Microsoft

Microsoft Partner Project Ready

Data Integration and Transformation with Data Factory in Microsoft Fabric

Day 01 of 02

Technofocus

Course Plan and Learning Objectives

Day 1

240 mins

Module 1 - Introduction to Data Factory in Microsoft Fabric

- Microsoft Fabric - The unified data platform for AI transformation
- Data Factory in Microsoft Fabric

Module 2 - Ingest data with Data Factory in Microsoft Fabric

- Data ingestion with Pipelines
- Copy activity with Pipelines
- Use parameters and expressions in Pipelines
- Data ingestion with Copy Job
- Mirroring databases in Microsoft Fabric
- Execute, monitor and troubleshoot Pipelines
- Data Pipeline storage event triggers
- REST API capabilities and CI/CD for Pipelines
- Fabric Pipelines vs. ADF/Synapse Pipelines

Hands-on labs

- Use case 01: Implementing Medallion Architecture with Data Factory in Microsoft Fabric for scalable data processing

Day 2

240 mins

Module 3 - Data Transformation with Dataflows Gen2

- Dataflows Gen2 in Microsoft Fabric
- Fast Copy in Dataflows Gen2
- Dataflow Gen2 with CI/CD and Git integration support
- Monitor your Dataflows
- Copilot for Data Factory in Microsoft Fabric
- Workflow Orchestration with Apache Airflow job

Module 4 - Migrate to Data Factory in Microsoft Fabric

- Plan your migration from Azure Data Factory to Data Factory in Microsoft Fabric
- Migrate from Dataflow Gen1 to Dataflow Gen2
- Move queries from Dataflow Gen1 to Dataflow Gen2
- Ingest data into Microsoft Fabric using the Azure Data Factory Copy Activity

Hands-on labs

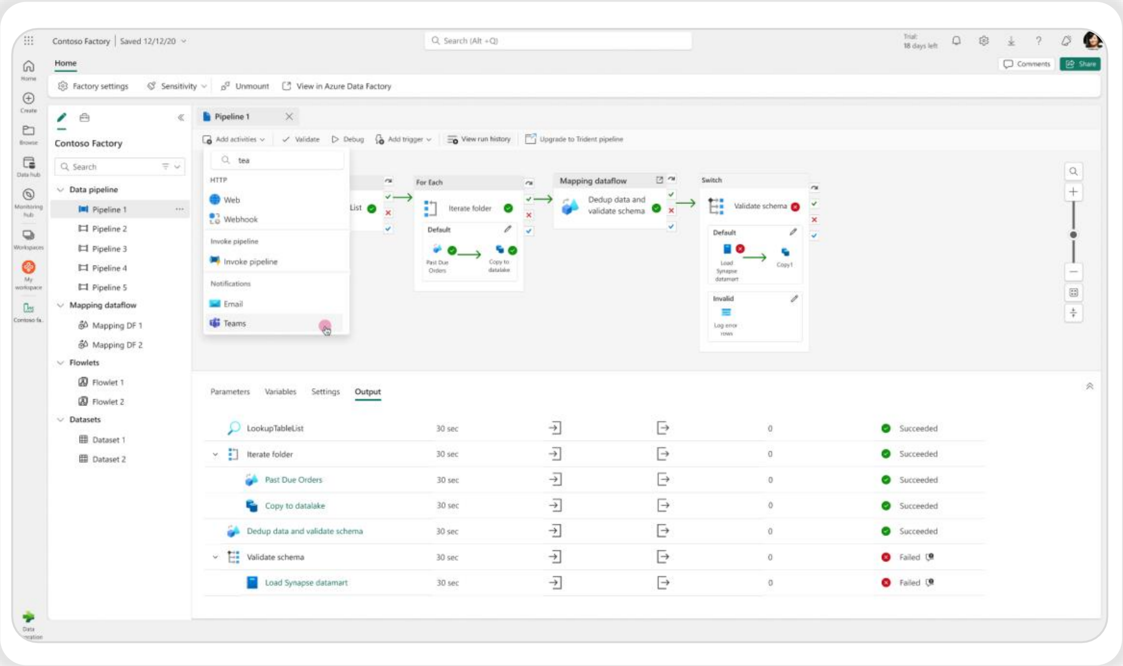
- Use Case 02: Data Factory solution for moving and transforming data with dataflows and data pipelines

04
Migrate to Data Factory in Microsoft Fabric

Plan your migration from Azure Data Factory to Data Factory in Microsoft Fabric

Migrate from Azure Data Factory to Data Factory in Microsoft Fabric

Design Principle



LookupTableList	30 sec	→	→	0	Succeeded
Iterate folder	30 sec	→	→	0	Succeeded
Past Due Orders	30 sec	→	→	0	Succeeded
Copy to datalake	30 sec	→	→	0	Succeeded
Dedup data and validate schema	30 sec	→	→	0	Succeeded
Validate schema	30 sec	→	→	0	Failed
Load Synapse datamart	30 sec	→	→	0	Failed

Frictionless Transition

Provides a seamless integration of Azure Data Factory into Microsoft Fabric, allowing users to easily manage their data pipelines.

Modernized Experience

Provides a modernized user interface and pipeline management experience for the pipelines available within the mounted Azure data factory

New Capabilities

Provides with exciting opportunities to streamline their data integration processes and unblock new capabilities.

Migration benefits

- Easy routing of messages during pipeline execution with email and Teams activities

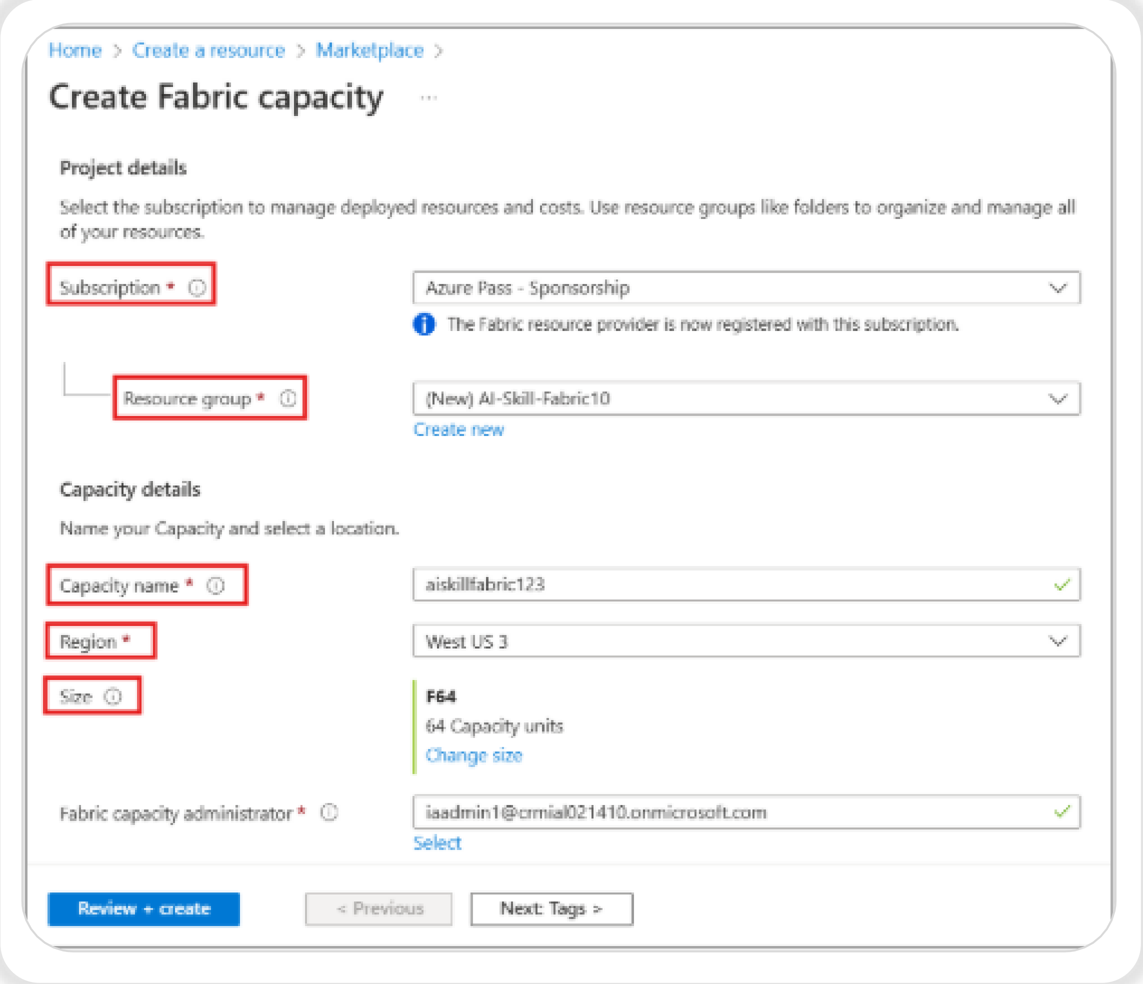
- Built-in continuous integration and delivery (CI/CD) features (deployment pipelines) don't require external integration with Git repositories

- Workspace integration with your OneLake data lake enables single-pane-of-glass easy analytics management

- Refreshing your semantic data models is easy in Fabric with a fully integrated pipeline activity

License requirement

- Fabric pipelines and Dataflow Gen2 require a Microsoft Fabric premium capacity workspace



Create Fabric capacity ...

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all of your resources.

Subscription * : Azure Pass - Sponsorship

The Fabric resource provider is now registered with this subscription.

Resource group * : (New) AI-Skill-Fabric10

Capacity details

Name your Capacity and select a location.

Capacity name * : aiskillfabric123

Region * : West US 3

Size : F64
64 Capacity units

Fabric capacity administrator * : iaadmin1@cmial021410.onmicrosoft.com

Platform differences between ADF and Data Factory in Fabric

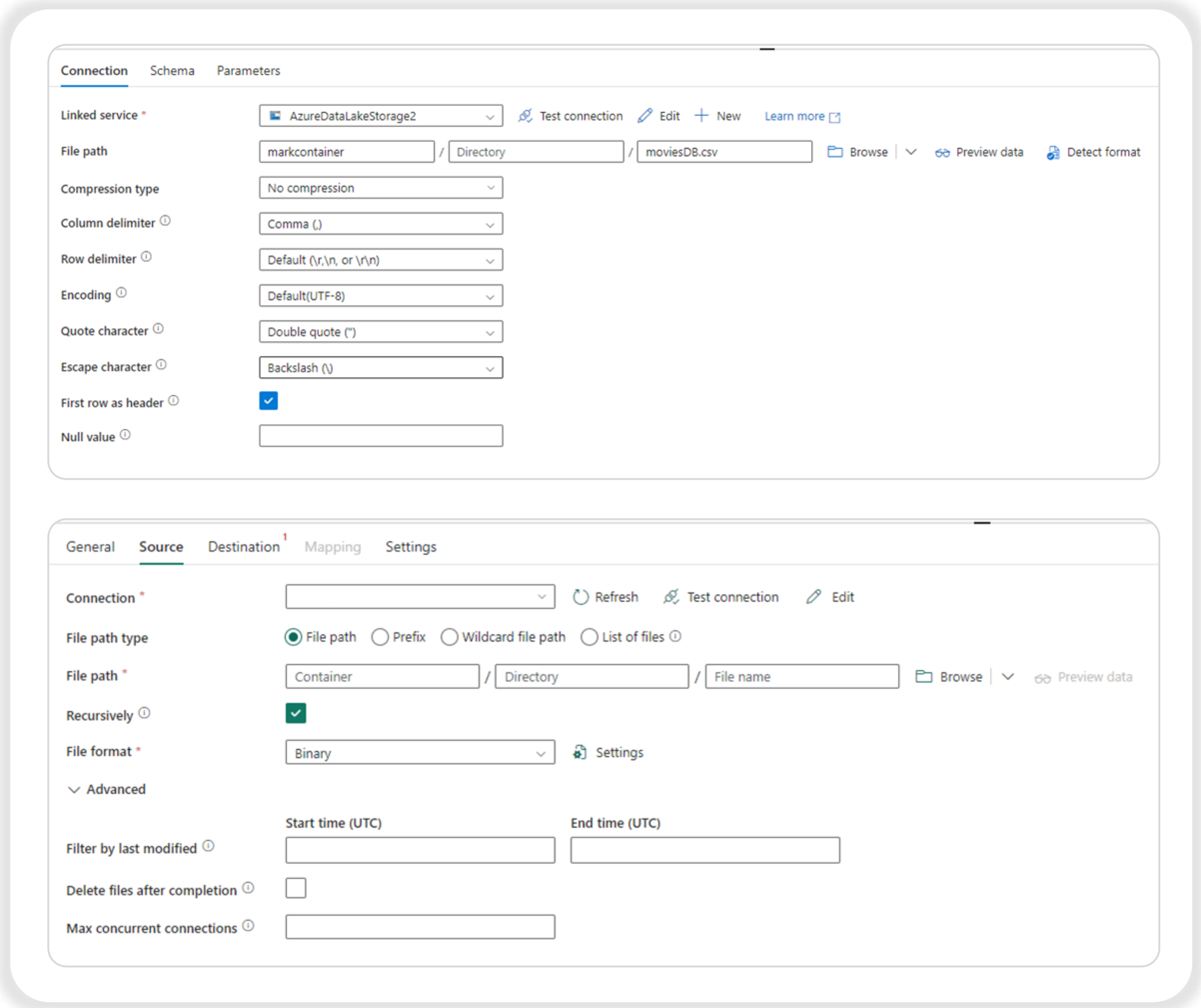
Features	Azure Data Factory	Data Factory in Fabric
	PaaS product	SaaS product
Integration Runtimes	IRs are configuration objects that represent compute	Default is to use cloud-based compute
Pipelines	Fundamental component for the primary workflow and orchestration of your ADF processes	Pipelines in Fabric Data Factory are nearly identical to ADF. The JSON definition of pipelines differs slightly
Linked Services	Linked Services define the connectivity properties needed to connect to your data stores	Recreate definitions as Connections
Datasets	Datasets define the shape, location, and contents of your data in ADF	Don't exist as entities in Fabric
Dataflows	Data Flows are built on the Synapse Spark infrastructure	Power Query based Dataflows

Platform differences between ADF and Data Factory in Fabric

Features	Azure Data Factory	Data Factory in Fabric
Triggers	Features are similar in Fabric although the underlying implementation is different	Triggers only exist as a pipeline concept. Data Activator framework is used
Debugging	Separate debug mode you find in ADF pipelines and data flows	Debugging pipelines is simpler in Fabric than in ADF
Change Data Capture	Preview feature, move data quickly in an incremental manner	To migrate CDC artifacts to Fabric Data Factory, recreate the artifacts as Copy job
Azure Synapse Link	Not available in ADF	Recreate the Azure Synapse Link artifacts as Mirroring items
SQL Server Integration Services (SSIS)	Lift-and-shift your SSIS packages into the cloud using the ADF SSIS IR	Don't have the concept of IRs, so this functionality isn't possible today
Invoke pipeline activity	A common activity: Execute pipeline activity	Enhanced activity: Invoke pipeline activity

Migration scenario – Migrate ADF pipelines and data flows

- Modernize your ETL environment from the ADF factory PaaS model to the new Fabric SaaS model
- The primary factory items to migrate: pipelines and data flows
- Other fundamental factory elements to migrate: linked services, integration runtimes, datasets, and triggers



Linked service * AzureDataLakeStorage2

File path markcontainer / Directory / moviesDB.csv

Compression type No compression

Column delimiter Comma (,)

Row delimiter Default (\r,\n, or \r\n)

Encoding Default(UTF-8)

Quote character Double quote (")

Escape character Backslash (\)

First row as header ☑

Null value

Connection *

File path type ● File path ○ Prefix ○ Wildcard file path ○ List of files

File path * Container / Directory / File name

Recursively ☑

File format * Binary

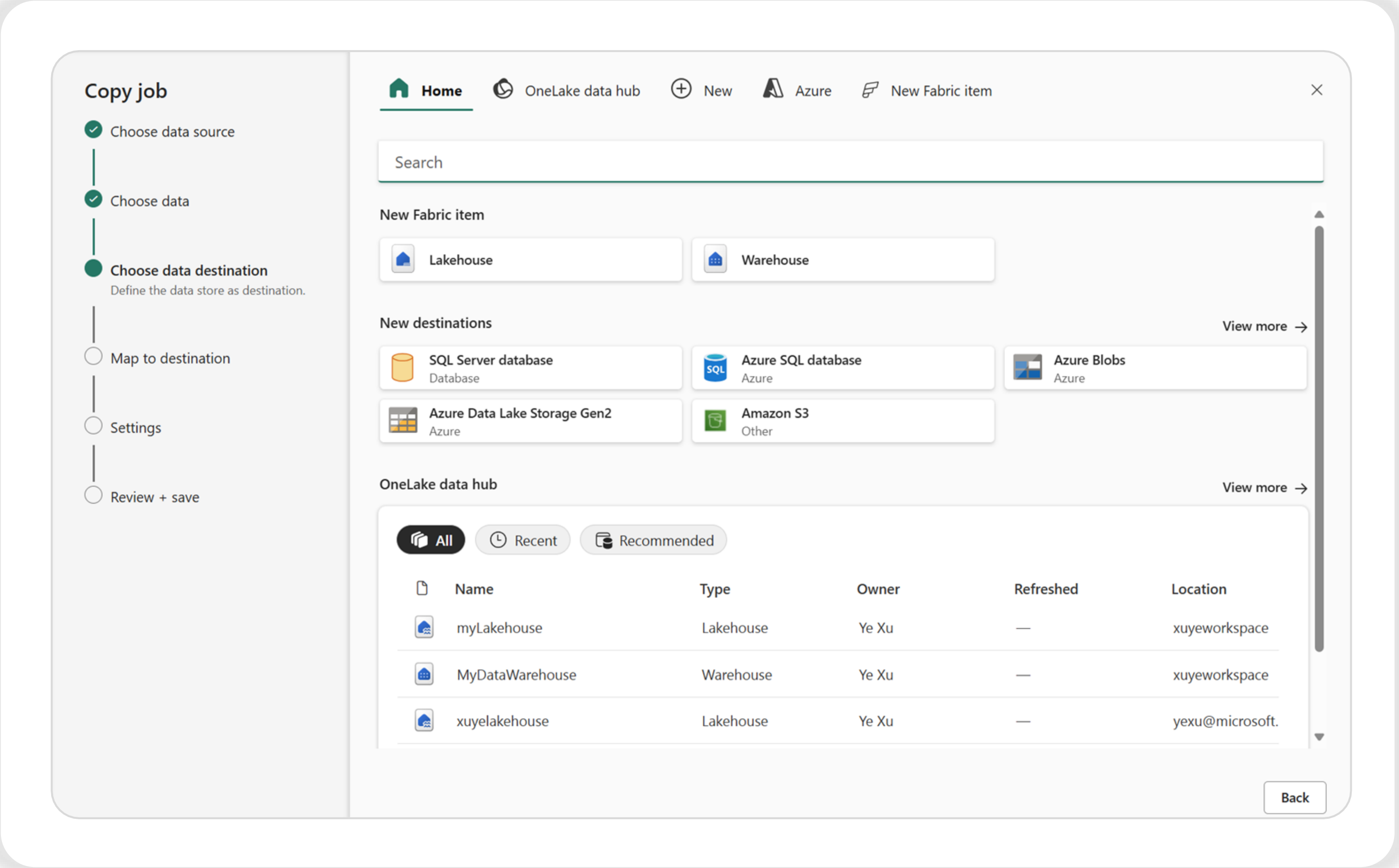
Filter by last modified Start time (UTC) End time (UTC)

Delete files after completion ☐

Max concurrent connections

Migration scenario - Migrate ADF with CDC, SSIS, and Airflow

- These features serves different data integration needs, but require special attention when migrating
- CDC is a top-level ADF concept but in Fabric, you see this capability as the Copy job
- Airflow is available in both



Copy job

- Choose data source
- Choose data
- Choose data destination
  Define the data store as destination.
- Map to destination
- Settings
- Review + save

New Fabric item

Lakehouse	Warehouse

New destinations

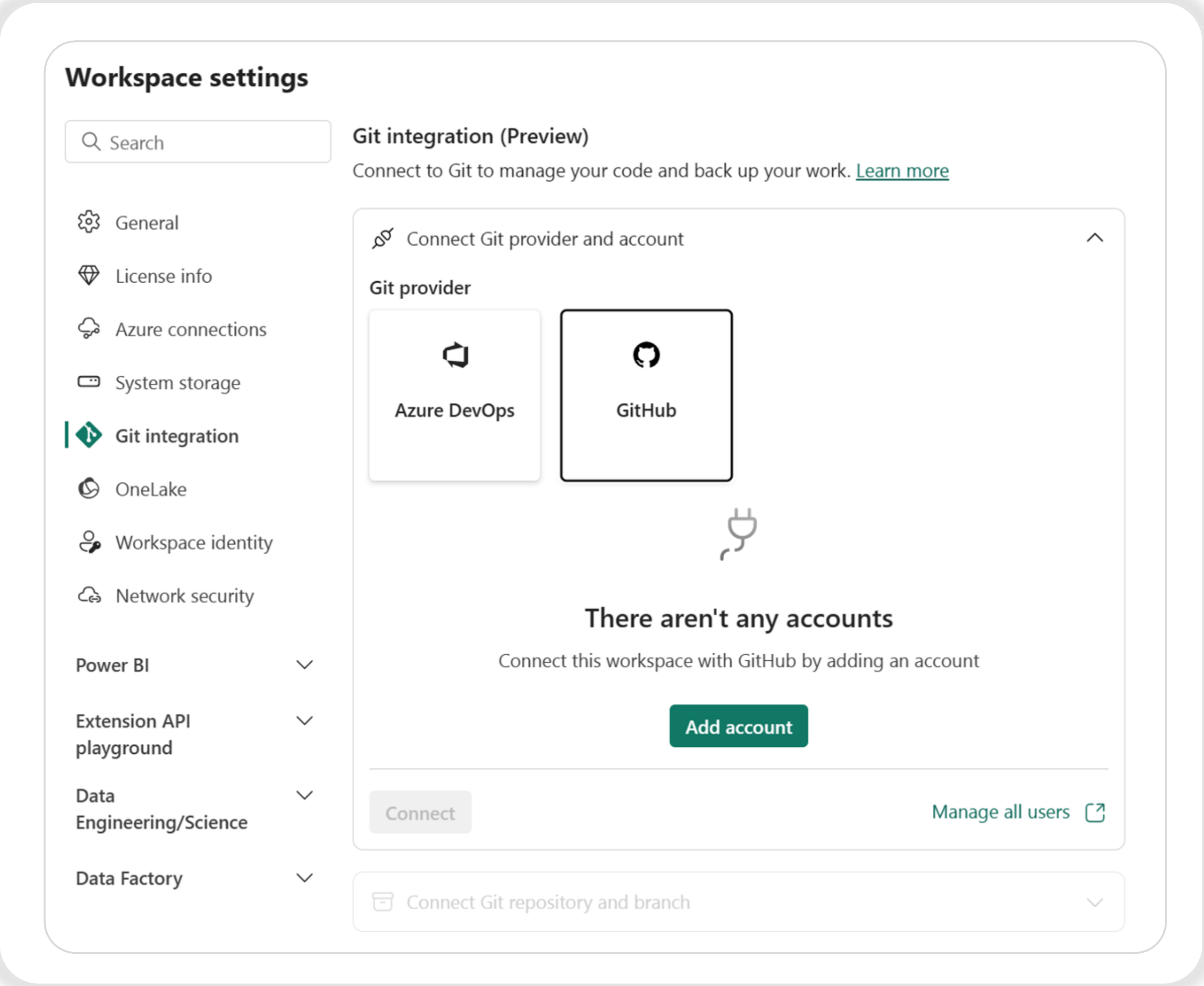
SQL Server database Database	Azure SQL database Azure	Azure Blobs Azure
Azure Data Lake Storage Gen2 Azure	Amazon S3 Other	

OneLake data hub

Name	Type	Owner	Refreshed	Location
myLakehouse	Lakehouse	Ye Xu	—	xuyeworkspace
MyDataWarehouse	Warehouse	Ye Xu	—	xuyeworkspace
xuyelakehouse	Lakehouse	Ye Xu	—	yexu@microsoft.

Migration scenario – Git-enabled Data Factory migration to Fabric

- Fabric provides two primary ways to enable CI/CD, both at the workspace level: Git integration and built-in deployment pipelines
- Existing Git repo from ADF doesn't work with Fabric



Workspace settings

Git integration (Preview)

Connect to Git to manage your code and back up your work. Learn more

Connect Git provider and account

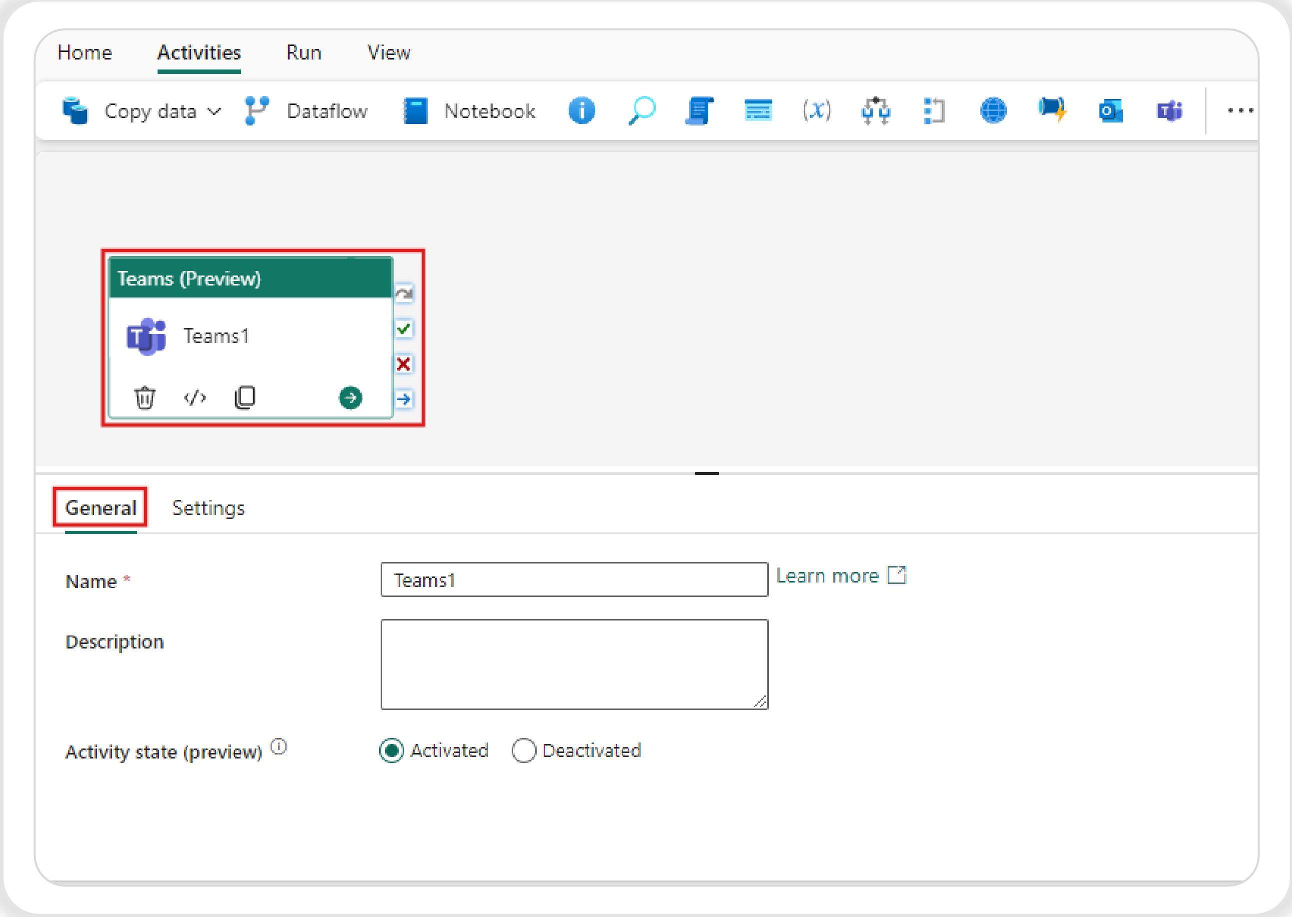
Git provider

There aren't any accounts

Connect this workspace with GitHub by adding an account

Activity continuity

- Data Factory in Fabric aims to offer comprehensive capabilities while maintaining compatibility with ADF
- New activities in Fabric Data Factory
  - Outlook
  - Teams
  - Semantic model refresh
  - Dataflow Gen2



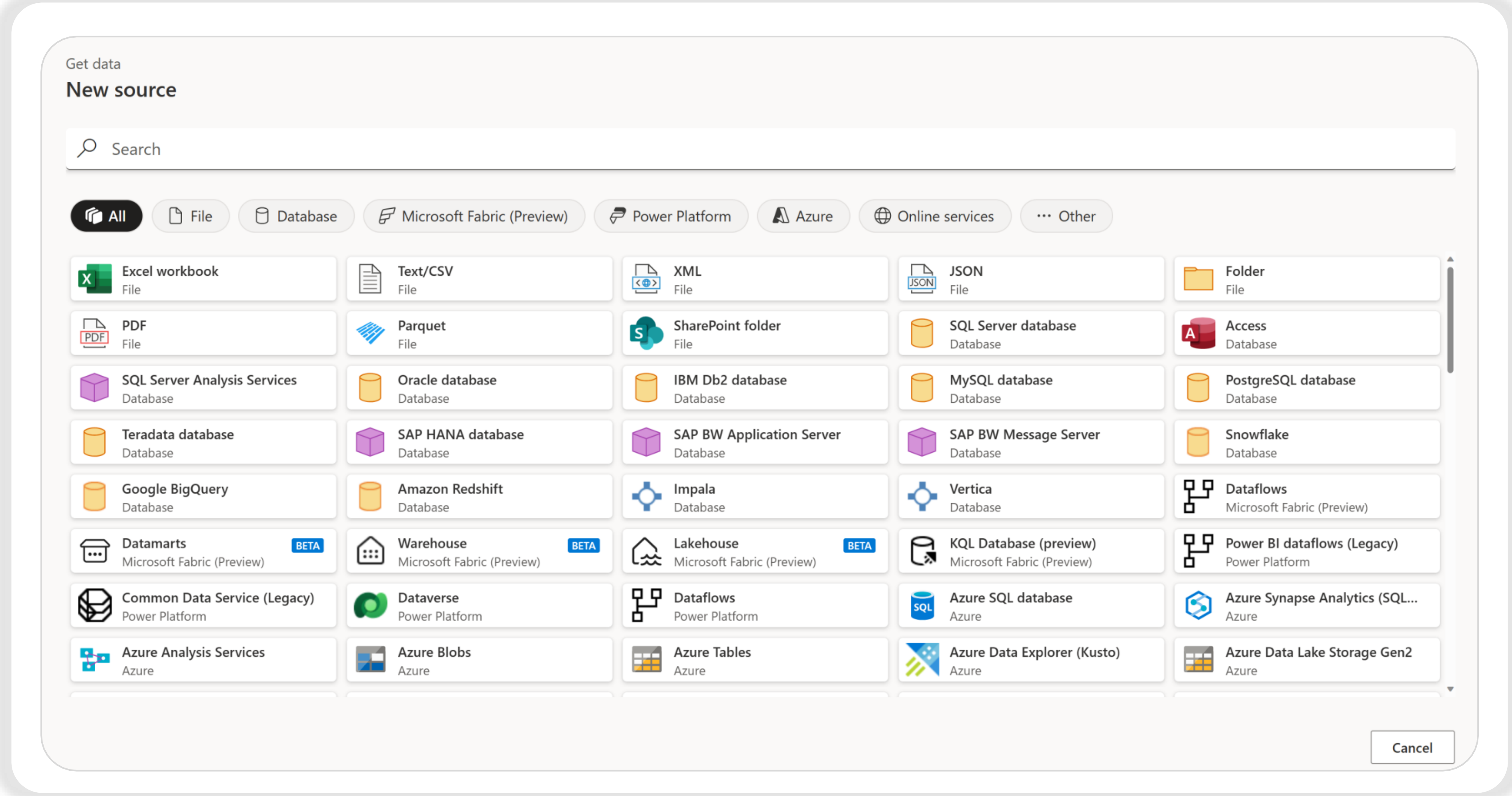
Name * Teams1 Learn more

Description

Activity state (preview) ● Activated ○ Deactivated

Connector continuity

With Data Factory in Microsoft Fabric, data pipelines provide connectivity to a rich set of data sources



Get data

New source

Search

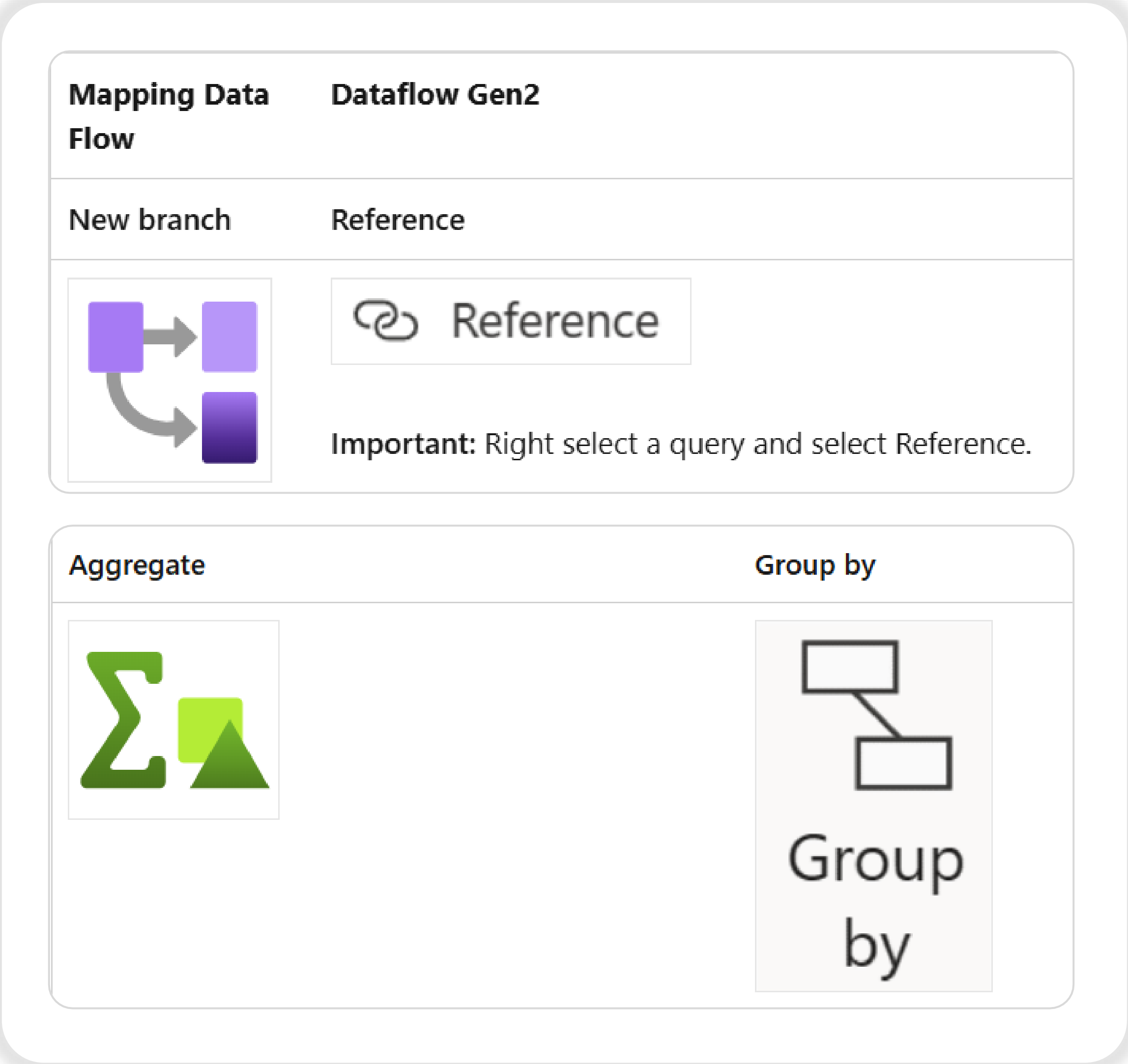
All | File | Database | Microsoft Fabric (Preview) | Power Platform | Azure | Online services | ··· Other

Excel workbook File	Text/CSV File	XML File	JSON File	Folder File
PDF File	Parquet File	SharePoint folder File	SQL Server database Database	Access Database
SQL Server Analysis Services Database	Oracle database Database	IBM Db2 database Database	MySQL database Database	PostgreSQL database Database
Teradata database Database	SAP HANA database Database	SAP BW Application Server Database	SAP BW Message Server Database	Snowflake Database
Google BigQuery Database	Amazon Redshift Database	Impala Database	Vertica Database	Dataflows Microsoft Fabric (Preview)
Datamarts BETA Microsoft Fabric (Preview)	Warehouse BETA Microsoft Fabric (Preview)	Lakehouse BETA Microsoft Fabric (Preview)	KQL Database (preview) Microsoft Fabric (Preview)	Power BI dataflows (Legacy) Power Platform
Common Data Service (Legacy) Power Platform	Dataverse Power Platform	Dataflows Power Platform	Azure SQL database Azure	Azure Synapse Analytics (SQL... Azure
Azure Analysis Services Azure	Azure Blobs Azure	Azure Tables Azure	Azure Data Explorer (Kusto) Azure	Azure Data Lake Storage Gen2 Azure

Cancel

Mapping Data Flow transformations to the Dataflow Gen2

- When first starting to author Dataflows, you can also use the Global search box
- Mapping data flow transformations supported in Dataflow Gen2
  - Multiple inputs/outputs
  - Schema modifier
  - Formatters
  - Row modifier
  - Flowlets
  - Destination

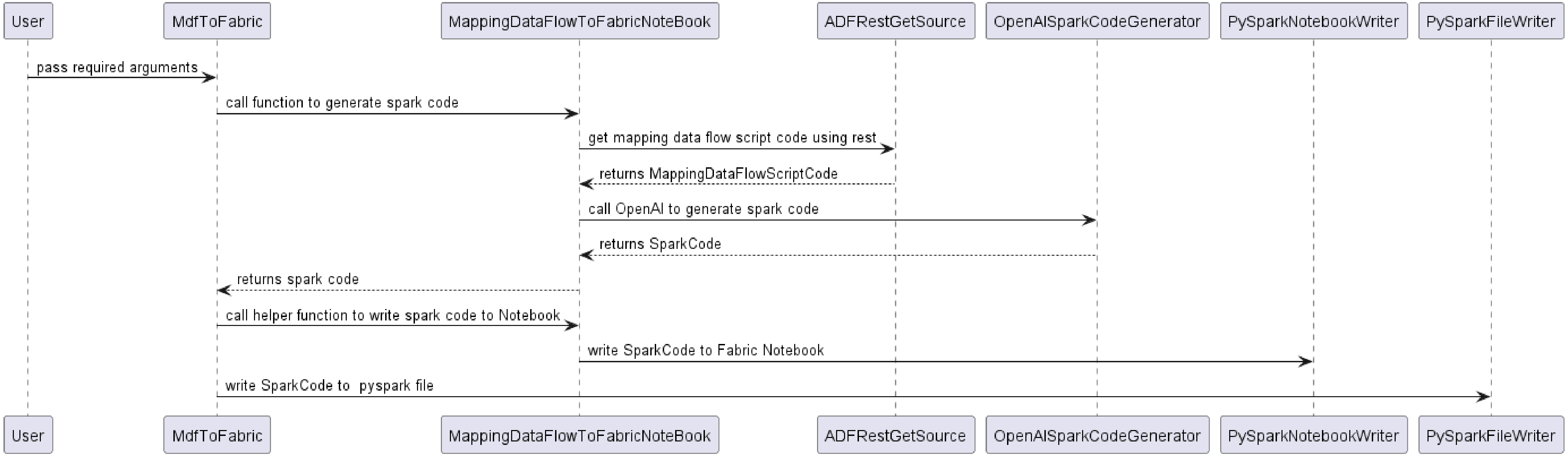


Mapping Data Flow	Dataflow Gen2
New branch	Reference
	Reference **Important:** Right select a query and select Reference.

Aggregate	Group by
	Group by

Mapping data flows to PySpark Notebooks

The Mapping data flow to Microsoft Fabric notebook is a command line tool to convert Mapping data flow script code to Spark Scala and PySpark code


User
MdfToFabric
MappingDataFlowToFabricNoteBook
ADFRestGetSource
OpenAISparkCodeGenerator
PySparkNotebookWriter
PySparkFileWriter
pass required arguments
call function to generate spark code
get mapping data flow script code using rest
returns MappingDataFlowScriptCode
call OpenAI to generate spark code
returns SparkCode
returns spark code
call helper function to write spark code to Notebook
write SparkCode to Fabric Notebook
write SparkCode to pyspark file
User
MdfToFabric
MappingDataFlowToFabricNoteBook
ADFRestGetSource
OpenAISparkCodeGenerator
PySparkNotebookWriter
PySparkFileWriter

Mounting

Within Fabric, you will have the ability to mount existing Synapse Pipeline/Azure Data Factory artifact. This allows you to still preserve the feature in ADF but also leverage new features in Microsoft Fabric.

- The mount provides the interface UI in Fabric to allow customers to develop/edit their existing ADF investment and it is NOT migration.

- The mount provide the interface UI in Microsoft Fabric platform to allow customer to edit

- Development can happen within Fabric or within the existing ADF workspace.

- User is allowed to add activity that is only available in Microsoft Fabric – E-mail notification, Team Activity, destination output, etc

- Everything user do right now, whether custom Dev/Ops, meta driven pipeline should work with mount, since this is nothing for than a UI, but still using the Azure platform.

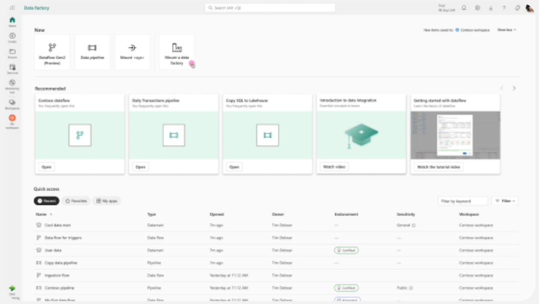
- The backend is still using the Azure platform, until it is migrated to the Microsoft Fabric platform

- ADF customer can also invoke activity in Microsoft Fabric.

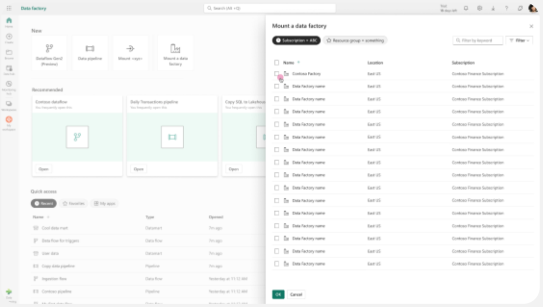
Use Case - Mount/Unmount

This feature allows users to select Azure Data Factory as their preferred mount point, providing them with an easy and seamless way to use their ADF in Fabric

Start from Fabric Data Integration homepage



Choose target data factories



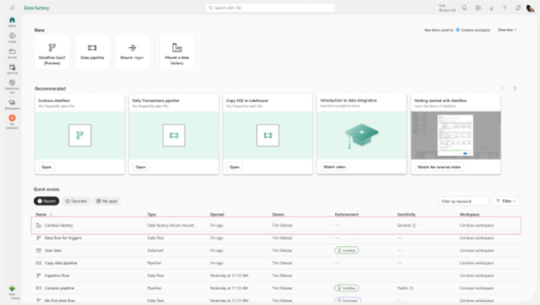
Mount a data factory

Subscription = ABC · Resource group = something · Filter by keyword · Filter

Name	Location	Subscription
Contoso Factory	East US	Contoso Finance Subscription
Data Factory name	East US	Contoso Finance Subscription
Data Factory name	East US	Contoso Finance Subscription
Data Factory name	East US	Contoso Finance Subscription
Data Factory name	East US	Contoso Finance Subscription
Data Factory name	East US	Contoso Finance Subscription
Data Factory name	East US	Contoso Finance Subscription
Data Factory name	East US	Contoso Finance Subscription
Data Factory name	East US	Contoso Finance Subscription
Data Factory name	East US	Contoso Finance Subscription
Data Factory name	East US	Contoso Finance Subscription
Data Factory name	East US	Contoso Finance Subscription
Data Factory name	East US	Contoso Finance Subscription
Data Factory name	East US	Contoso Finance Subscription
Data Factory name	East US	Contoso Finance Subscription
Data Factory name	East US	Contoso Finance Subscription

OK · Cancel

Mount to the Fabric workspace



Quick access

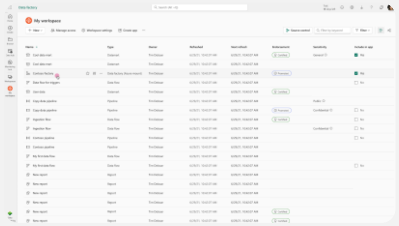
Name	Type	Opened	Owner	Endorsement	Sensitivity	Workspace
Contoso factory	Data factory (Azure mount)	7m ago	Tim Deboer	—	General	Contoso workspace
Data flow for triggers	Data flow	7m ago	Tim Deboer	—	—	Contoso workspace
User data	Datamart	7m ago	Tim Deboer	Certified	—	Contoso workspace
Copy data pipeline	Pipeline	7m ago	Tim Deboer	—	—	Contoso workspace
Ingestion flow	Data flow	Yesterday at 11:12 AM	Tim Deboer			Contoso workspace
Contoso pipeline	Pipeline	Yesterday at 11:12 AM	Tim Deboer	Certified	Public	Contoso workspace
My first data flow	Data flow	Yesterday at 11:12 AM	Tim Deboer	Promoted	—	Contoso workspace

- The mount can be filter at the subscription, resource group. It allows user to mount into the Fabric WS with just one click.
- User also has the option to unmount to remove from the Fabric WS
- This mount option will also be available in existing ADF/Synapse Pipeline to the Fabric WS

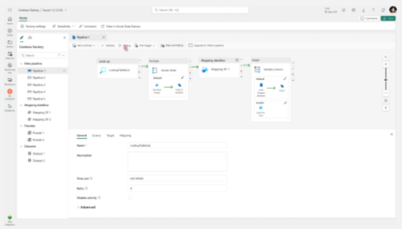
Use Case - Frictionless Transition & Hybrid Development

This feature provides a seamless integration of Azure Data Factory into the Fabric platform, allowing users to easily manage their data pipelines and work with their data in a centralized location.

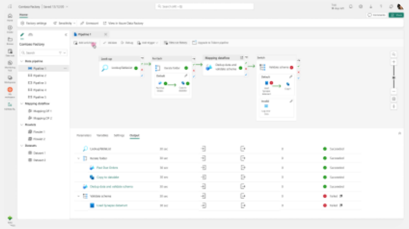
Open a mounted data factory with object explore view



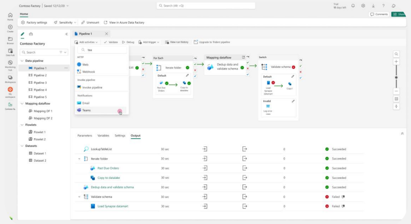
Select a pipeline view on the canvas



Run a mounted pipeline



Add a Teams activity

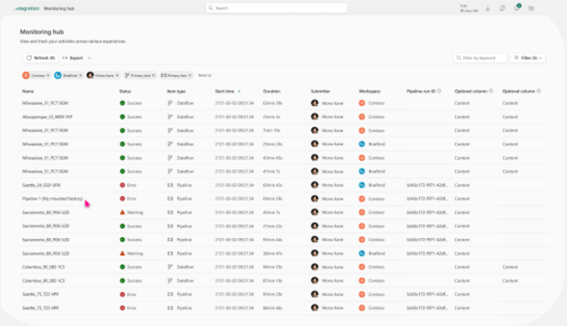


- User can select a mounted ADF in the Fabric WS, all the elements of ADF is visible
- All the changes made in Fabric will be sync back in ADF
- Customer can run the pipeline in Trident allowing them to experience both platforms at the same time

Use Case - Central Monitoring

This feature provides a central place for monitoring of all pipeline run without having to switch between different platforms.

View mounted data factory pipeline in monitoring hub



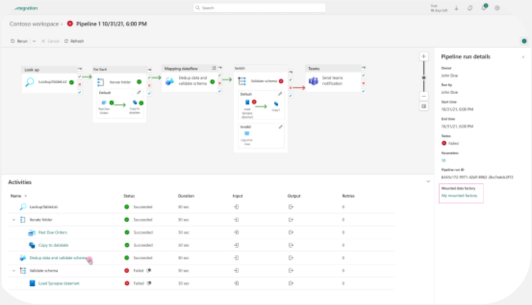
integration Monitoring hub

Monitoring hub

View and track your activities across various experiences.

Name	Status	Item type	Start time	Duration	Submitter	Workspace	Pipeline run ID	Optional column	Optional column
Milwaukee_31_PC7 0GM	Success	Dataflow	2121-03-02 09:21:34	63min 38s	Mona Kane	Contoso		Content	Content
Albuquerque_23_WID9 7KP	Success	Dataflow	2121-03-02 09:21:34	23min 3s	Mona Kane	Contoso		Content	Content
Milwaukee_31_PC7 0GM	Success	Dataflow	2121-03-02 09:21:34	7min 10s	Mona Kane	Contoso		Content	Content
Milwaukee_31_PC7 0GM	Success	Dataflow	2121-03-02 09:21:34	25min 20s	Mona Kane	Bradford		Content	Content
Milwaukee_31_PC7 0GM	Success	Dataflow	2121-03-02 09:21:34	43min 45s	Mona Kane	Contoso		Content	Content
Milwaukee_31_PC7 0GM	Success	Dataflow	2121-03-02 09:21:34	47min 7s	Mona Kane	Bradford		Content	Content
Seattle_24_GQ1 8RN	Error	Pipeline	2121-03-02 09:21:34	60min 42s	Mona Kane	Bradford	b343c172-f971-42df...	Content	
Pipeline 1 (My mounted factory)	Error	Pipeline	2121-03-02 09:21:34	69min 58s	Mona Kane	Contoso	b343c172-f971-42df...	Content	
Sacramento_80_PD6 5ZD	Warning	Pipeline	2121-03-02 09:21:34	45min 7s	Mona Kane	Contoso	b343c172-f971-42df...	Content	
Sacramento_80_PD6 5ZD	Success	Pipeline	2121-03-02 09:21:34	27min 22s	Mona Kane	Contoso	b343c172-f971-42df...	Content	
Sacramento_80_PD6 5ZD	Success	Pipeline	2121-03-02 09:21:34	95min 44s	Mona Kane	Contoso	b343c172-f971-42df...	Content	
Sacramento_80_PD6 5ZD	Warning	Pipeline	2121-03-02 09:21:34	38min 47s	Mona Kane	Bradford	b343c172-f971-42df...	Content	
Columbus_90_SB3 1CS	Success	Dataflow	2121-03-02 09:21:34	54min 18s	Mona Kane	Contoso		Content	Content
Columbus_90_SB3 1CS	Success	Dataflow	2121-03-02 09:21:34	87min 19s	Mona Kane	Contoso		Content	Content
Seattle_75_T22 4PR	Error	Pipeline	2121-03-02 09:21:34	95min 28s	Mona Kane	Contoso	b343c172-f971-42df...	Content	
Seattle_75_T22 4PR	Error	Pipeline	2121-03-02 09:21:34	96min 46s	Mona Kane	Contoso	b343c172-f971-42df...	Content	

View mapping dataflow execution details in monitoring page



Contoso workspace > Pipeline 1 10/31/21, 6:00 PM

Pipeline run details

Owner
John Doe

Run by
John Doe

Start time
10/31/21, 6:00 PM

End time
10/31/21, 6:00 PM

Status
Failed

Parameters
10

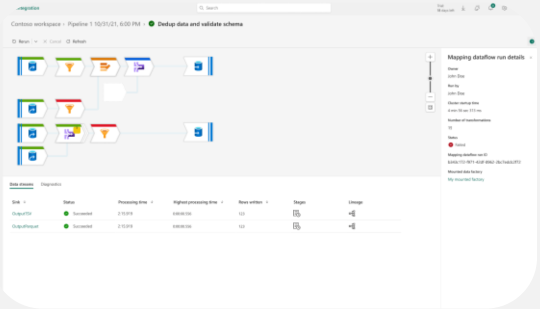
Pipeline run ID
b343c172-f971-42df-8962-2bc7eddc2f12

Mounted data factory
My mounted factory

Activities

Name	Status	Duration	Input	Output	Retries
LookupTableList	Succeeded	30 sec			0
Iterate folder	Succeeded	30 sec			0
Past Due Orders	Succeeded	30 sec			0
Copy to datalake	Succeeded	30 sec			0
Dedup data and validate schema	Succeeded	30 sec			0
Validate schema	Failed	30 sec			0
Load Synapse dataset	Failed	30 sec			0

View mounted data factory pipeline execution details in pipeline monitoring page



Contoso workspace > Pipeline 1 10/31/21, 6:00 PM > Dedup data and validate schema

Mapping dataflow run details

Owner
John Doe

Run by
John Doe

Cluster startup time
4 min 56 sec 333 ms

Number of transformations
11

Status
Failed

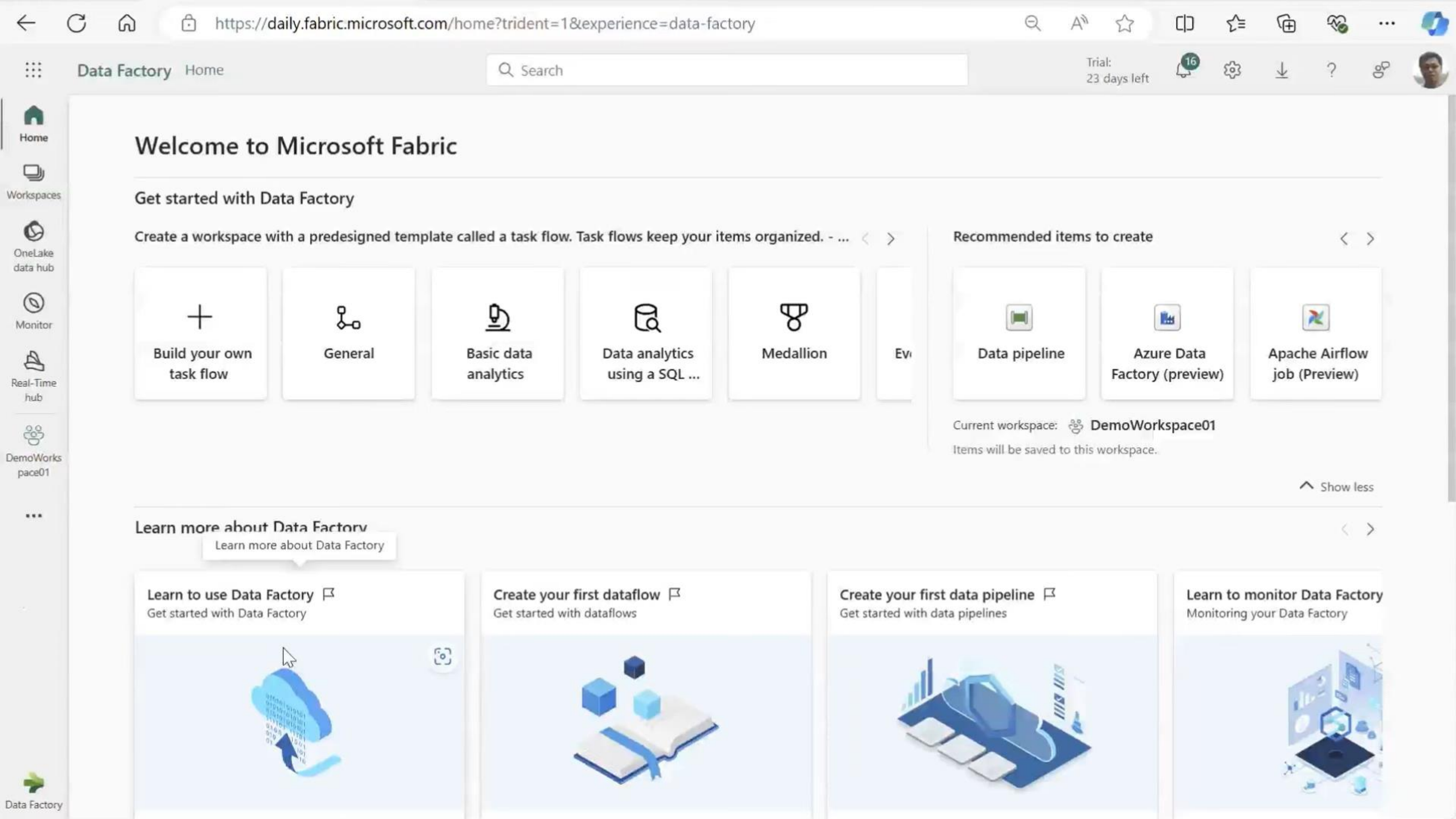
Mapping dataflow run ID
b343c172-f971-42df-8962-2bc7eddc2f12

Mounted data factory
My mounted factory

Sink	Status	Processing time	Highest processing time	Rows written	Stages	Lineage
OutputTSV	Succeeded	2:15:919	0:00:00.556	123		
OutputParquet	Succeeded	2:15:919	0:00:00.556	123		

Demo

Mount your existing ADF instances directly to a Fabric workspace



Welcome to Microsoft Fabric

Get started with Data Factory

Create a workspace with a predesigned template called a task flow. Task flows keep your items organized. - ...

Recommended items to create

Current workspace: DemoWorkspace01

Items will be saved to this workspace.

Learn more about Data Factory

Learn to use Data Factory
Get started with Data Factory

Create your first dataflow
Get started with dataflows

Create your first data pipeline
Get started with data pipelines

Learn to monitor Data Factory
Monitoring your Data Factory

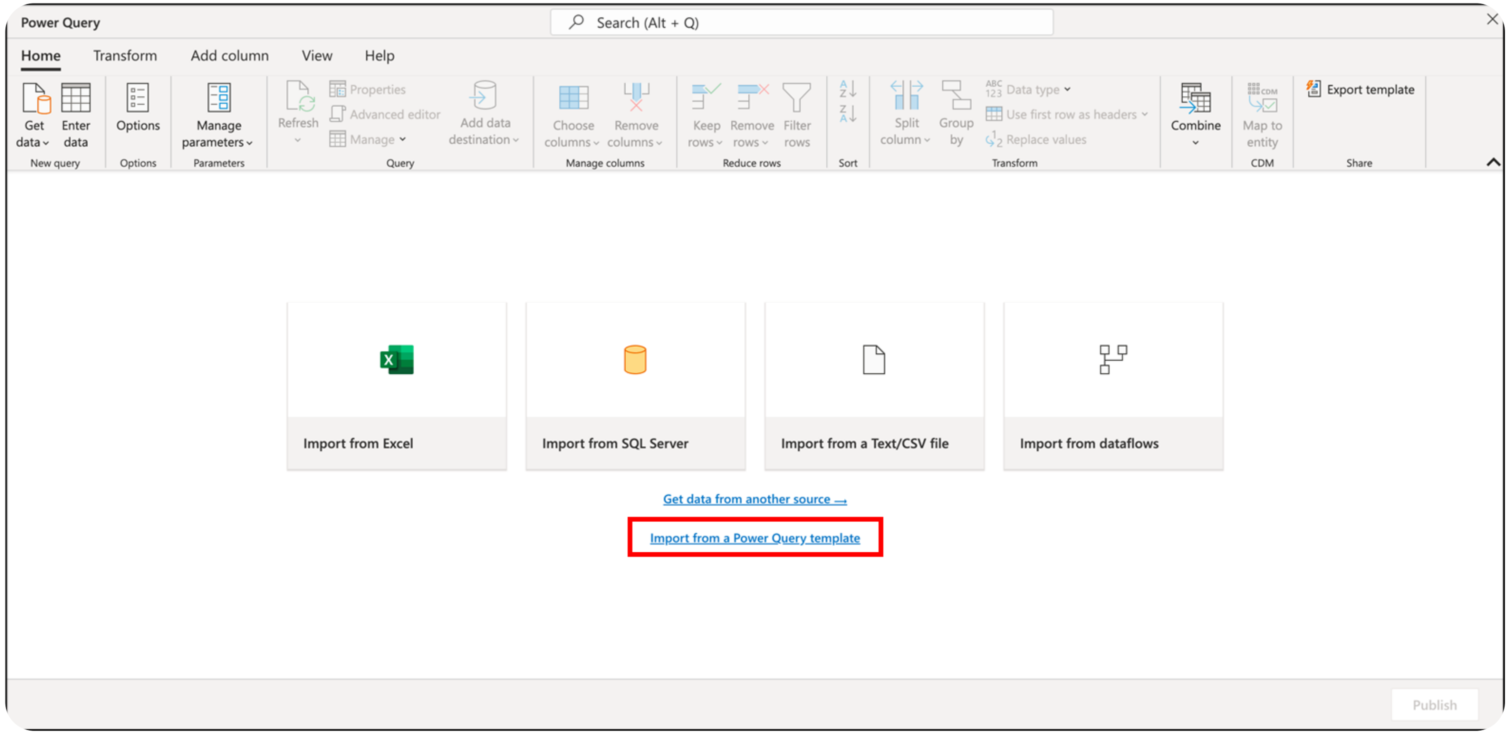
Migrate from Dataflow Gen1 to Dataflow Gen2

Feature overview

Feature	Dataflow Gen2	Dataflow Gen1
Author dataflows with Power Query	✓	✓
Shorter authoring flow	✓	
AutoSave and background publishing	✓	
Data destinations	✓	
Improved monitoring and refresh history	✓	
Integration with data pipelines	✓	
High-scale compute	✓	
Get Data via Dataflows connector	✓	✓
Direct Query via Dataflows connector		✓
Incremental refresh	✓	✓
AI Insights support		✓

Migration guidance

- Compile an inventory of your dataflows and dependent items
    - Dataflows as a source in Power BI
    - Dataflows as a source in Power Apps
    - Dataflows as a source in Excel
- Consider using Power Query templates



Migrate from Dataflow Gen1 to Dataflow Gen2: Migration scenarios

- **Personal or team usage**

  Use dataflows to automate data ingestion and preparation tasks, allowing them to focus on data analysis and insights

- **Departmental usage**

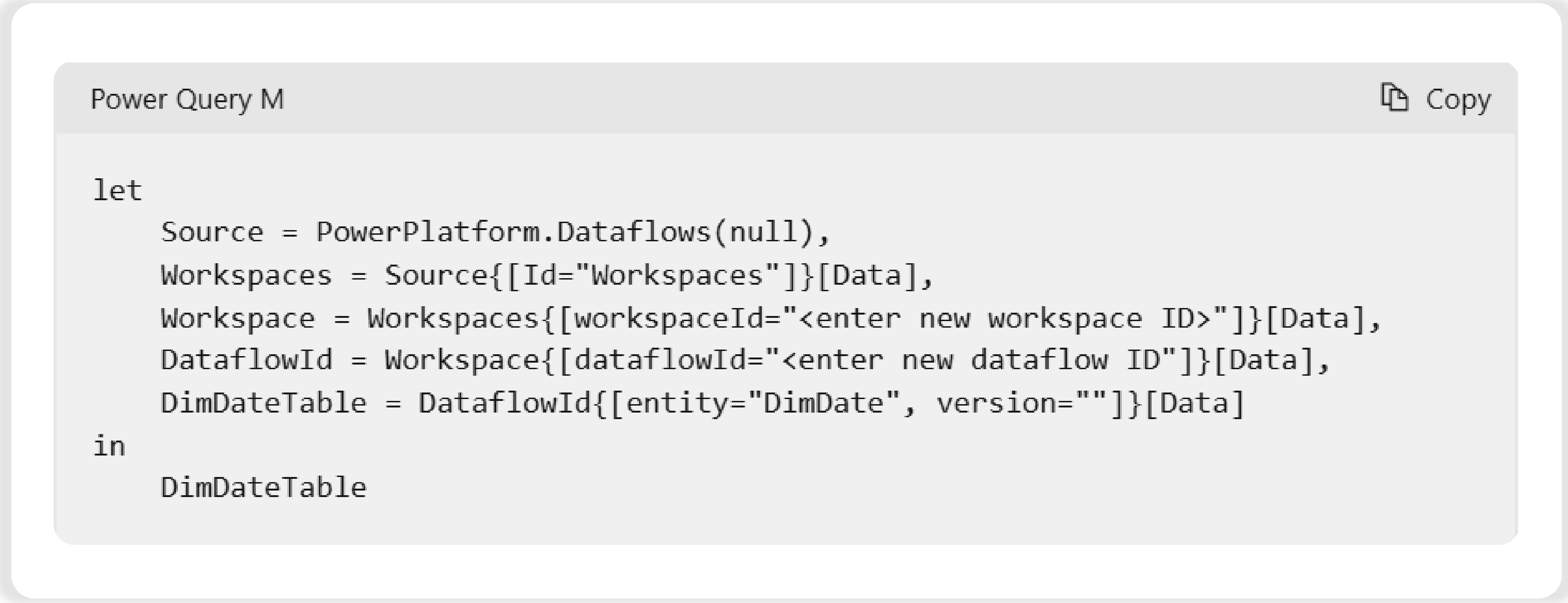
  Use dataflows to manage larger data sources and complex transformations

- **Enterprise usage**

  Use dataflows in ingesting vast amounts of data across multiple departments at scale

Migration scenario - Personal or team usage

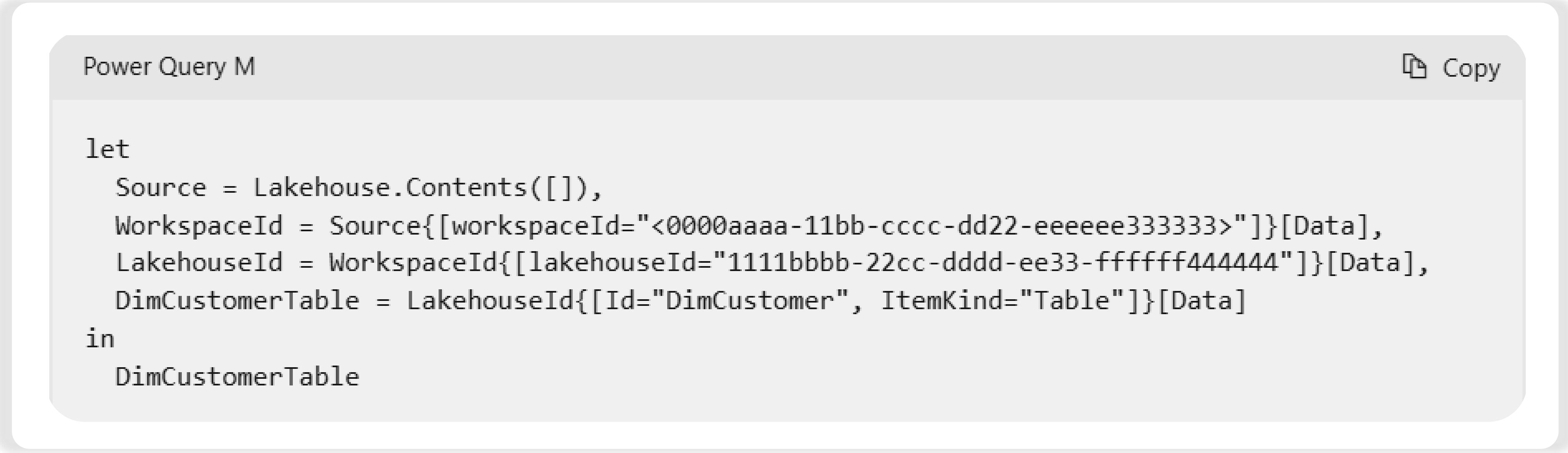
- **Problem Statement**
  - The dataflow creators want to take advantage of the advanced capabilities of Dataflow Gen2 for authoring purposes
  - At the same time, they plan to temporarily continue using dataflow tables as a data source during a phased migration
- **Solution**
  - Update the workspace ID, if a new workspace is created to store the new dataflow
  - Update existing solutions from the original (Gen1) dataflow ID to the new (Gen2) dataflow ID



```
let
    Source = PowerPlatform.Dataflows(null),
    Workspaces = Source{[Id="Workspaces"]}[Data],
    Workspace = Workspaces{[workspaceId="<enter new workspace ID>"]}[Data],
    DataflowId = Workspace{[dataflowId="<enter new dataflow ID"]}[Data],
    DimDateTable = DataflowId{[entity="DimDate", version=""]}[Data]
in
    DimDateTable
```

Migration scenario - Departmental usage

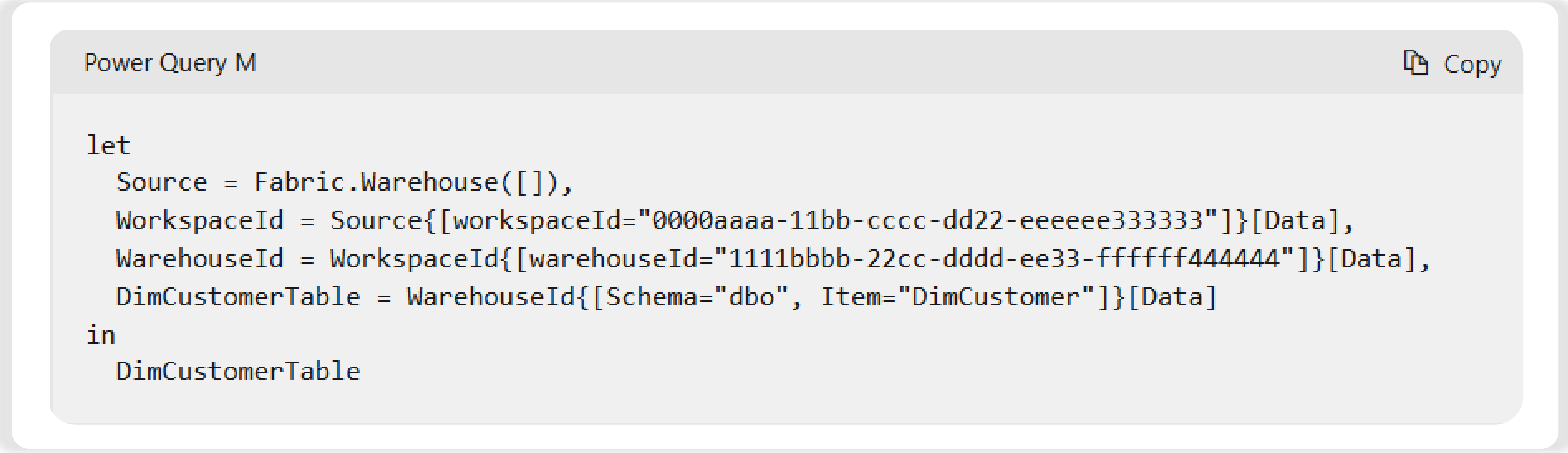
- **Problem Statement**
  - The dataflow creators want to take advantage of the advanced capabilities of Dataflow Gen2 for authoring
  - Efficiently sharing and outputting the dataflow tables to a Fabric lakehouse
  - This method takes advantage of OneLake shortcuts
- **Solution**
  - Replace linked tables with OneLake shortcuts, which provide downstream consumers with direct access to the data
  - Update existing solutions and transition queries by replacing the PowerPlatform.Dataflows or PowerBI.Dataflows functions with the Lakehouse.Contents data access function in Fabric



```
let
  Source = Lakehouse.Contents([]),
  WorkspaceId = Source{[workspaceId="<0000aaaa-11bb-cccc-dd22-eeeeee333333>"]}[Data],
  LakehouseId = WorkspaceId{[lakehouseId="1111bbbb-22cc-dddd-ee33-ffffff444444"]}[Data],
  DimCustomerTable = LakehouseId{[Id="DimCustomer", ItemKind="Table"]}[Data]
in
  DimCustomerTable
```

Migration scenario - Enterprise usage

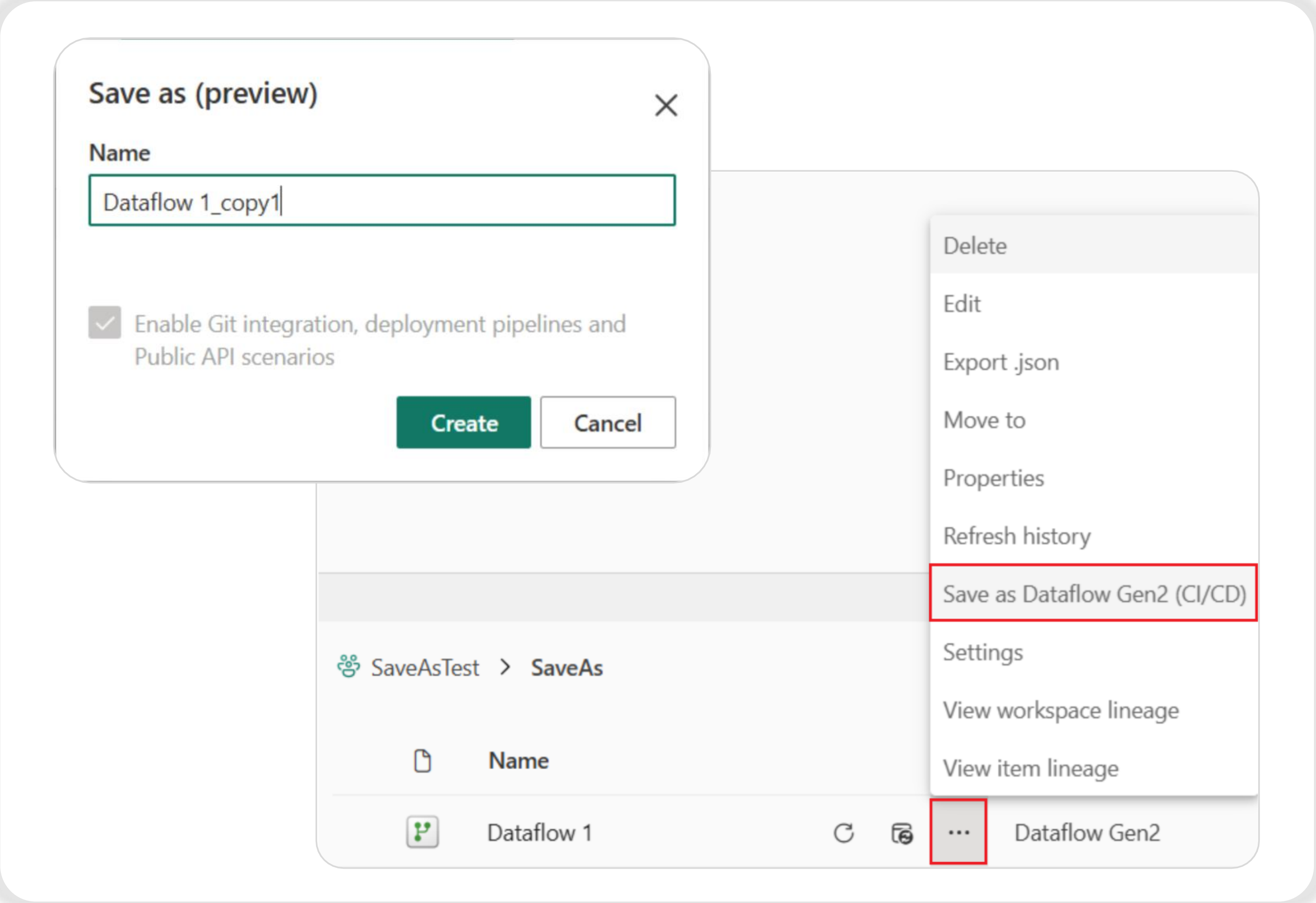
- **Problem Statement**
  - The dataflow creators want to take advantage of the advanced capabilities of Dataflow Gen2 for authoring
  - Outputting and sharing dataflow tables from a Fabric warehouse that has granular user permissions
- **Solution**
  - Grant data access through the SQL compute engine's granular permissions, which provide more selective access to certain users by restricting access to specific tables and schemas, as well as implementing RLS and CLS
  - Update existing solutions and transition queries by replacing the PowerPlatform.Dataflows or PowerBI.Dataflows function with the Fabric.Warehouse data access function in Fabric



```
let
  Source = Fabric.Warehouse([]),
  WorkspaceId = Source{[workspaceId="0000aaaa-11bb-cccc-dd22-eeeeee333333"]}[Data],
  WarehouseId = WorkspaceId{[warehouseId="1111bbbb-22cc-dddd-ee33-ffffff444444"]}[Data],
  DimCustomerTable = WarehouseId{[Schema="dbo", Item="DimCustomer"]}[Data]
in
  DimCustomerTable
```

Migrate to Dataflow Gen2 (CI/CD) using Save As (Preview)

- Save a Dataflow Gen2 or Gen2 (CI/CD) as a new Dataflow Gen2 (CI/CD)
- Save a Dataflow Gen1 as a new Dataflow Gen2 (CI/CD)



Save as (preview)

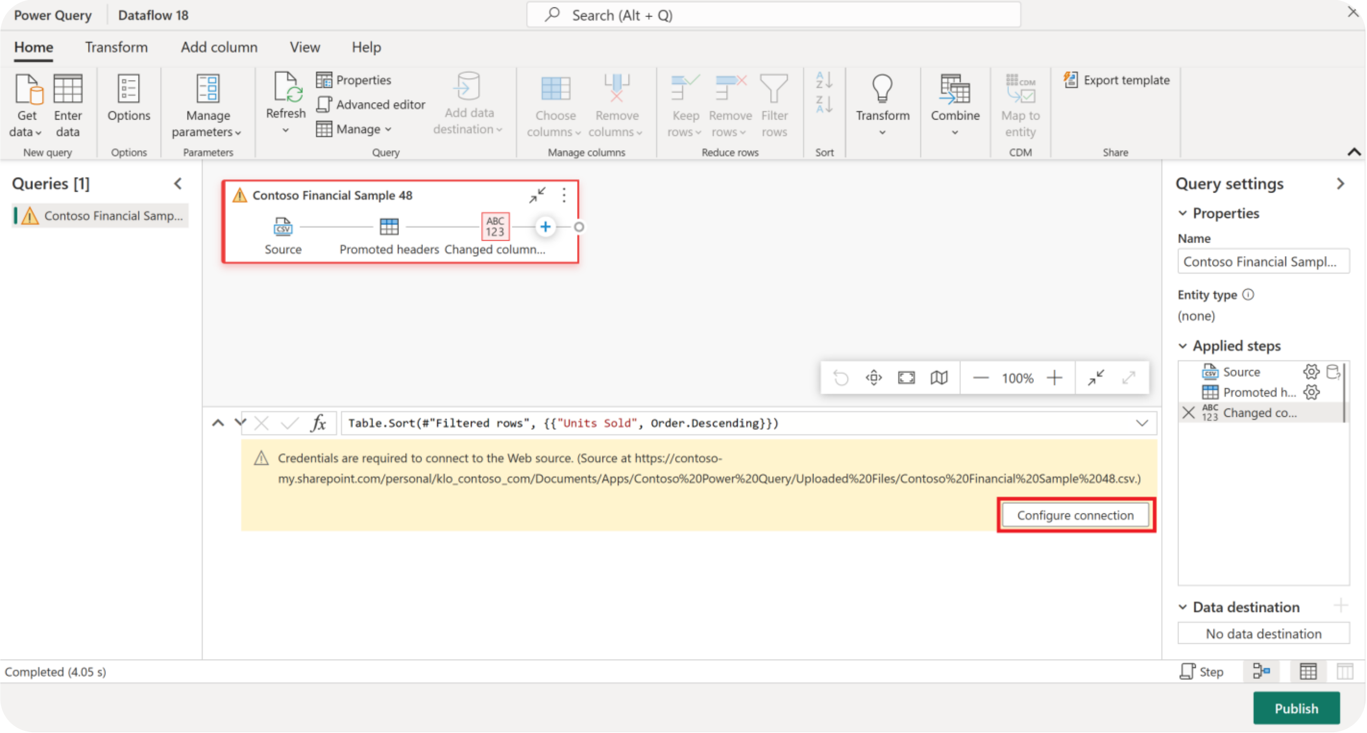
Name: Dataflow 1_copy1

☑ Enable Git integration, deployment pipelines and Public API scenarios

Move queries from Dataflow Gen1 to Dataflow Gen2

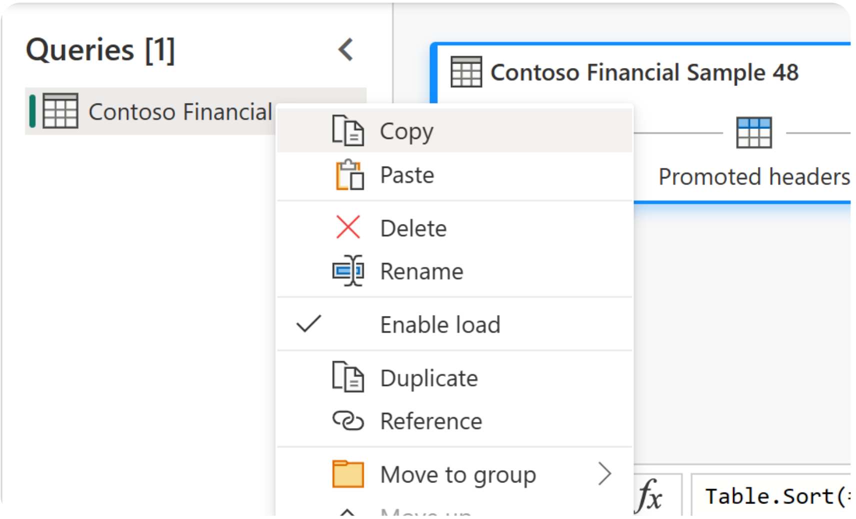
Move queries from Dataflow Gen1 to Dataflow Gen2

- Use the export template feature
- Copy and paste existing Dataflow Gen1 queries



```
Table.Sort(#"Filtered rows", {{"Units Sold", Order.Descending}})
```

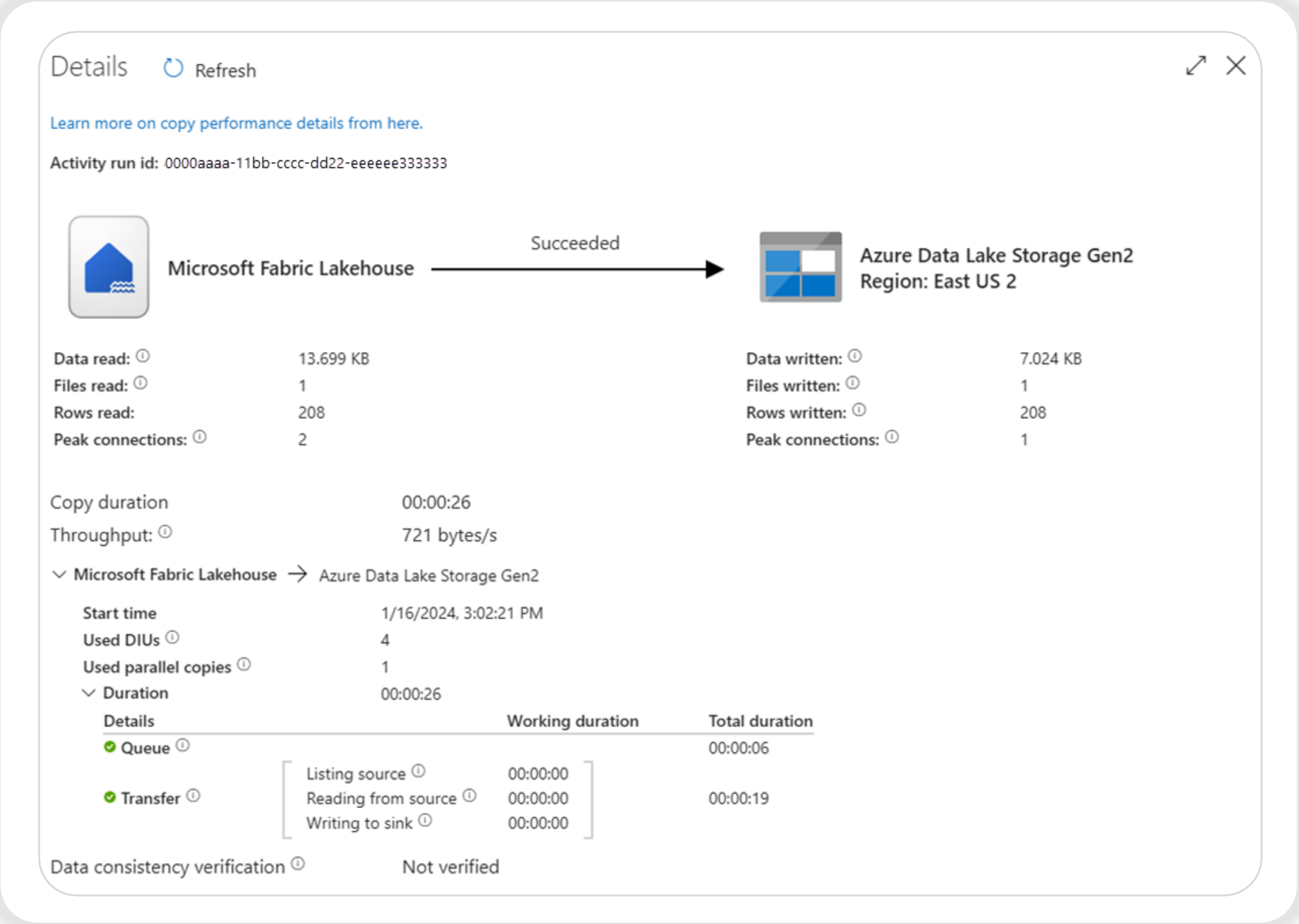
Credentials are required to connect to the Web source. (Source at https://contoso-my.sharepoint.com/personal/klo_contoso_com/Documents/Apps/Contoso%20Power%20Query/Uploaded%20Files/Contoso%20Financial%20Sample%2048.csv.)



Ingest data into Microsoft Fabric using the Azure Data Factory Copy Activity

Ingest data into Microsoft Fabric using the Azure Data Factory Copy Activity

- Set up authentication
- Write to a Fabric Lakehouse table with an ADF pipeline
- Read from a Fabric Lakehouse table with an ADF pipeline



Learn more on copy performance details from here.

Activity run id: 0000aaaa-11bb-cccc-dd22-eeeeee333333

Microsoft Fabric Lakehouse → Succeeded → Azure Data Lake Storage Gen2
Region: East US 2

Data read:	13.699 KB	Data written:	7.024 KB
Files read:	1	Files written:	1
Rows read:	208	Rows written:	208
Peak connections:	2	Peak connections:	1

Copy duration 00:00:26

Throughput: 721 bytes/s

Microsoft Fabric Lakehouse → Azure Data Lake Storage Gen2

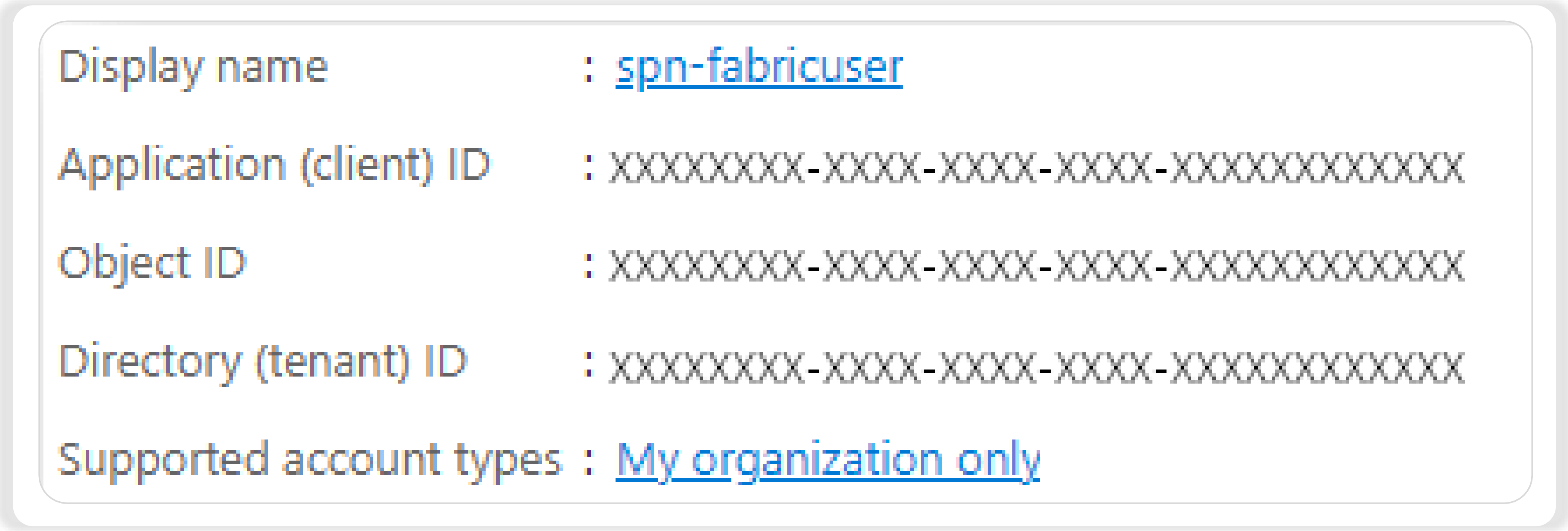
Start time	1/16/2024, 3:02:21 PM
Used DIUs	4
Used parallel copies	1
Duration	00:00:26

Details		Working duration	Total duration
✓ Queue			00:00:06
✓ Transfer	Listing source	00:00:00	00:00:19
	Reading from source	00:00:00	
	Writing to sink	00:00:00	

Data consistency verification Not verified

Set up authentication

Create or use an existing app registration service principal (SPN)



Display name : spn-fabricuser

Application (client) ID : XXXXXXXX-XXXX-XXXX-XXXX-XXXXXXXXXXXX

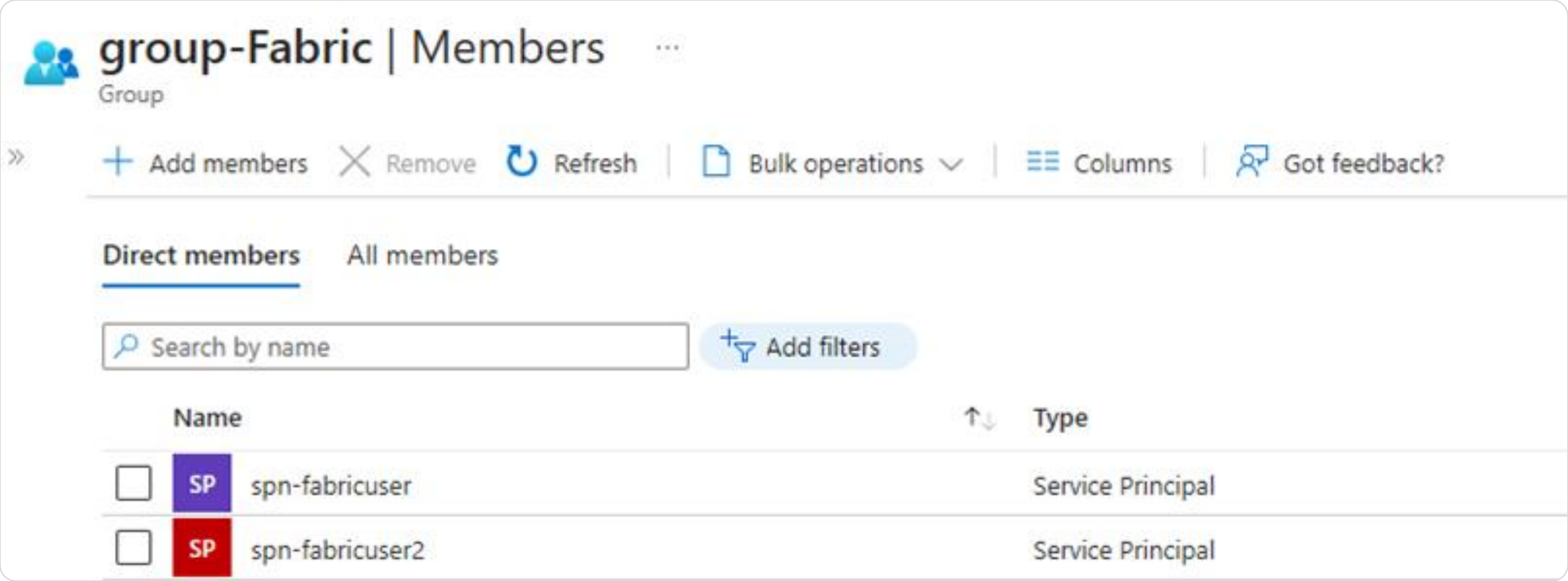
Object ID : XXXXXXXX-XXXX-XXXX-XXXX-XXXXXXXXXXXX

Directory (tenant) ID : XXXXXXXX-XXXX-XXXX-XXXX-XXXXXXXXXXXX

Supported account types : My organization only

Set up authentication

Create a new Microsoft Entra Security Group or use an existing one then add the SPN to it



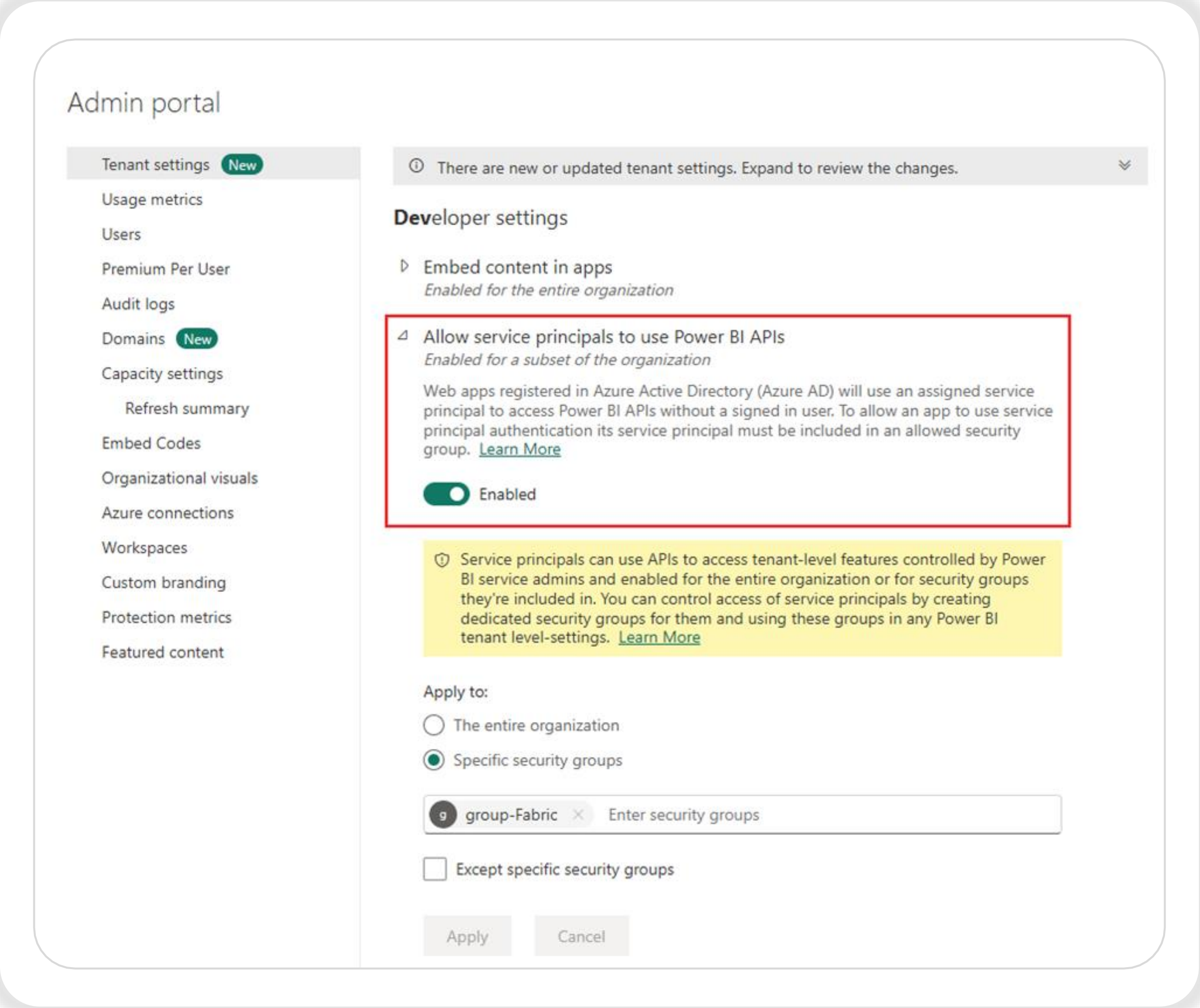
group-Fabric | Members

Group

	Name	Type
□	SP spn-fabricuser	Service Principal
□	SP spn-fabricuser2	Service Principal

Set up authentication

On Power BI admin portal, **allow service principals and** add the Security Group



Admin portal

There are new or updated tenant settings. Expand to review the changes.

Developer settings

Embed content in apps
Enabled for the entire organization

Allow service principals to use Power BI APIs
Enabled for a subset of the organization

Web apps registered in Azure Active Directory (Azure AD) will use an assigned service principal to access Power BI APIs without a signed in user. To allow an app to use service principal authentication its service principal must be included in an allowed security group. Learn More

Enabled

Service principals can use APIs to access tenant-level features controlled by Power BI service admins and enabled for the entire organization or for security groups they're included in. You can control access of service principals by creating dedicated security groups for them and using these groups in any Power BI tenant level-settings. Learn More

Apply to:

○ The entire organization

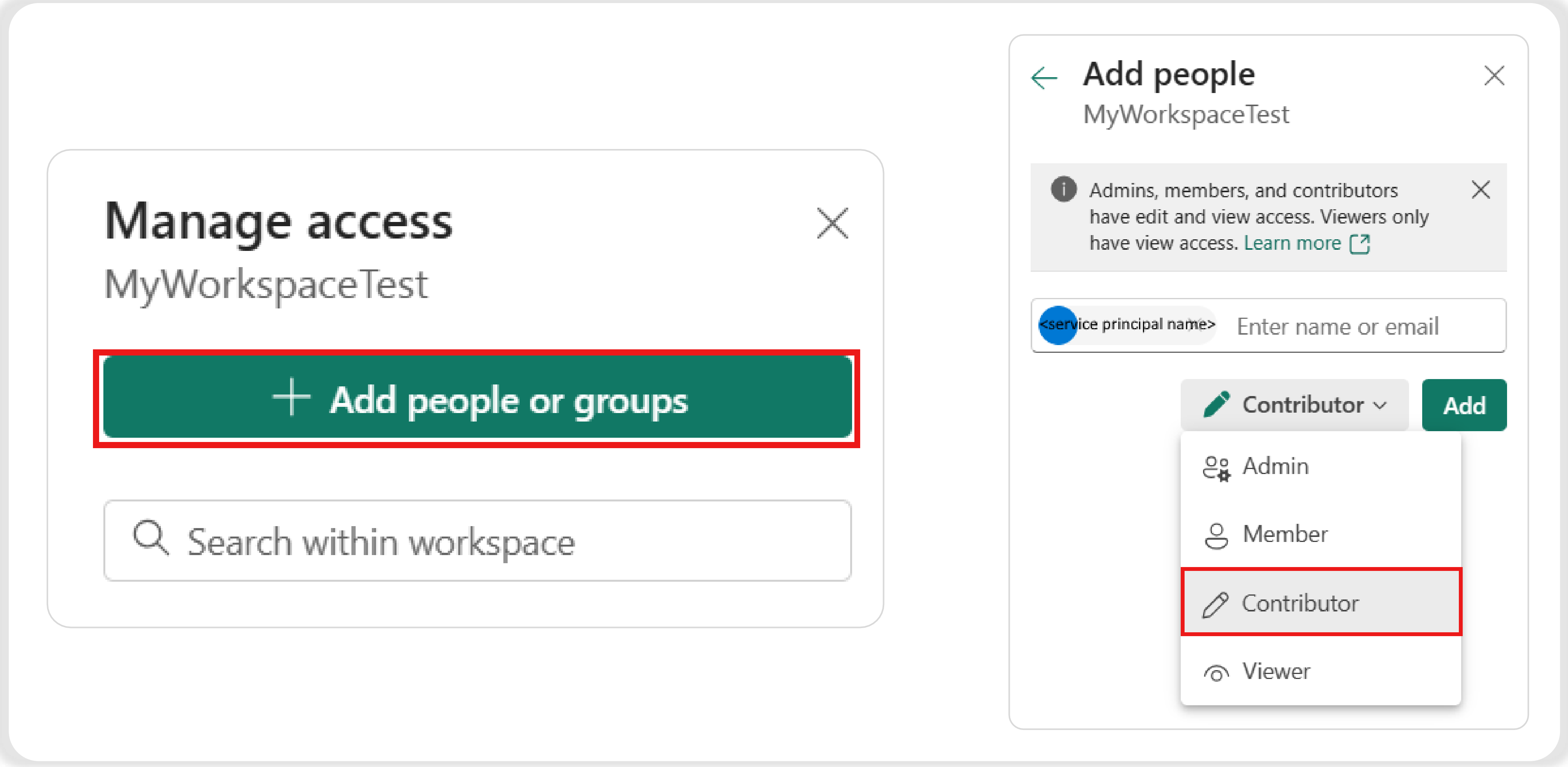
● Specific security groups

group-Fabric × Enter security groups

□ Except specific security groups

Set up authentication

Add the SPN or the service group to the workspace



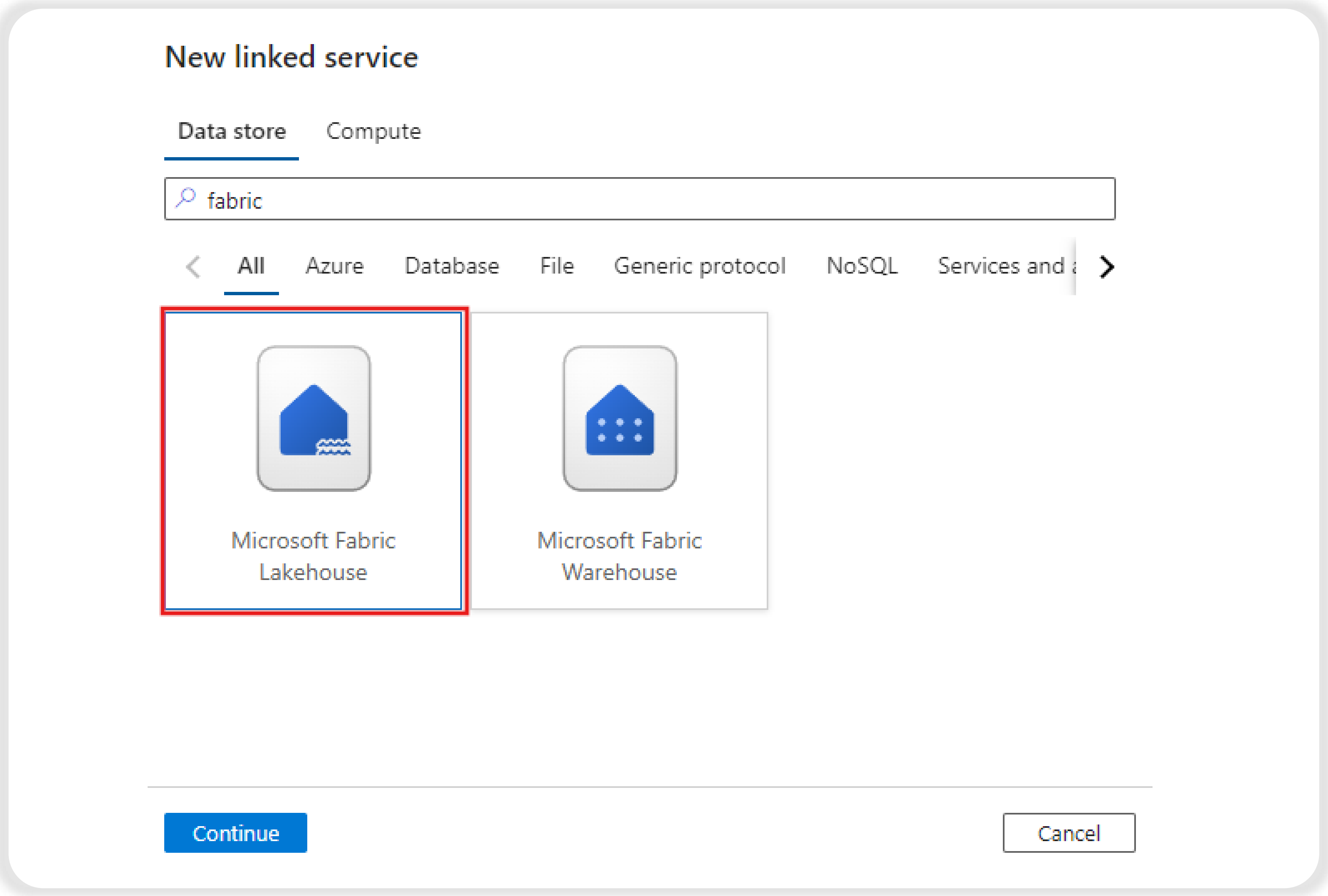
Admins, members, and contributors have edit and view access. Viewers only have view access. Learn more

<service principal name> Enter name or email

Contributor

Set up authentication

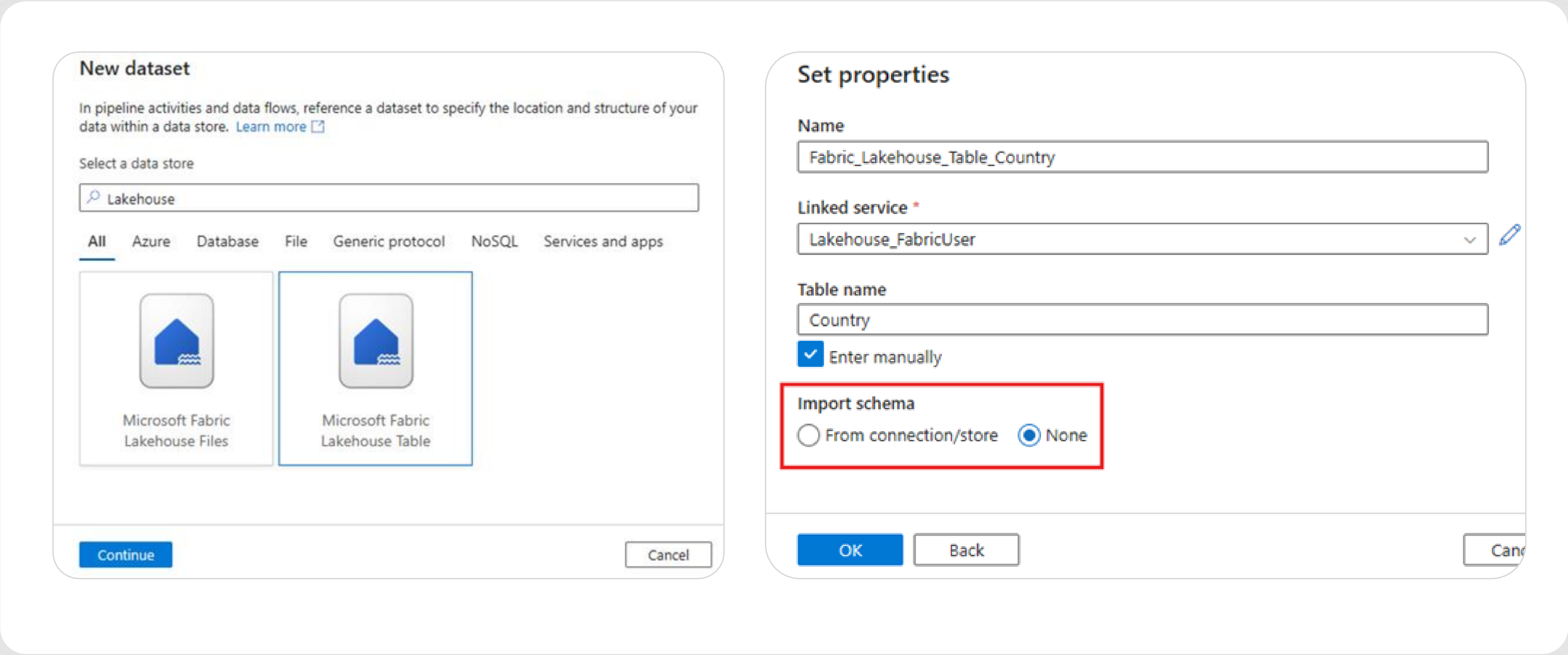
From Azure Data Factory, create a new Microsoft Fabric Lakehouse linked service



New linked service

Set up authentication

Create a Dataset that references the Microsoft Fabric Lakehouse linked service



New dataset

In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. Learn more

Select a data store

Lakehouse

Set properties

Name

Fabric_Lakehouse_Table_Country

Linked service *

Lakehouse_FabricUser

Table name

Country

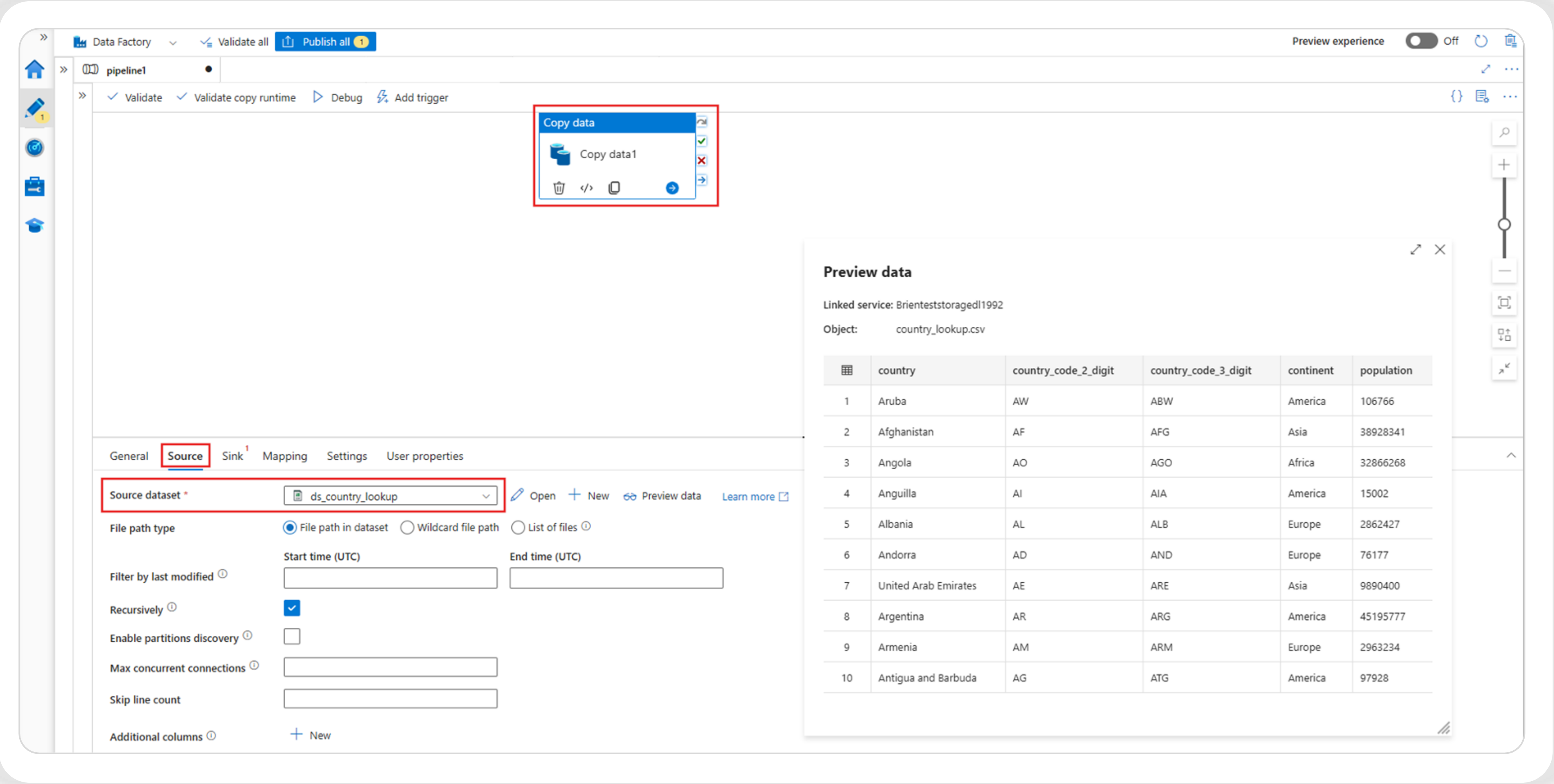
☑ Enter manually

Import schema

○ From connection/store ● None

Write to a Fabric Lakehouse table with an ADF pipeline

Create a new pipeline and add a Copy activity to the pipeline canvas



General | Source | Sink 1 | Mapping | Settings | User properties

Source dataset * ds_country_lookup — Open · New · Preview data · Learn more

File path type: ● File path in dataset ○ Wildcard file path ○ List of files

Filter by last modified — Start time (UTC): — End time (UTC):

Recursively ☑

Enable partitions discovery ☐

Max concurrent connections

Skip line count

Additional columns + New

Preview data

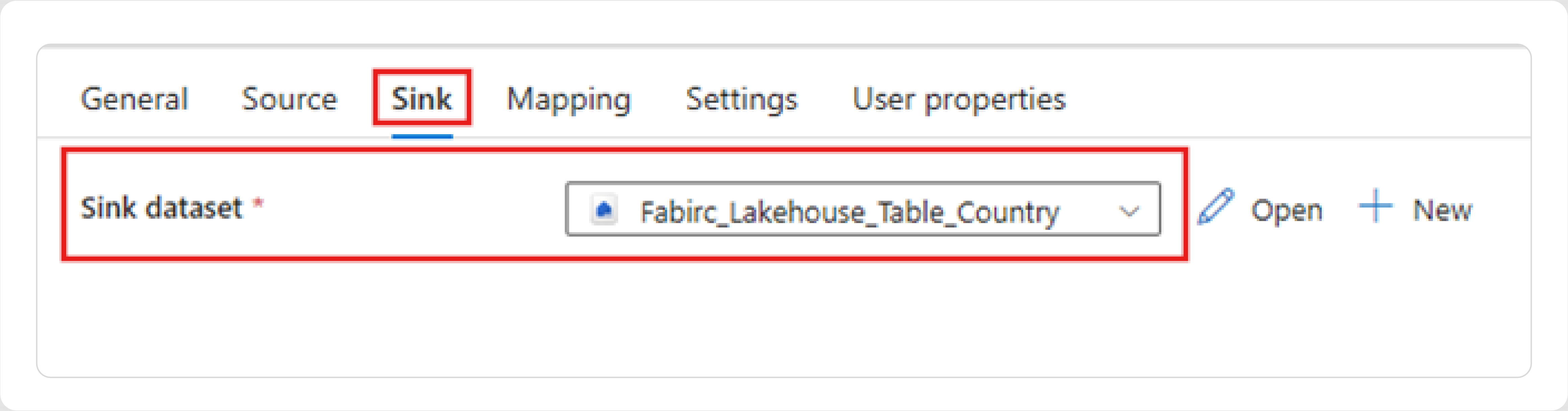
Linked service: Brienteststoragedl1992

Object: country_lookup.csv

	country	country_code_2_digit	country_code_3_digit	continent	population
1	Aruba	AW	ABW	America	106766
2	Afghanistan	AF	AFG	Asia	38928341
3	Angola	AO	AGO	Africa	32866268
4	Anguilla	AI	AIA	America	15002
5	Albania	AL	ALB	Europe	2862427
6	Andorra	AD	AND	Europe	76177
7	United Arab Emirates	AE	ARE	Asia	9890400
8	Argentina	AR	ARG	America	45195777
9	Armenia	AM	ARM	Europe	2963234
10	Antigua and Barbuda	AG	ATG	America	97928

Write to a Fabric Lakehouse table with an ADF pipeline

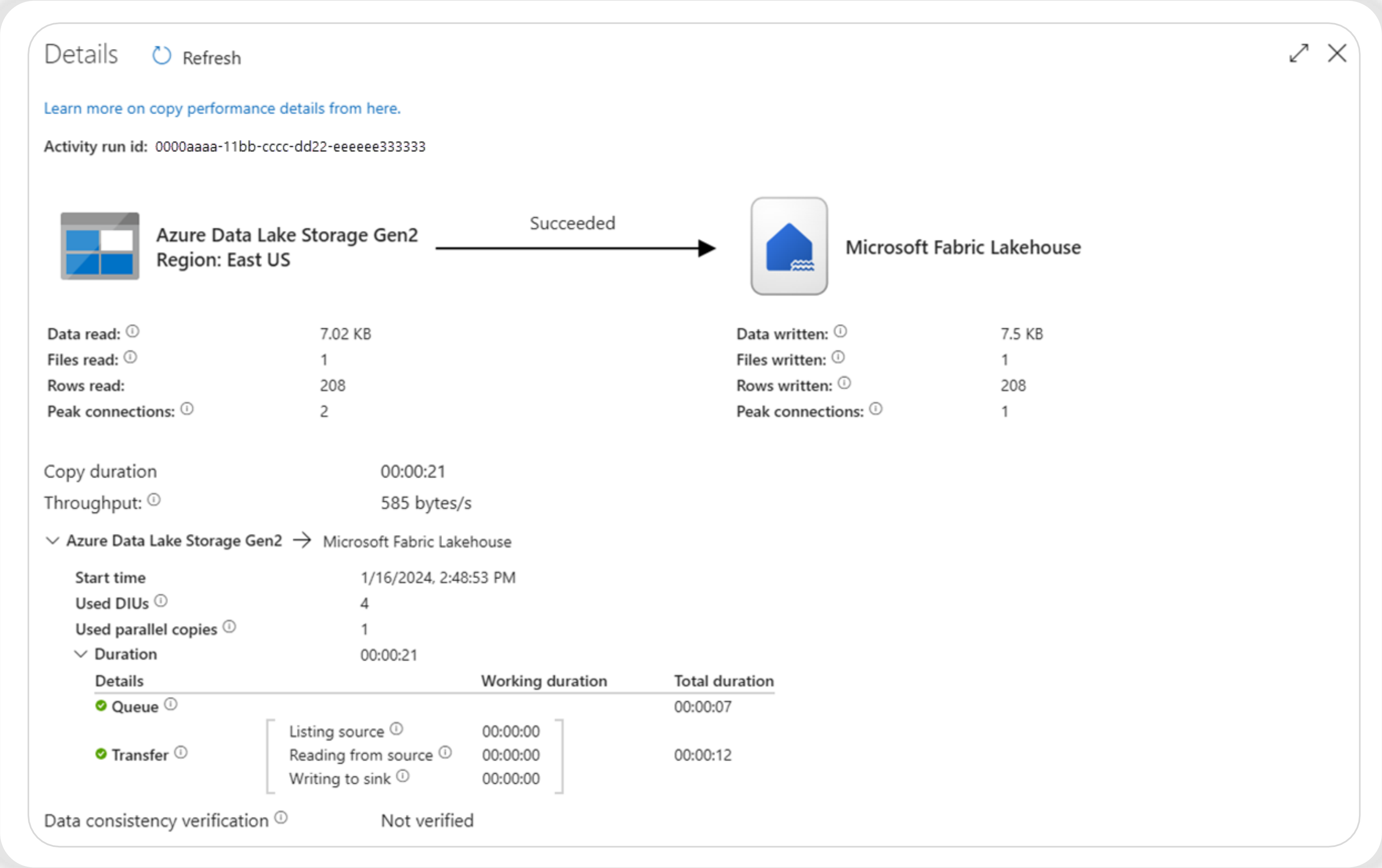
From the Copy activity's Sink tab and select the Fabric Lakehouse dataset



Sink dataset * Fabirc_Lakehouse_Table_Country

Write to a Fabric Lakehouse table with an ADF pipeline

Run the pipeline



Learn more on copy performance details from here.

Activity run id: 0000aaaa-11bb-cccc-dd22-eeeeee333333

Azure Data Lake Storage Gen2
Region: East US

Microsoft Fabric Lakehouse

Data read:	7.02 KB
Files read:	1
Rows read:	208
Peak connections:	2

Data written:	7.5 KB
Files written:	1
Rows written:	208
Peak connections:	1

Copy duration	00:00:21
Throughput:	585 bytes/s

Azure Data Lake Storage Gen2 → Microsoft Fabric Lakehouse

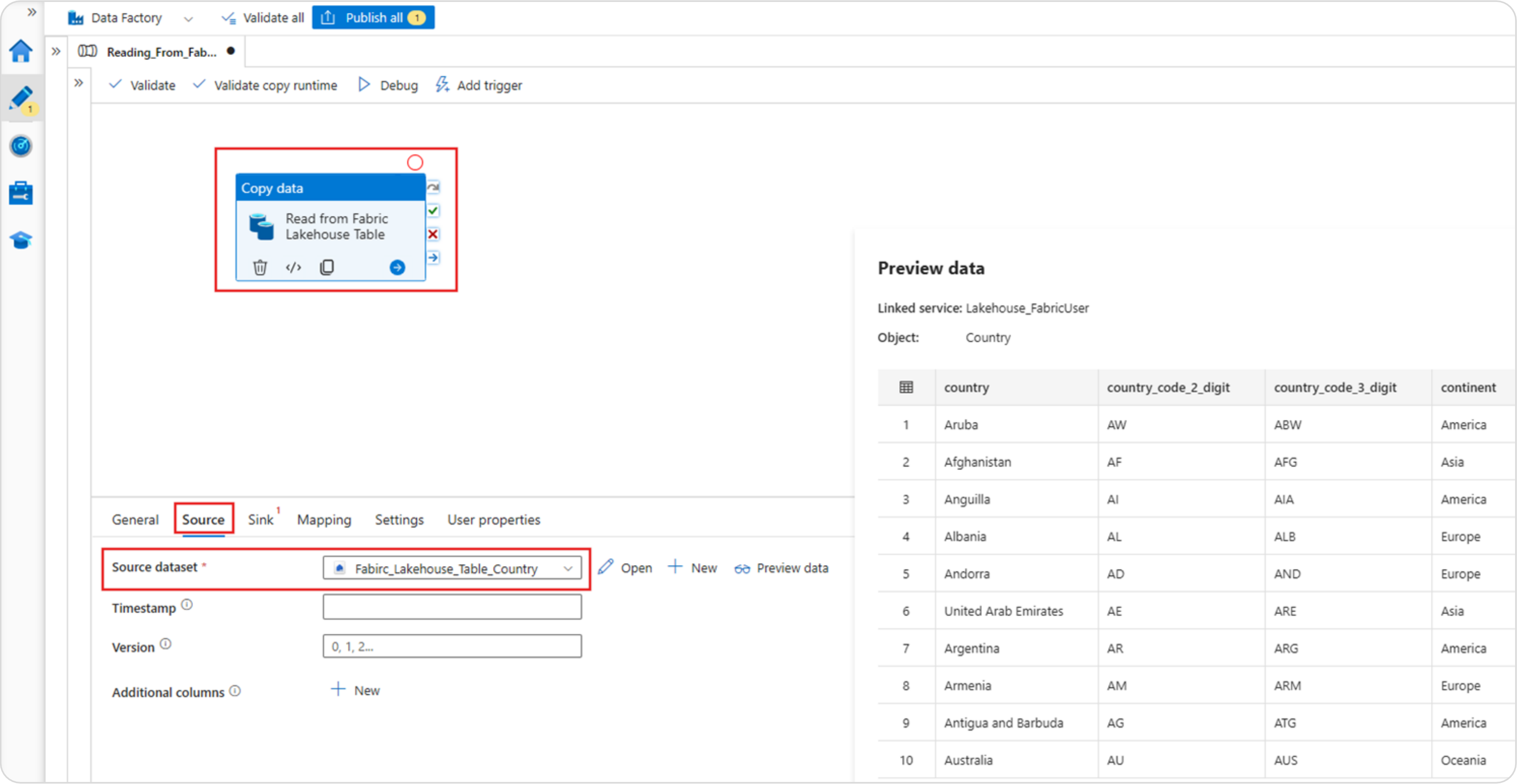
Start time	1/16/2024, 2:48:53 PM
Used DIUs	4
Used parallel copies	1
Duration	00:00:21

Details		Working duration	Total duration
Queue			00:00:07
Transfer	Listing source	00:00:00	00:00:12
	Reading from source	00:00:00	
	Writing to sink	00:00:00	

Data consistency verification   Not verified

Read from a Fabric Lakehouse table with an ADF pipeline

Create a new pipeline and add a Copy activity to the pipeline canvas



Source dataset * Fabirc_Lakehouse_Table_Country

Timestamp

Version 0, 1, 2...

Additional columns

Preview data

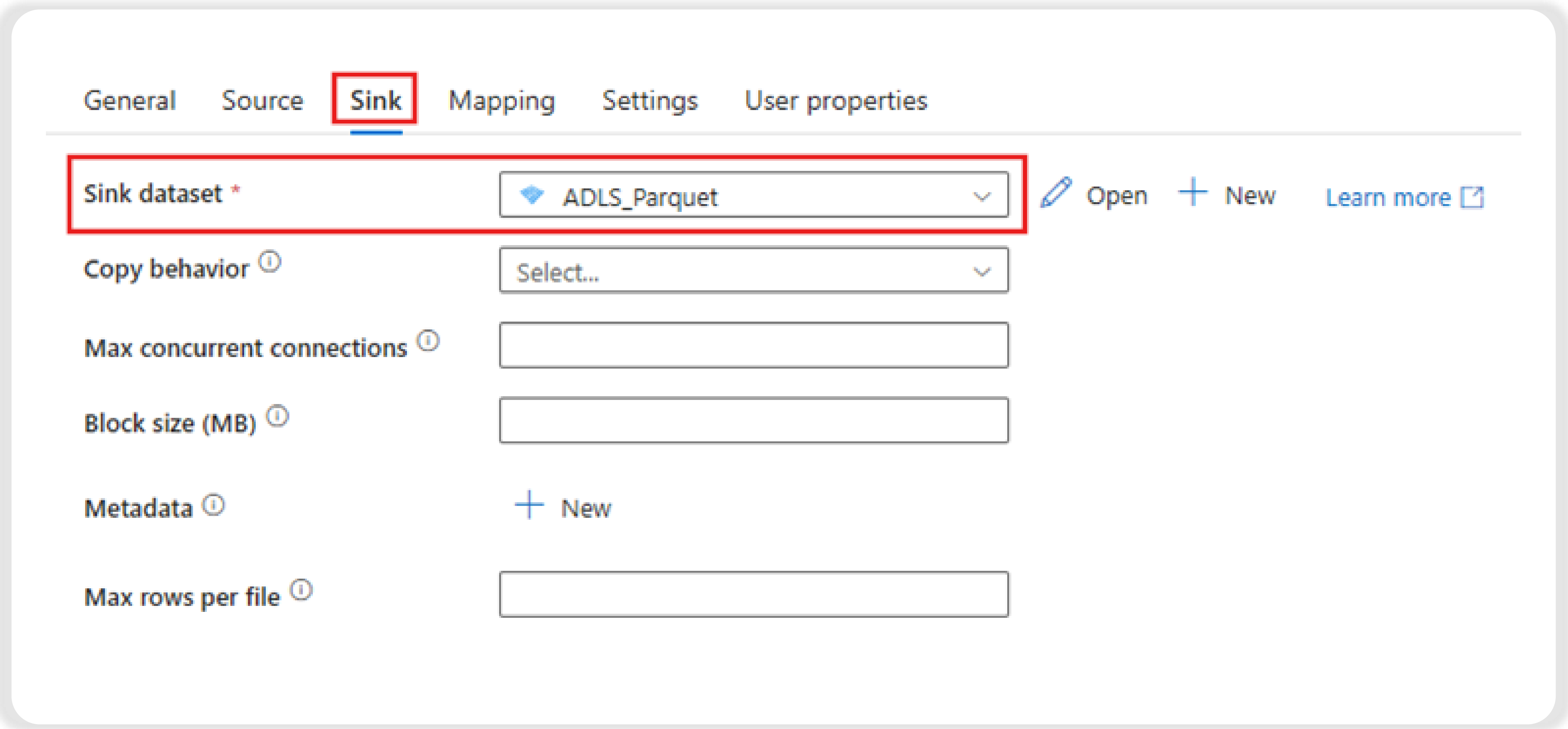
Linked service: Lakehouse_FabricUser

Object: Country

	country	country_code_2_digit	country_code_3_digit	continent
1	Aruba	AW	ABW	America
2	Afghanistan	AF	AFG	Asia
3	Anguilla	AI	AIA	America
4	Albania	AL	ALB	Europe
5	Andorra	AD	AND	Europe
6	United Arab Emirates	AE	ARE	Asia
7	Argentina	AR	ARG	America
8	Armenia	AM	ARM	Europe
9	Antigua and Barbuda	AG	ATG	America
10	Australia	AU	AUS	Oceania

Read from a Fabric Lakehouse table with an ADF pipeline

Copy activity's Sink tab and select the destination dataset



Sink dataset * ADLS_Parquet

Copy behavior: Select...

Max concurrent connections:

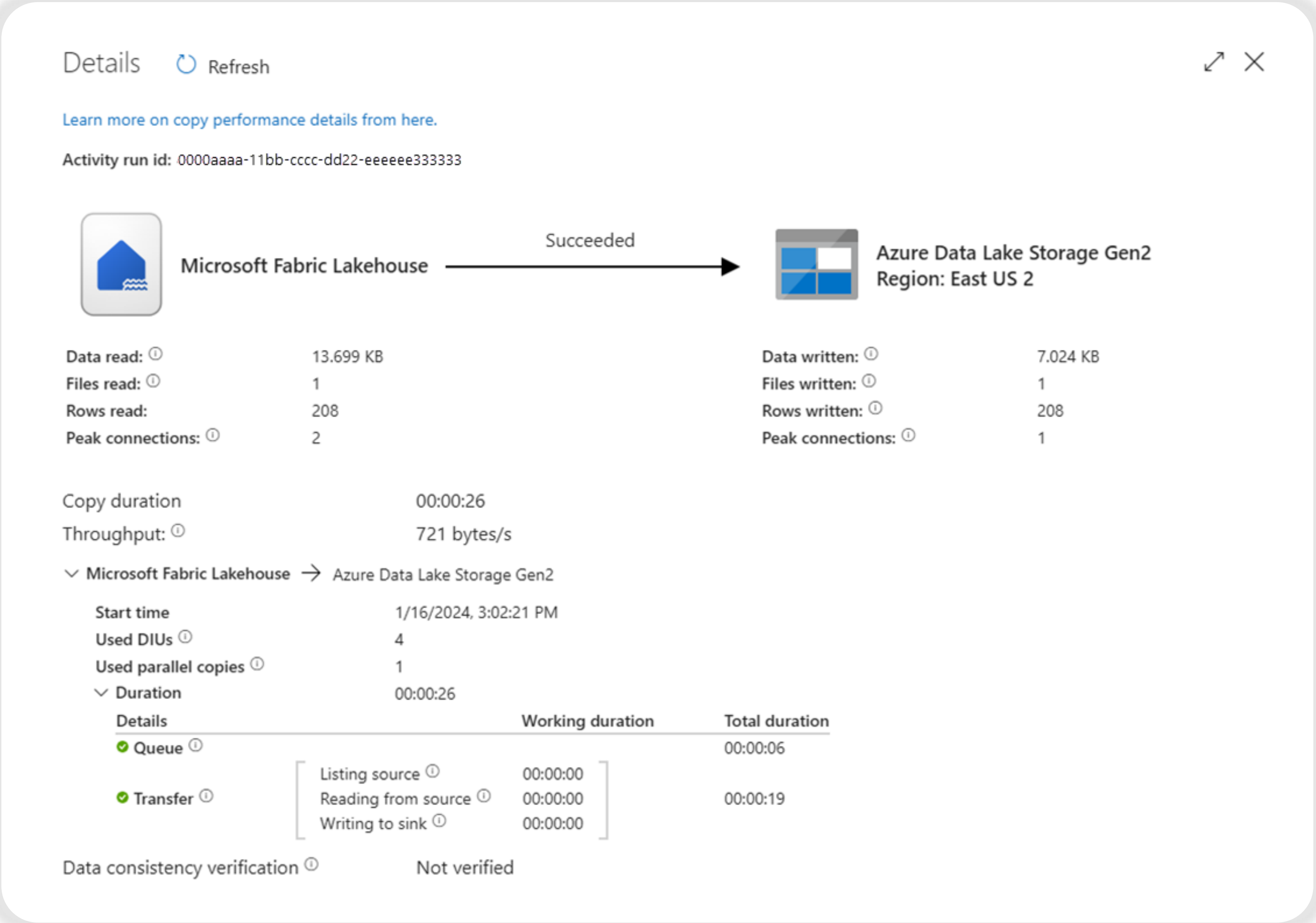
Block size (MB):

Metadata: + New

Max rows per file:

Read from a Fabric Lakehouse table with an ADF pipeline

Run the pipeline



Details Refresh

Learn more on copy performance details from here.

Activity run id: 0000aaaa-11bb-cccc-dd22-eeeeee333333

Microsoft Fabric Lakehouse → Succeeded → Azure Data Lake Storage Gen2
Region: East US 2

Data read:	13.699 KB
Files read:	1
Rows read:	208
Peak connections:	2

Data written:	7.024 KB
Files written:	1
Rows written:	208
Peak connections:	1

Copy duration	00:00:26
Throughput:	721 bytes/s

Microsoft Fabric Lakehouse → Azure Data Lake Storage Gen2

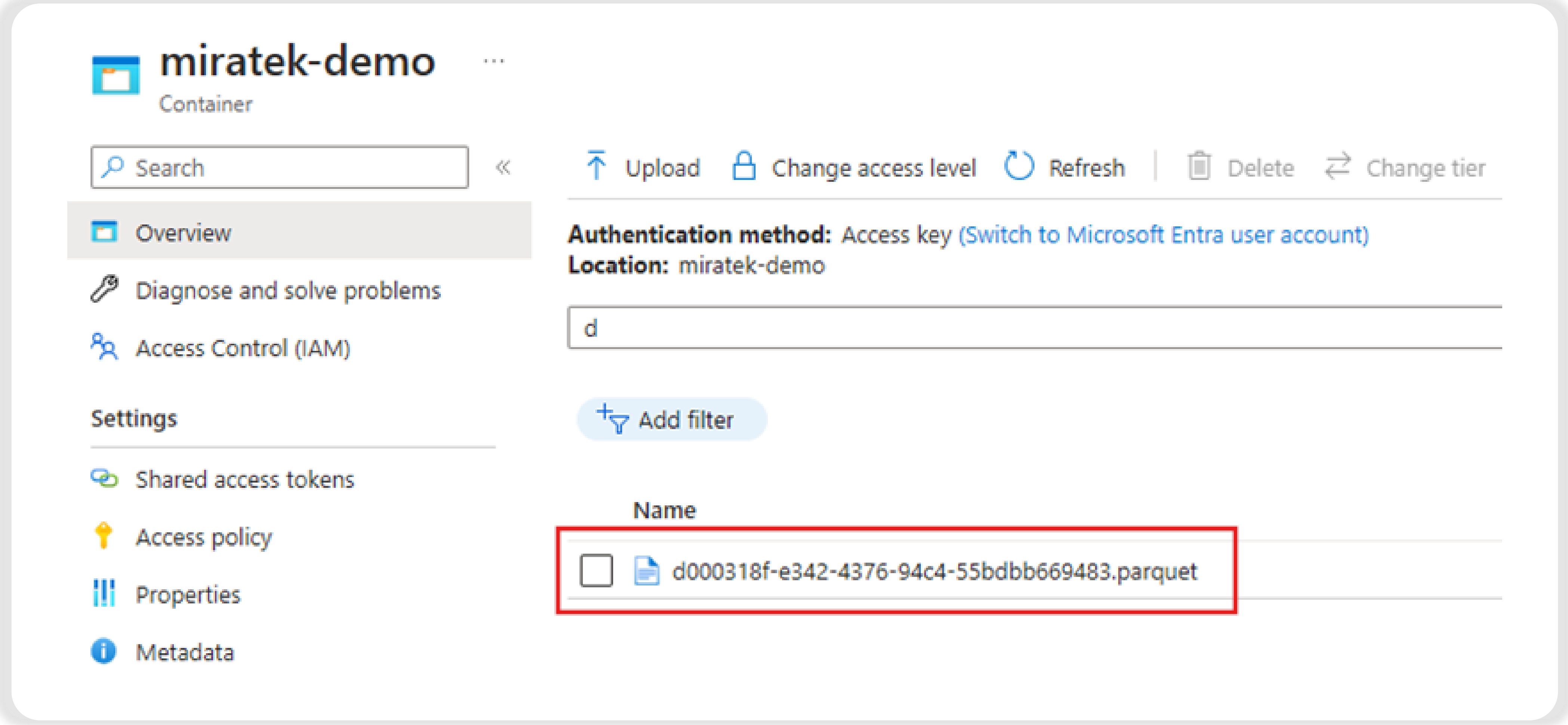
Start time	1/16/2024, 3:02:21 PM
Used DIUs	4
Used parallel copies	1
Duration	00:00:26

Details		Working duration	Total duration
✓ Queue			00:00:06
✓ Transfer	Listing source	00:00:00	00:00:19
	Reading from source	00:00:00	
	Writing to sink	00:00:00	

Data consistency verification	Not verified

Read from a Fabric Lakehouse table with an ADF pipeline

Inspect the Parquet file in ADLS Gen2



Authentication method: Access key (Switch to Microsoft Entra user account)
Location: miratek-demo

d

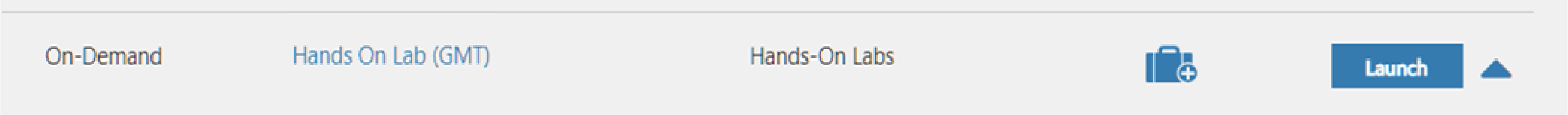
□ d000318f-e342-4376-94c4-55bdbb669483.parquet

Hands-on labs

Hands-on Lab

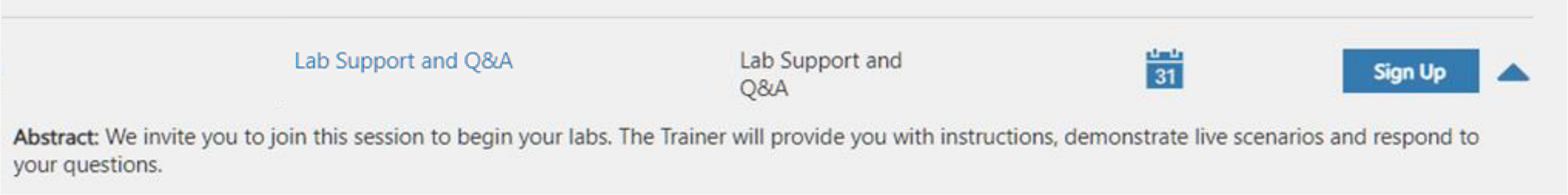
⚠ Please note that the labs are only open to a limited number of Microsoft partner participants and are offered on a **first-come-first-served basis**.

- **Step 1** : Click on **Launch** in the event curriculum to launch your labs



On-Demand Hands On Lab (GMT) Hands-On Labs

- **Step 2 :** Join the **Live session for Lab Support and Q&A Session**



Lab Support and Q&A Lab Support and Q&A

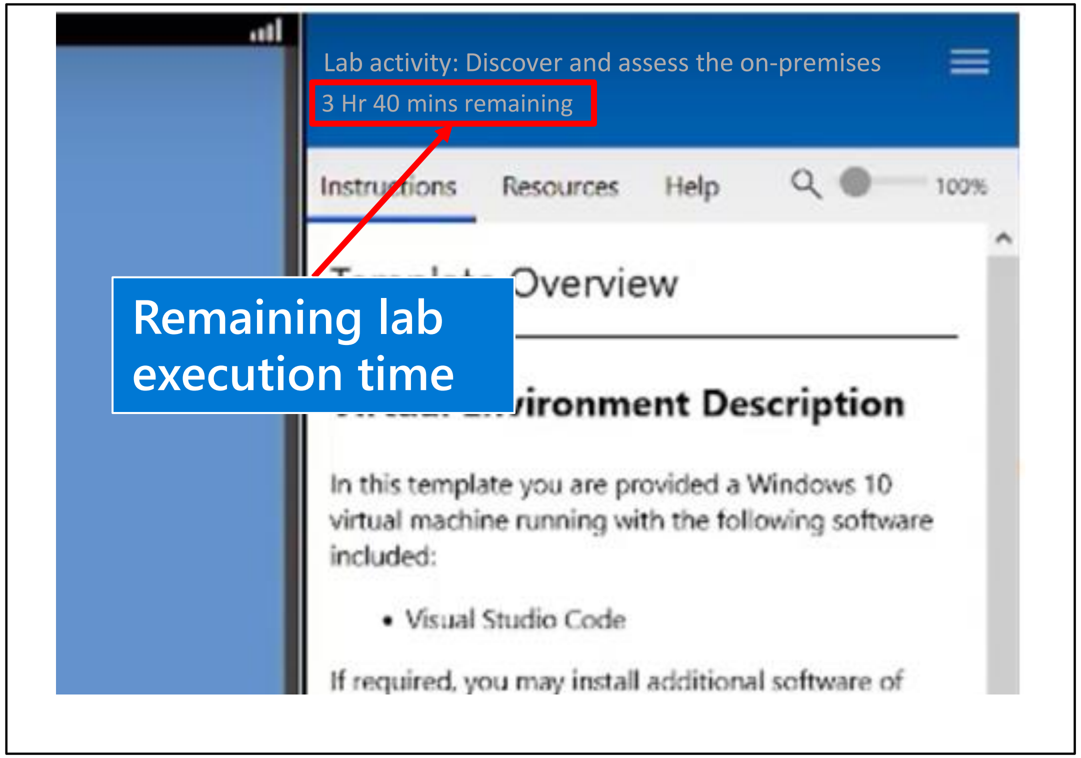
Abstract: We invite you to join this session to begin your labs. The Trainer will provide you with instructions, demonstrate live scenarios and respond to your questions.

**For all Lab logistical support issues, please go to:* [*https://www.skillable.com/customer-support/*](https://www.skillable.com/customer-support/) *and select the "Open a Support Ticket"*

Important lab information

Track remaining lab execution time

- Once launched, a lab is available for active execution for the time displayed in the screenshot.



Lab activity: Discover and assess the on-premises

3 Hr 40 mins remaining

Overview

ironment Description

In this template you are provided a Windows 10 virtual machine running with the following software included:

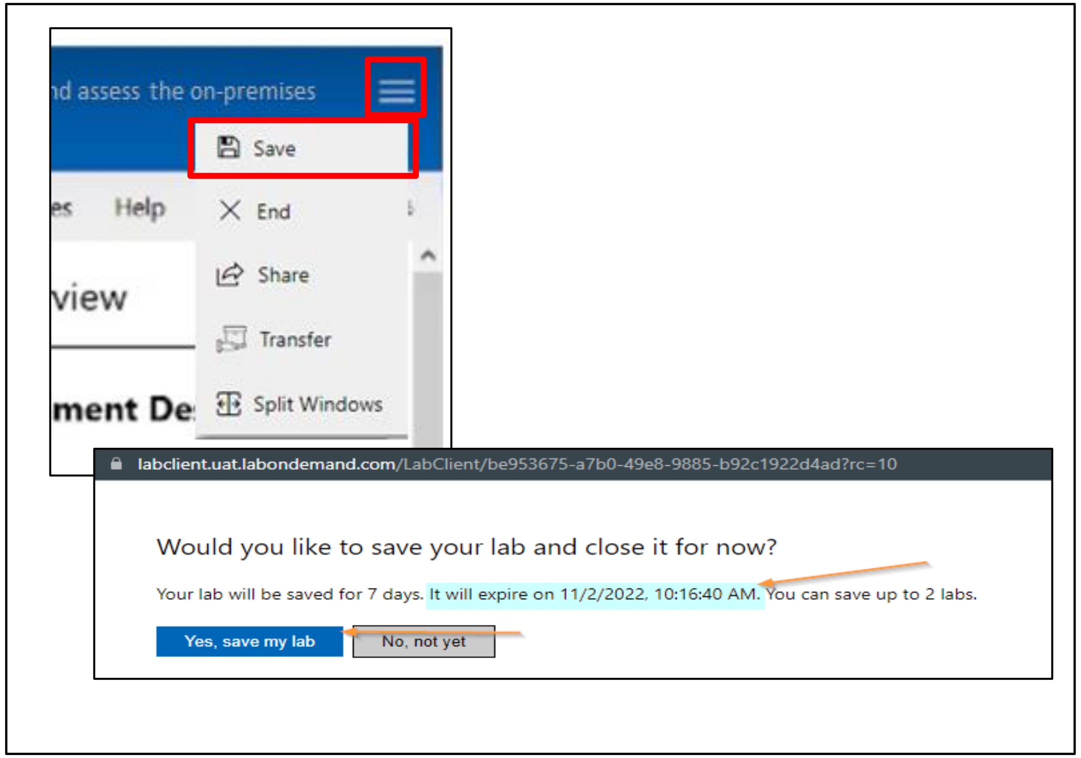
- Visual Studio Code

If required, you may install additional software of

Remaining lab execution time

Save your Lab

- Ensure that you SAVE the lab after each break.
- Do not fail to relaunch within the specified time to avoid losing saved labs.



labclient.uat.labondemand.com/LabClient/be953675-a7b0-49e8-9885-b92c1922d4ad?rc=10

Would you like to save your lab and close it for now?

Your lab will be saved for 7 days. It will expire on 11/2/2022, 10:16:40 AM. You can save up to 2 labs.

Once the allocated lab time expires, it can not be extended.

Hands-on labs

- **Lab 02**

  Use Case 02: Data Factory solution for moving and transforming data with dataflows and data pipelines

Microsoft

Technofocus

Thank you