsoft Azure Sponsorship (MVP)

Subscription ID : 2635e25c-2d0f-4b09-ad52-1a62452a94e9

Tags (edit) : Add tags

Resource name **acapacity**

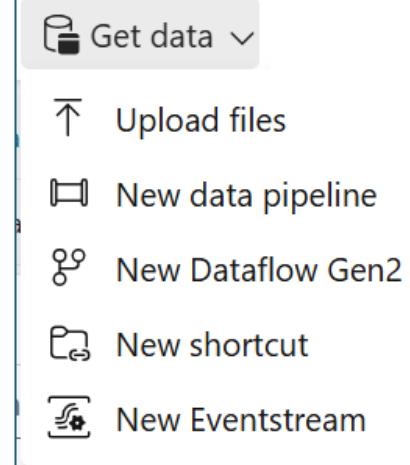
SKU **F2**

Demos

One Lake - Acquiring Data

1 Manually

Get data



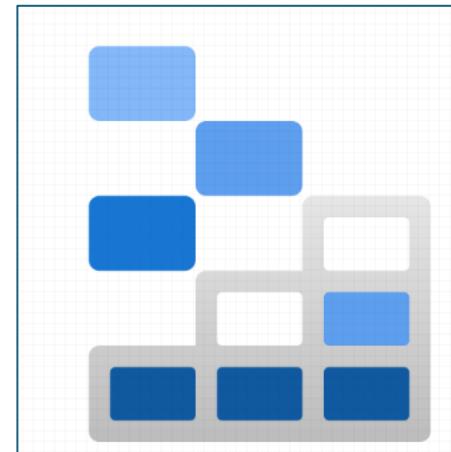
OneLake File Explorer

OneLake File Explorer

The OneLake file explorer application seamlessly integrates OneLake with Windows File Explorer

[Download OneLake app](#)

Azure Storage Explorer



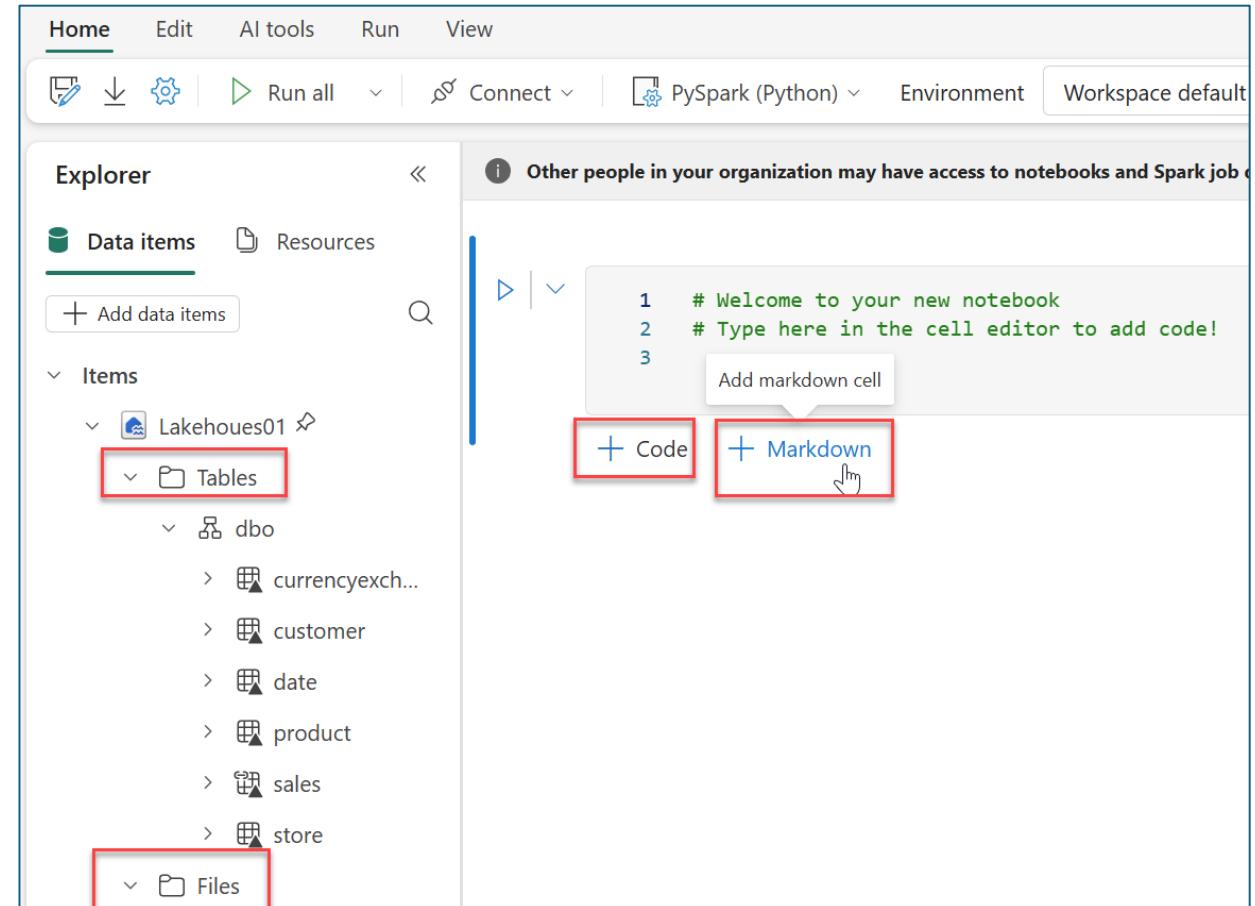
One Lake - Acquiring Data

2 Data Engineering (Notebooks)

Get Data

Prepare data

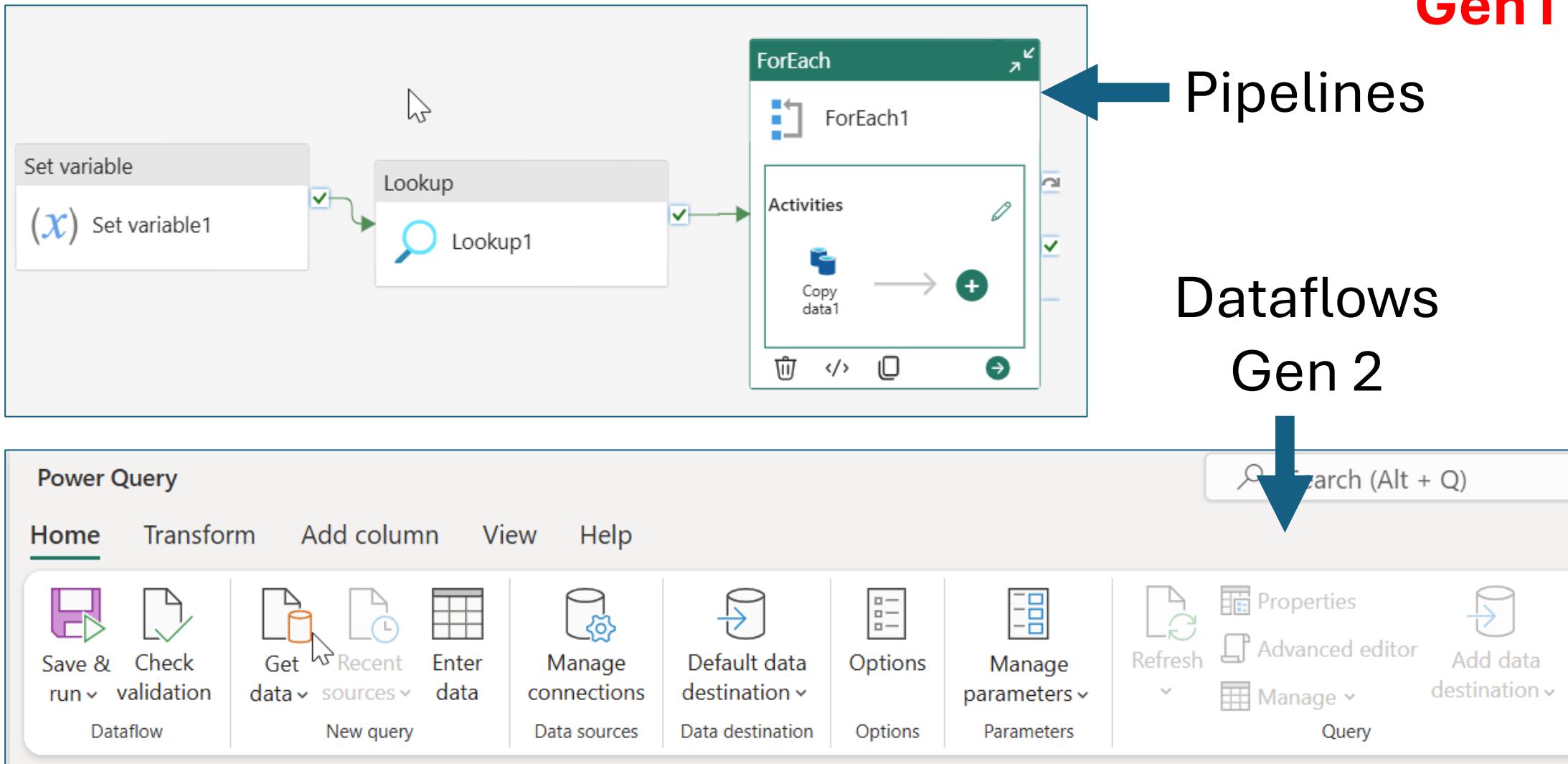
Analyse and Train Data



One Lake - Acquiring Data

3 Data Integration

What about
Dataflow
Gen1?



One Lake - Acquiring Data (Pipelines)

3 Data Integration

Fabric Data Factory (SaaS) – **Similar to**

Azure Data Factory (ADF) – PaaS

Synapse Pipelines – PaaS

SaaS – (Simplifies) v PaaS (more control)

No Datasets

No Linked Services

Ingesting data and Orchestration

Can call – Notebooks , Dataflow Gen2, Refreshes

One Lake - Acquiring Data (dataflows)

3 Data Integration

Want to create a Dataflow Gen2 instead? ×

Dataflow Gen2 significantly enhances the data ingestion and transformation capabilities available in Dataflow Gen1, with new features such as Output Destinations, Copilot, Fast Copy (for large-scale data ingestion), a new High-Scale Dataflows Transformation engine, VNET Gateway support, enhanced Refresh History and Diagnostics experiences and more.

[Learn more](#)

Don't show this again, always create a Dataflow Gen1

Yes, create a Dataflow Gen2 No, create a Dataflow Gen1

One Lake - Acquiring Data (dataflows)

3 Data Integration

New Dataflow Gen2

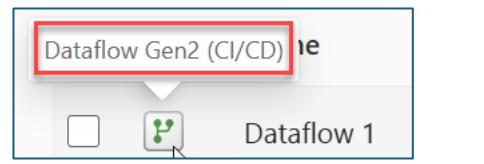
Name
Dataflow 2

Enable Git integration, deployment pipelines and Public API scenarios

Create Cancel

```
graph TD; A[Dataflow Gen2 (CI/CD)] --> B[Dataflow 1]; A --> C[Dataflow 2]
```

Gen 2 – CI/CD



Gen 2



Gen 1



Prepare data

Clean, transform, extract, and load your data for analysis and modeling tasks.

Dataflow Gen1

Prep, clean, and transform data.

Dataflow Gen2

Prep, clean, and transform data.

Others

Find unique or third-party provided functionality that builds on Fabric's core capabilities.

Streaming dataflow

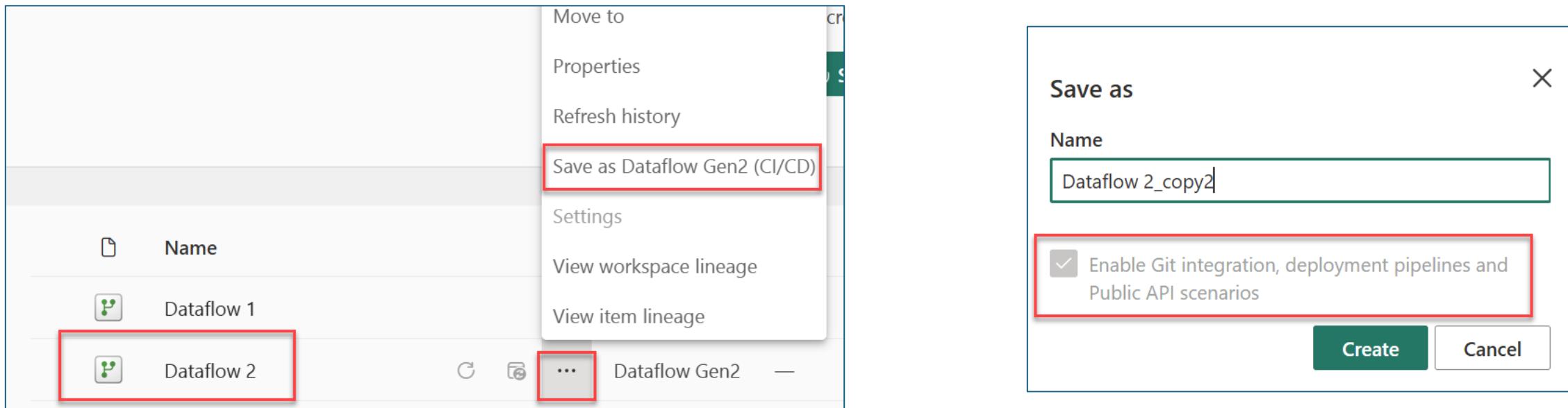
Combine and transform streaming data.

```
graph LR; A[Dataflow Gen1] --- B[Prep, clean, and transform data.]; A --- C[★]; A[Dataflow Gen2] --- D[Prep, clean, and transform data.]; A --- E[★]; A[Streaming dataflow] --- F[Combine and transform streaming data.]; A --- G[★]
```

One Lake - Acquiring Data (dataflows)

3 Data Integration

Upgrade a ‘Dataflow Gen 2’ to ‘Dataflow Gen 2 CI/CD’



Creates a copy of the dataflow gen2 with a new name
Upgrade (copy) a dataflow gen1 to dataflow gen 2 CI/CD

One Lake - Acquiring Data (Data Flows)

3 Data Integration

Power BI Data Flow (Dataflow Gen1) v Dataflow Gen 2

Use Dataflow Gen 2 for new development

Better performance

Fast copy capabilities

Simpler creation process

Multiple output destinations

Works with pipelines

One Lake - Acquiring Data (Data Flows)

4 Notebooks

PySpark Notebooks

Multi Language Support

PySpark

Spark SQL

Scala

SparkR

Python Notebooks

One Lake - Acquiring Data

4 Shortcuts

New shortcut

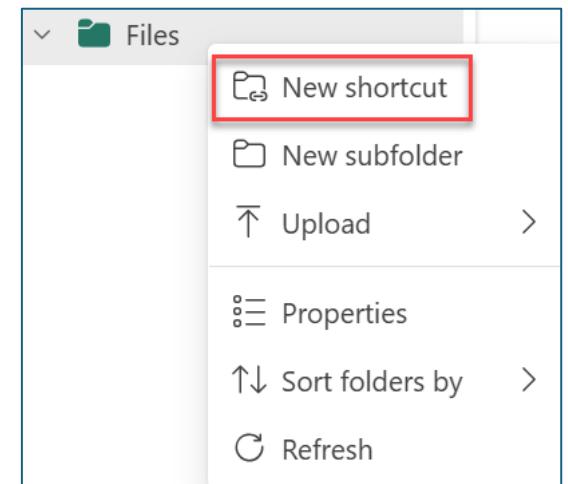
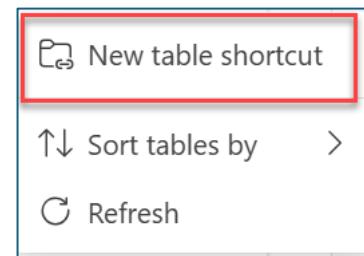
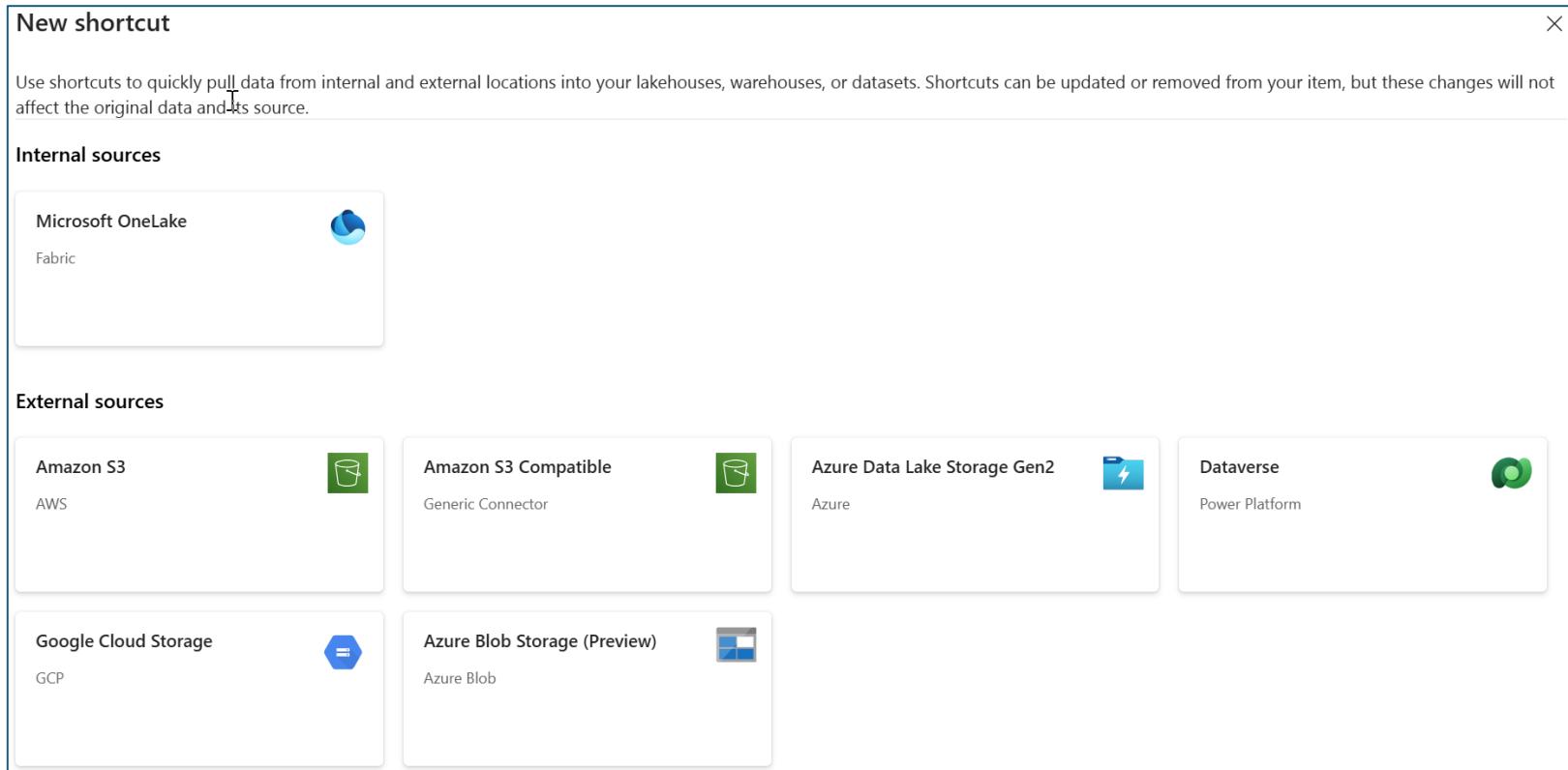
Use shortcuts to quickly pull data from internal and external locations into your lakehouses, warehouses, or datasets. Shortcuts can be updated or removed from your item, but these changes will not affect the original data and its source.

Internal sources

- Microsoft OneLake

External sources

- Amazon S3
- Amazon S3 Compatible
- Azure Data Lake Storage Gen2
- Dataverse
- Google Cloud Storage
- Azure Blob Storage (Preview)



One Lake - Acquiring Data

5 Mirroring

Get data

Ingest batch and real-time data into a single location within your Fabric workspace.

Mirrored Azure Cosmos DB (pre...) 

Easily replicate data from an existing source into an analytics-friendly format.



Mirrored Azure Database for Po... 

Easily replicate data from an existing source into an analytics-friendly format.



Mirrored Azure Databricks catalog 

Explore Unity Catalog Tables



Mirrored Azure SQL Database 

Easily replicate data from an existing source into an analytics-friendly format.



Mirrored Azure SQL Managed In... 

Easily replicate data from an existing source into an analytics-friendly format.



Mirrored database 

Easily replicate data from an existing source into an analytics-friendly format.



Mirrored Snowflake 

Easily replicate data from an existing source into an analytics-friendly format.

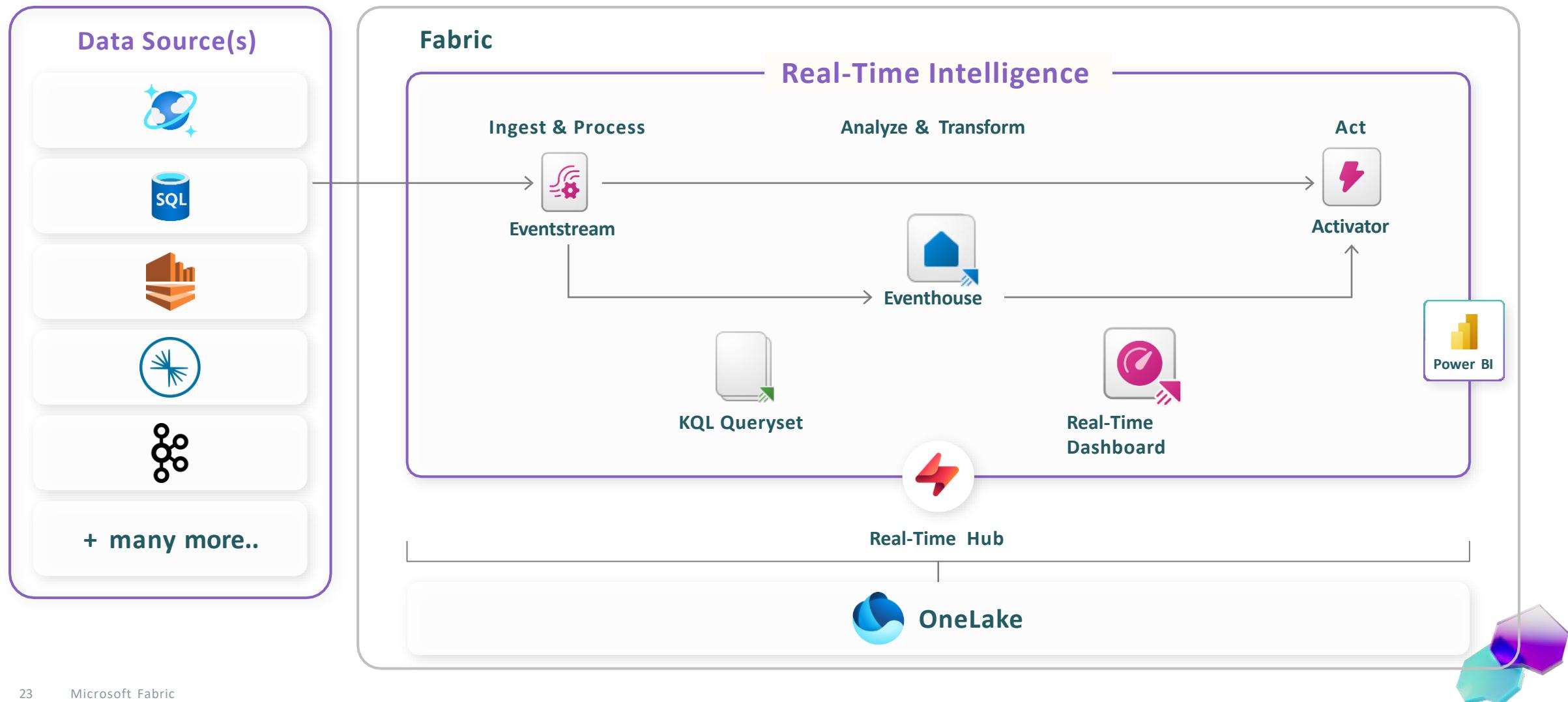


Mirrored SQL Server (preview) 

Easily replicate data from an existing source into an analytics-friendly format.



Components of Fabric's Real-Time Intelligence



One Lake - Acquiring Data

6 Real Time

A KQL database
within an
Eventhouse



Turn on
OneLake Availability
for the
Eventhouse

One Lake - Acquiring Data

7 Semantic Model

Upload an existing Semantic Model

Connect to it – example using a notebook

Acquire data

[Read data from semantic models and write data that semantic models can consume using Spark - Microsoft Fabric | Microsoft Learn](#)

One Lake - Acquiring Data

8 Programmatically

ReST API's

Various API's

Fabric CLI

One Lake - Acquiring Data

9 SQL on Fabric

Azure SQL Database inside Fabric

2 End Points

Database – Read & Write

DatabaseWarehouse – Read Only

One Lake - Acquiring Data

10 Warehouse - TSQL

COPY

```
COPY INTO [ ].[ ]  
FROM 'https://...'  
WITH (  
    FILE_TYPE = 'CSV',  
    FIRSTROW = 2  
)
```

OPENROWSET

```
SELECT *  
FROM OPENROWSET(BULK  
'https://...',  
                FORMAT='CSV',  
                HEADER_ROW=True,  
                ROW_TERMINATOR='\n',  
                FIELD_TERMINATOR=',') AS data;
```

CTAS

```
CREATE TABLE ...  
AS  
SELECT DATEPART(YEAR, updated) AS [year],  
       DATEPART(MONTH, updated) AS [month],  
       DATEPART(DAY, updated) AS [dayofmonth],  
       *  
FROM dbo.[...];
```

One Lake - Acquiring Data

- | | | | |
|---|--------------------|----|---------------|
| 1 | Manually | 9 | SQL on Fabric |
| 2 | Data Integration | 10 | Warehouse |
| | - Notebooks | | COPY |
| 3 | Data Engineering | | OPENROWSET |
| | - (Data) Pipelines | | T-SQL (CTAS) |
| | - Dataflow Gen 2 | 11 | Mount ADF |
| 4 | Mirroring | | |
| 5 | Shortcuts | | |
| 6 | Semantic Model | | |
| 7 | Programmatically | | |
| 8 | Real Time | | |

Manually

Data Integration

- Notebooks

Data Engineering

- (Data) Pipelines
- Dataflow Gen2

Mirroring

Shortcuts

Semantic Model

Programmatically

Real Time

SQL on Fabric

Warehouse

- COPY
- OPENROWSET
- T-SQL (CTAS)

Mount ADF

Thank You

Slides



<https://tinyurl.com/3r6jjbkh>