



Microsoft Partner Project Ready

# Data Integration and Transformation with Data Factory in Microsoft Fabric

Day 01 of 02



# 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

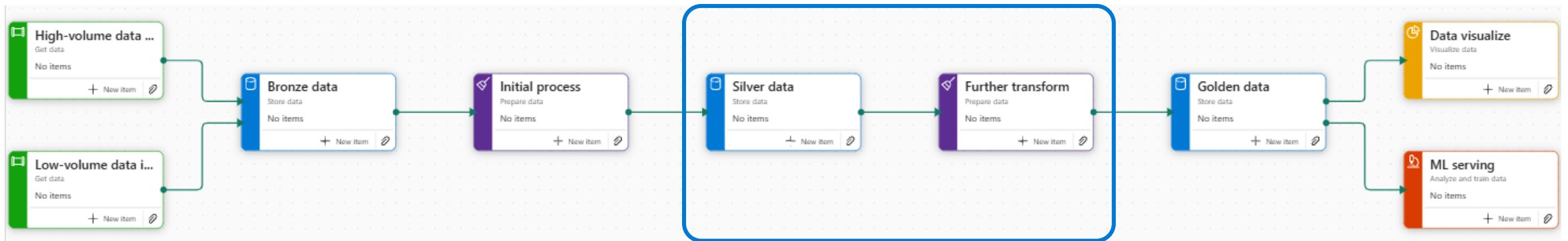
- Use Case 02: Data Factory solution for moving and transforming data with dataflows and data pipelines



03

# Data Transformation with Dataflows Gen2

# Medallion Architecture in Microsoft Fabric Data Factory



# Dataflows Gen2 in Microsoft Fabric



# Create a dataflow

- Create a dataflow from your Microsoft Fabric workspace
- Get data
- Apply transformations
- Configure destination settings for transformed data

## New item

Select an item type

☆ Favorites **📄 All items**

🔍 Filter by item type

### Get data

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

**Copy job (preview)**  
Makes it easy to copy data in Fabric. Includes full copy, incremental copy, and event-based copy modes.  
☆

**Data pipeline**  
Ingest data at scale and schedule data workflows.  
☆

**Dataflow Gen1**  
Prep, clean, and transform data.  
☆

**Dataflow Gen2**  
Prep, clean, and transform data.  
☆

**Spark Job Definition**  
Define, schedule, and manage your Apache Spark jobs for big data processing.  
☆

**Eventstream**  
Capture, transform, and route real-time event stream to various destinations in desired format with no-code experience.  
☆

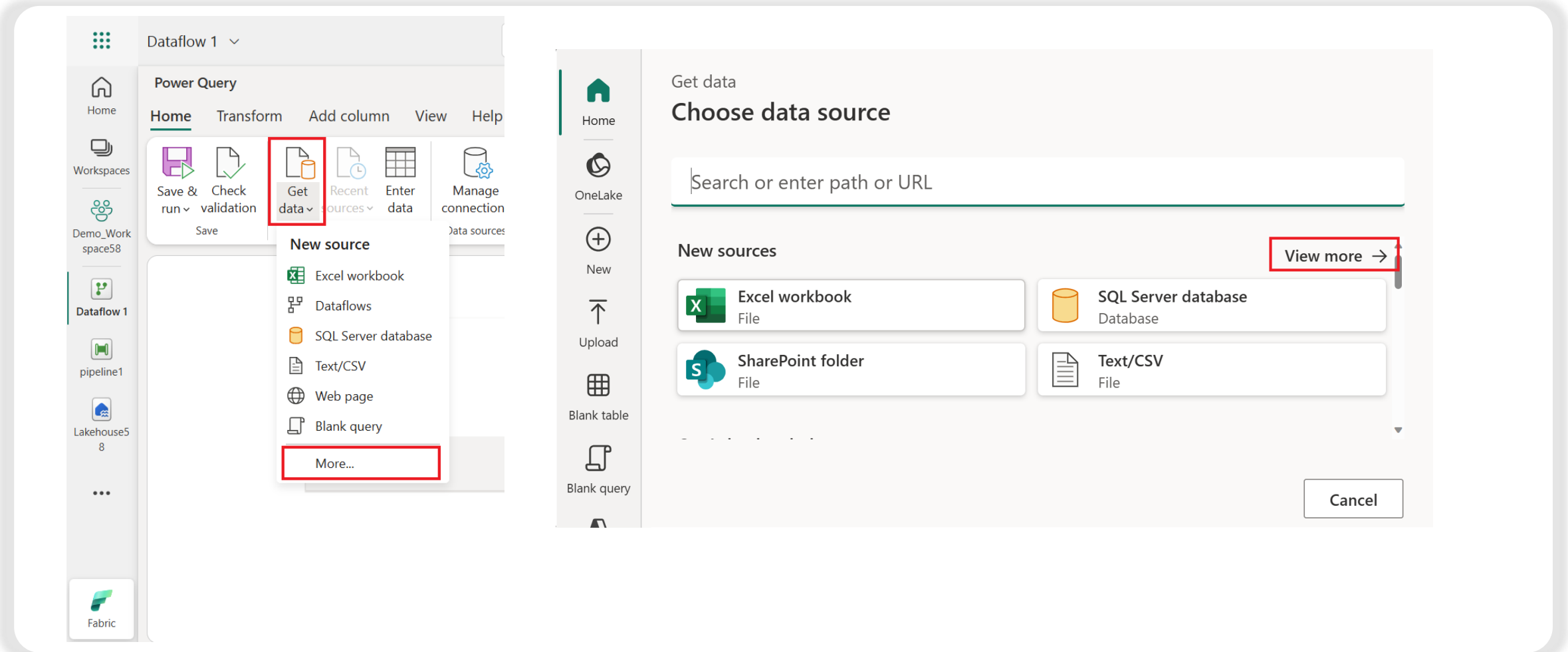
**Mirrored Azure Cosmos DB (p...**  
Easily replicate data from an existing source into an analytics-friendly format.  
☆

**Mirrored Azure Databricks cat...**  
Explore Unity Catalog Tables  
☆

**Mirrored Azure SQL Database ...**  
Easily replicate data from an existing source into an analytics-friendly format.  
☆

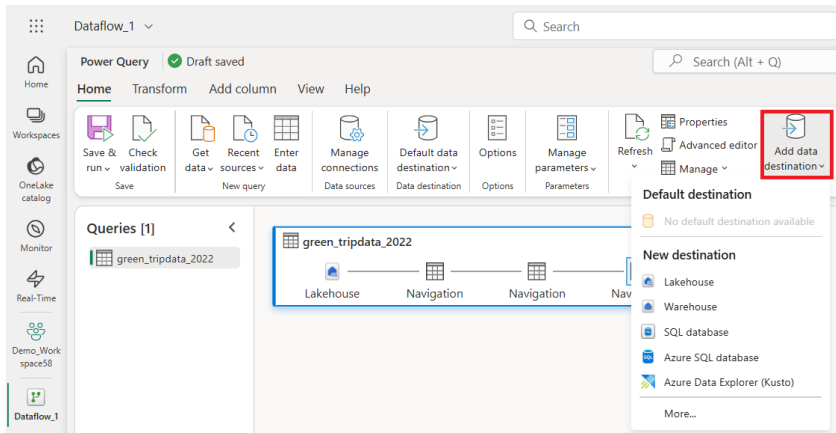
# Get data for your dataflow

Dataflows (Gen2) connect to various data sources and perform transformations in Power Query Online

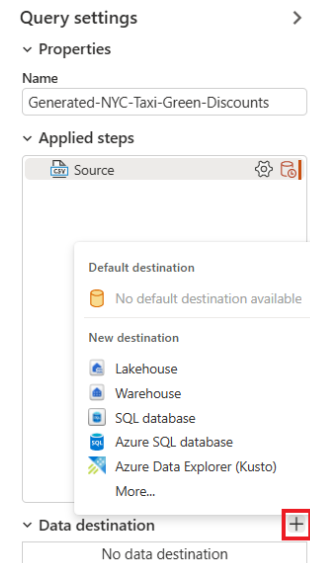


# Entry points to specify the data destination

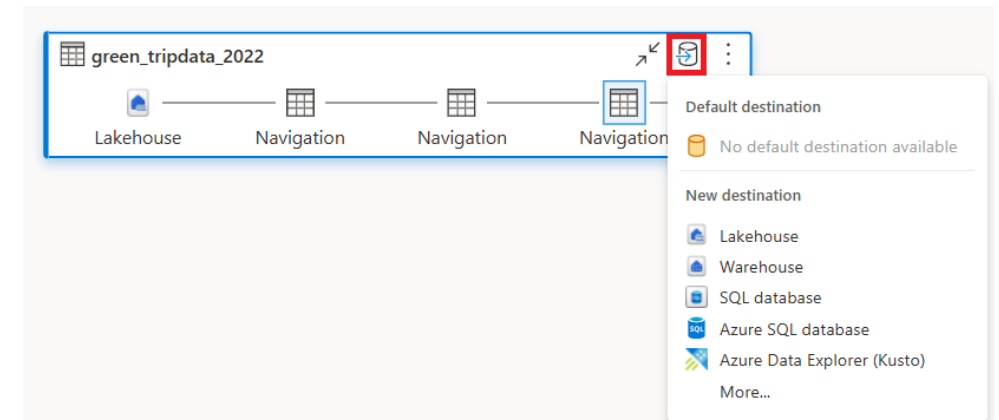
To specify the data destination



Through the top ribbon



Through query settings

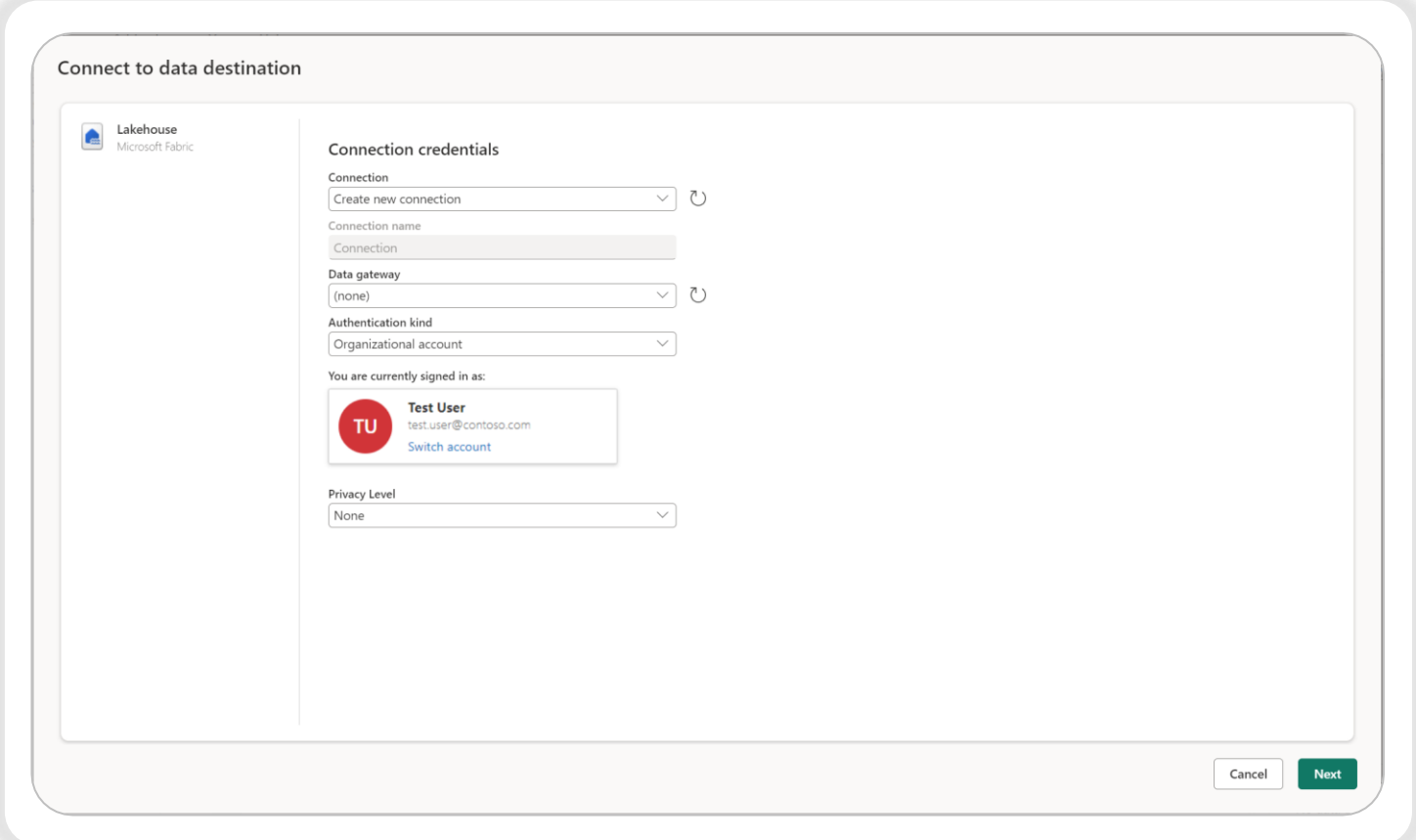


Through the diagram view



# Connect to the data destination

- Connecting to the data destination is similar to connecting to a data source
- Connections can be used for both reading and writing your data



The screenshot shows a 'Connect to data destination' dialog box. On the left, there is a Lakehouse icon and the text 'Lakehouse Microsoft Fabric'. The main area is titled 'Connection credentials' and contains several fields: 'Connection' (a dropdown menu with 'Create new connection' selected), 'Connection name' (a text input field with 'Connection' entered), 'Data gateway' (a dropdown menu with '(none)' selected), and 'Authentication kind' (a dropdown menu with 'Organizational account' selected). Below these fields, it says 'You are currently signed in as:' followed by a user card for 'Test User' (test.user@contoso.com) with a 'Switch account' link. At the bottom, there is a 'Privacy Level' dropdown menu set to 'None'. In the bottom right corner, there are 'Cancel' and 'Next' buttons.

Connect to data destination

Lakehouse  
Microsoft Fabric

**Connection credentials**

Connection  
Create new connection

Connection name  
Connection

Data gateway  
(none)

Authentication kind  
Organizational account

You are currently signed in as:

**Test User**  
test.user@contoso.com  
[Switch account](#)

Privacy Level  
None

Cancel Next

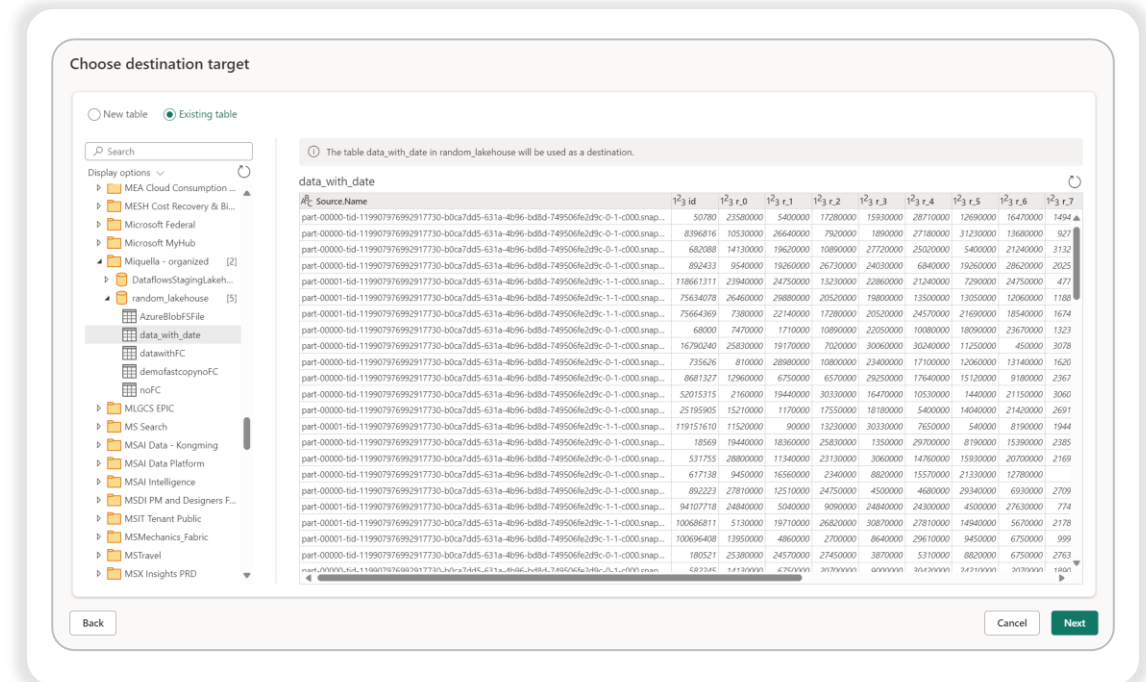
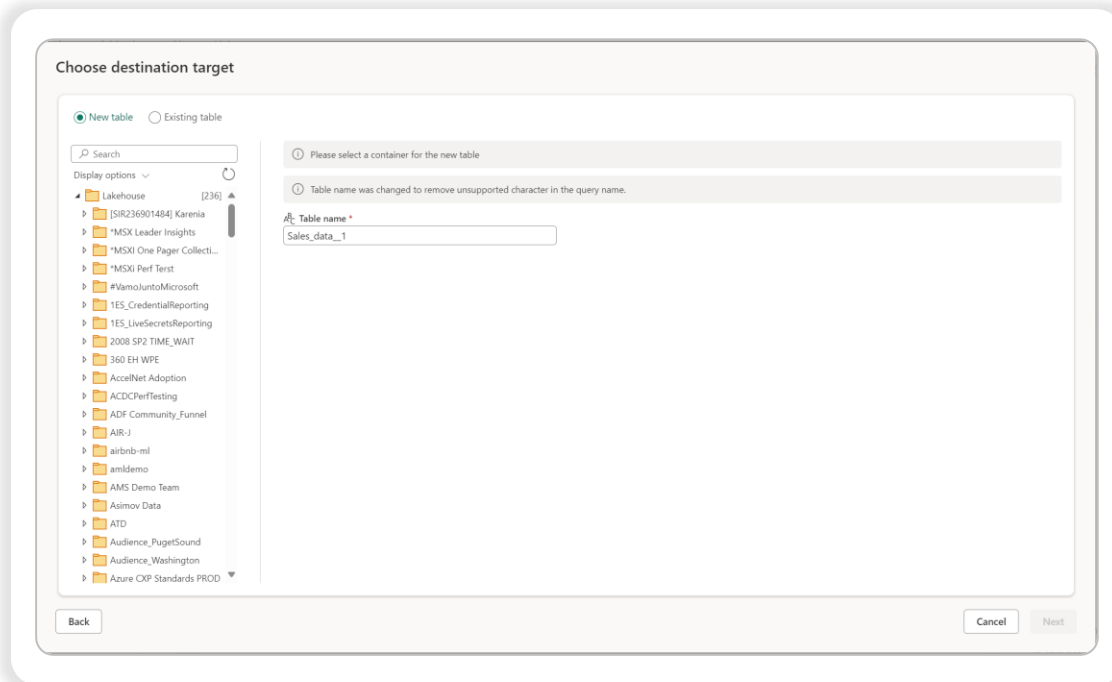
# Load into your data destination

- Create a new table

New table is created in your data destination

- Pick an existing table

Table can't be recreated in any scenario



# Settings for new tables

Choose destination settings

☒ Use automatic settings

Column mapping

Source	Source type	Destination	Destination type
Column1	Whole number	Column1	Whole number
Column2	Text	Column2	Text
Column3	Date	Column3	Date

Back

Cancel

Save settings

**The automatic settings**

Update method replace

Managed mapping

Drop and recreate table

Choose destination settings

☐ Use automatic settings

Update method

Existing data   New data →   Append   Replace

Schema options on publish

Existing schema →   Dynamic schema   Fixed schema

Column mapping

Source	Source type	Destination	Destination type
Column1	Whole number	Column1	Whole number

Back

Cancel

Save settings

**Manual settings**

Get full control over how to load your data

Update methods: Replace and Append

Schema options on publish

# Dataflow refresh

- On-demand
- By setting up a refresh schedule

The screenshot displays the Microsoft Power Query application window. The ribbon at the top includes tabs for Home, Transform, Add column, View, and Help. The Home tab is active, showing various options like 'Get data', 'Enter data', 'Manage connections', 'Options', 'Manage parameters', 'Refresh', 'Advanced editor', 'Add data destination', 'Choose columns', 'Remove columns', 'Keep rows', 'Remove rows', 'Filter rows', 'Sort', 'Transform', 'Combine', and 'Map to entity CDM'. The 'Queries [1]' pane on the left shows a single query named 'Customers'. The main workspace displays a data flow diagram for the 'Customers' query, starting from a 'Source' step, followed by 'Changed column...', 'Renamed columns', 'Removed columns', and finally 'Filtered rows'. Below the diagram is a preview of the data table with columns: CustomerID, CompanyName, ContactName, ContactTitle, and Address. The status bar at the bottom indicates 'Completed (4.39 s) Columns: 5 Rows: 99+'. The 'Query settings' pane on the right shows the 'Properties' tab with the query name 'Customers', entity type 'Custom', and a list of applied steps: Source, Changed c..., Renamed c..., Removed c..., and Filtered rows. The 'Data destination' section shows 'SQL Server data...'. A 'Publish' button is located at the bottom right.

	CustomerID	CompanyName	ContactName	ContactTitle	Address
1	ALFKI	Adatum Corporation	Tyler Stein	Sales Representative	Obere Str. 57
2	ANATR	Adventure Works Cycles	Daniel Eifyn	Owner	Ayda. de la Constitución 2222
3	ANTON	Alpine Ski House	Ncumisa Mvubu	Owner	Mataderos 2312
4	AROUT	Bellows College	Sam Centrell	Sales Representative	120 Hanover Sq.
5	BERGS	Best For You Organics Company	Ryuu Yasuda	Order Administrator	Berguvsvägen 8
6	BLAUS	Contoso, Ltd.	Sara Folgueroles	Sales Representative	Forsterstr. 57
7	BLONP	Contoso Pharmaceuticals	Serena Davis	Marketing Manager	24, place Kléber
8	BOLID	Contoso Suites	Hailey Clark	Owner	C/ Araquil, 67
9	BONAP	Consolidated Messenger	Jessie Irwin	Owner	12, rue des Bouchers
10	BOTTM	Fabrikam, Inc.	Shawn Hughes	Accounting Manager	23 Tsawassen Blvd.
11	BSBEV	Fabrikam Residences	Yuu Shibata	Sales Representative	Fauntleroy Circus
12	CATSI	Fabrikam Residences	Yuu Shibata	Sales Representative	Fauntleroy Circus

# Incremental refresh in Dataflow Gen2 (Preview)

- Refresh only new or updated data
- Allows you to
  - Reduce refresh times
  - Enhance reliability
  - Minimize resource usage
- Supported data destinations supported
  - Fabric Warehouse
  - Azure SQL Database
  - Azure Synapse Analytics

### Incremental refresh (Preview feature) [?](#)

☒ Enable incremental refresh

Choose a Date/DateTime/DateTimeZone column to filter by \*

OrderDate [v](#)

Extract data from the past \*

2 [v](#) Weeks

Bucket size \*

Day [v](#)

Only extract new data when the maximum value in this column changes \*

ModifiedDate [v](#)

☐ Only extract data for concluded periods

▼ Advanced options

☒ Require incremental refresh query to fully fold

[OK](#) [Cancel](#)

# Use public parameters in Dataflow Gen2 (Preview)

- Dynamically control and customize dataflows
- Refresh dataflow by passing parameter values outside of the Power Query editor
- Pass parameters through the Fabric REST API or through native Fabric experiences

The screenshot displays the 'Settings' tab of the Dataflow Gen2 interface. It includes sections for 'Workspace' and 'Dataflow', each with a dropdown menu and a 'Refresh' button. Below these is a 'Dataflow parameters' section with a table for defining parameters. The table has columns for 'Name', 'Type', and 'Value'. A single parameter 'Region' is listed with a 'String' type and an 'Eastern' value. There is also a 'Treat as null' checkbox.

Name *	Type *	Value
Region	String	Eastern

☐ Treat as null

# Pass custom parameter values for refresh

- Ensure the public parameter mode is enabled for the dataflow item
- Use the updated Dataflow refresh activity to pass parameters via Data Pipelines

The screenshot displays the 'Manage parameters' dialog box, which is used to configure parameters for a dataflow item. The dialog is divided into two main sections: a left sidebar for navigation and a main content area for configuration.

**Left Sidebar (Options):**

- Global**
  - General
  - Data load
  - Diagnostics
- Dataflow**
  - Data load
  - Regional settings
  - Privacy
  - Scale
  - Parameters** (selected)

**Main Content Area:**

**Parameters:** A list of parameters is shown. The 'Region' parameter is selected, indicated by a blue icon and a 'New' link.

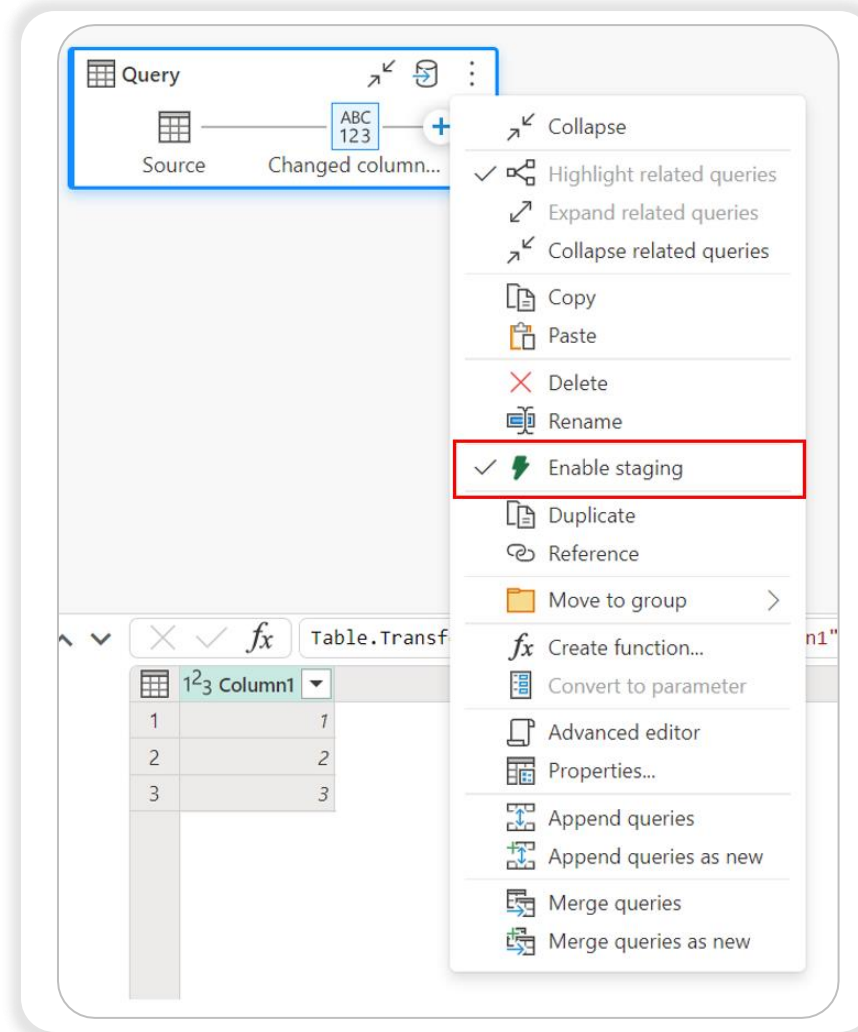
**Configuration Form:**

- Public parameter mode:** A message states 'Public parameter mode is enabled. [Learn more](#)'.
- Name:** A text input field containing 'Region'.
- Description:** A text input field.
- Required:** A checkbox labeled 'Required' is checked.
- Type:** A dropdown menu showing 'Text'.
- Suggested values:** A dropdown menu showing 'Any value'.
- Current value:** A text input field showing 'Eastern'.

**Buttons:** At the bottom of the dialog, there are 'OK' and 'Cancel' buttons.

# Dataflow Gen2 staging artifacts

- DataflowsStagingLakehouse
- DataflowsStagingWarehouse
- Use the dataflow connector in Power BI, Excel, or other dataflows, if you need to access data



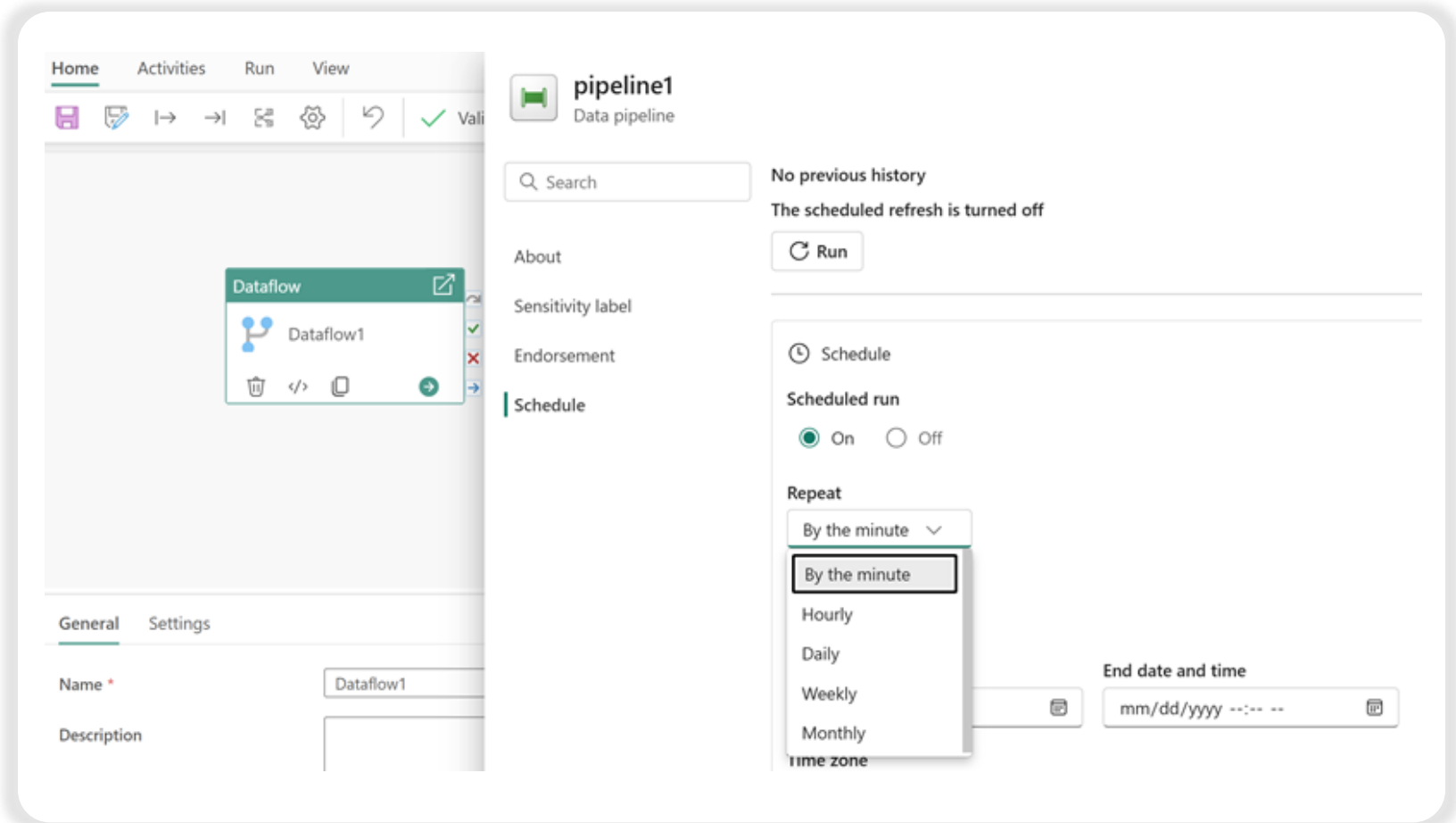


# Decision Guide: Copy Activity vs Dataflow vs Spark

	Pipeline Copy Activity	Dataflow Gen 2	Spark Notebook Job Definition
Use case	Data lake and data warehouse migration, data ingestion Light weight transformation	Data transformation, data wrangling, data profiling Data ingestion	Data transformation, data processing, data profiling, data profiling Data ingestion
Primary developer persona	Data engineer, data integrator	Data engineer, data integrator, business analyst	Data engineer data scientist, data developer
Primary developer skillset	ETL, SQL, JSON	ETL, M, SQL	Spark (Scala, PySpark, Spark SQL, R)
Code written	<a href="#">No code</a> Low code	No code Low code	Code
Data volume	Low to High	Low to <a href="#">High</a>	High
Development interface	Wizard Canvas	Power Query	Notebook Spark job <a href="#">definition</a>
<a href="#">Sources</a>	30+ <a href="#">connectors</a>	150+ <a href="#">connectors</a>	Hundreds of Spark libraries
Destinations	18+ connectors	Lakehouse Azure SQL database Azure Data explorer Azure Synapse analytics	Hundreds of Spark libraries
Transformation complexity	Low: Lightweight - type conversion, column mapping, merge/split files, flatten hierarchy	Low to high: 300+ transformation functions	<a href="#">Low to high:</a> Support for native Spark and open-source libraries

# Integrate Dataflows Gen2 and Pipelines in Microsoft Fabric

- Useful when you need to perform additional operations on the transformed data
- Some common activities in Data engineering and Data Factory:
  - Copy data
  - Incorporate Dataflow
  - Add Notebook
  - Get metadata
  - Execute a script or stored procedure





# Demo

Use a dataflow in a pipeline

# Fast Copy in Dataflows Gen2



# Introduction to Fast copy

Ingest terabytes of data with the easy experience of dataflows

The screenshot displays a dataflow configuration interface. On the left, a 'Source' step is configured for 'Azure Data Lake Storage'. A tooltip indicates that this step will be evaluated by the data source, with a 'Learn more' link. Below the tooltip, a table shows data rows with columns labeled  $1^2_3 r_1$  through  $1^2_3 r_5$ . The table contains six rows of numerical data. On the right, a list of 'Applied steps' is shown, including 'Source', 'Filtered hid...', 'Invoke cust...', 'Renamed c...', 'Removed o...', 'Expanded t...', and 'Changed c...'. Each step has a corresponding icon and a settings gear icon.

**Source**  
Azure Data Lake Storage

This step will be evaluated by the data source.  
[Learn more](#)

This step is going to be evaluated with fast copy.

$1^2_3 r_1$	$1^2_3 r_2$	$1^2_3 r_3$	$1^2_3 r_4$	$1^2_3 r_5$
27450000	3240000	26640000	22770000	2016
1080000	28800000	29070000	15210000	2547
16020000	24210000	26460000	15750000	846
2070000	24840000	12150000	3150000	1926
30870000	9720000	18270000	8010000	594
4320000	17640000	21420000	18990000	1458

**Applied steps**

- Source
- Filtered hid...
- Invoke cust...
- Renamed c...
- Removed o...
- Expanded t...
- Changed c...

# Fast copy | Supported Dataflow Gen2 connectors



- ADLS Gen2



- Blob storage



- Azure SQL DB



- Lakehouse



- PostgreSQL



- On premise SQL Server



- Warehouse



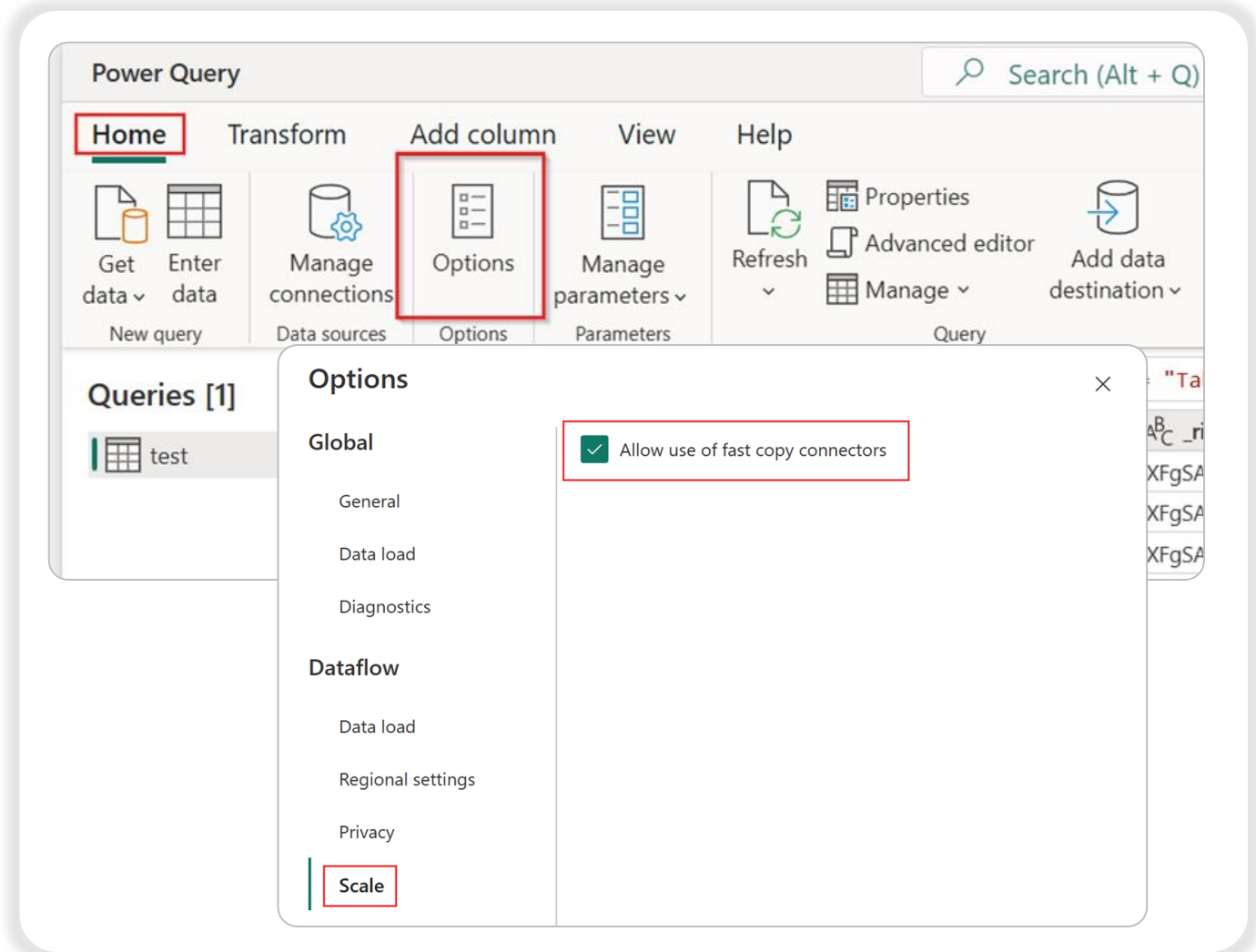
- Oracle



- Snowflake

# Use fast copy







- Navigate to the appropriate Fabric endpoint then to a premium workspace
- Create a dataflow Gen2
- On the **Home** tab of the new dataflow, select **Options**
- Select **Scale** and allow use of fast copy connectors



# Split your query to leverage fast copy

Part 1 - Data ingestion to staging

Part 2 - Large-scale transformation with SQL DW compute

Indicator	Icon	Description
This step is going to be evaluated with fast copy	<div><div> Choose columns</div><div>Choose columns</div></div> <div><div> This step is going to be evaluated with fast copy.</div><div><a href="#">Learn more</a></div></div>	The query up to this step supports fast copy
This step is not supported by fast copy	<div><div> Grouped rows</div><div>Group by</div></div> <div><div> This step is not supported by fast copy.</div><div><a href="#">Learn more</a></div></div>	This step doesn't support Fast Copy
One or more steps in your query are not supported by fast query	<div><div> Changed column type</div><div>Change type</div></div> <div><div> One or more steps in your query are not supported by fast copy.</div><div><a href="#">Learn more</a></div></div>	Some steps in this query support Fast Copy, while others don't

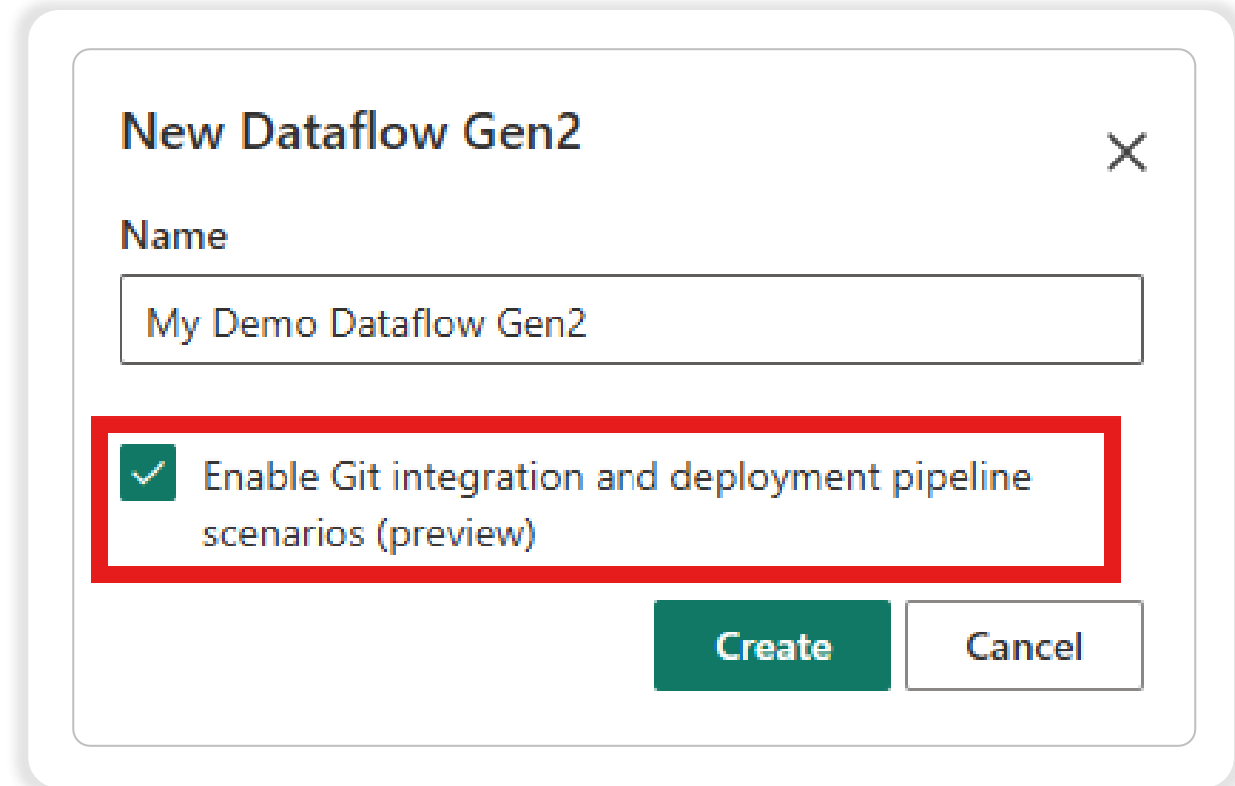


# Dataflow Gen2 with CI/CD and Git integration support



# What you can do with Dataflow Gen2 (CI/CD preview)

- Create, edit, and manage dataflows in a Git repository connected to your fabric workspace
- Automate the deployment of dataflows from your workspace to other workspaces
- Use the Fabric settings and scheduler to refresh and edit settings for Dataflow Gen2



**New Dataflow Gen2** ✕

**Name**




My Demo Dataflow Gen2

☒ Enable Git integration and deployment pipeline scenarios (preview)

**Create** **Cancel**

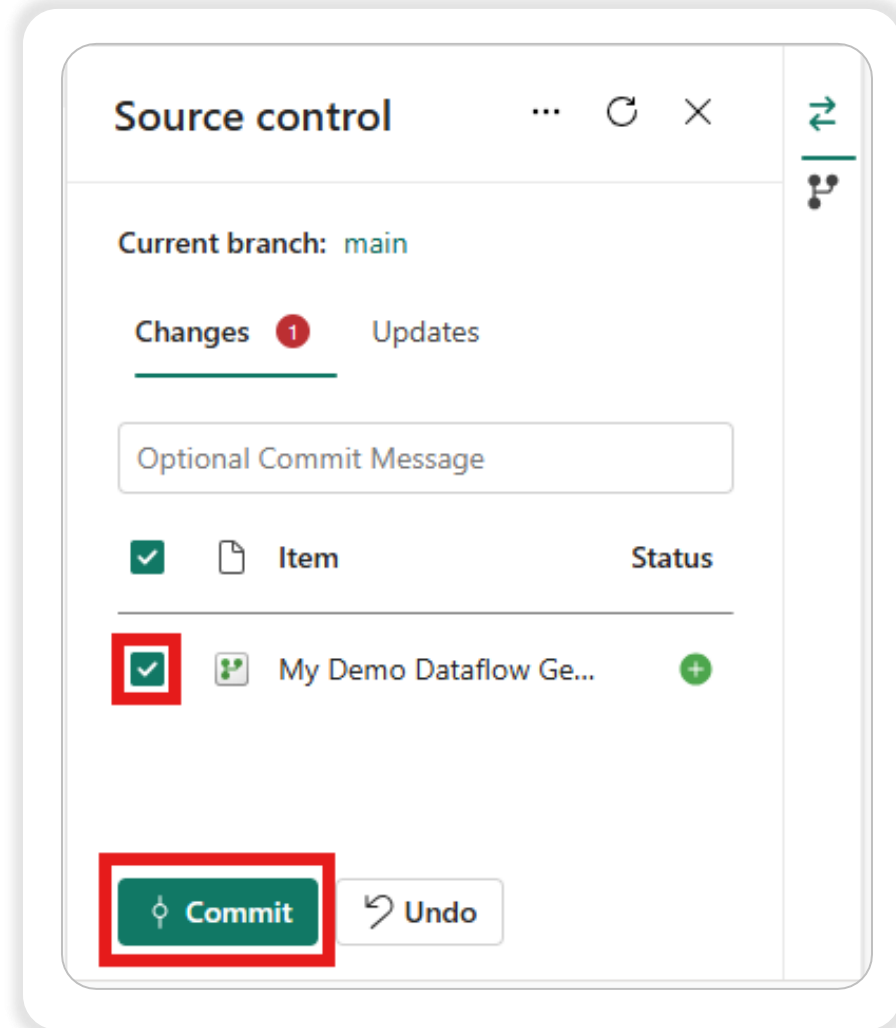
# Benefits of integration

- Automated Deployments
- Version Control
- Collaboration

	Name	Git status	Type
	Dataflow CICD	 Synced	Dataflow Gen2 (CI/CD, preview)

# Set up Dataflow Gen2 with CI/CD and GIT integration

- Enable GIT integration in your workspace settings before proceeding
- When creating a new Dataflow Gen2, there is now an option to enable CI/CD and GIT integration
- Once enabled, you will be using the new Dataflow Gen2 with CI/CD and GIT integration



# Steps to move existing dataflows over to the new Dataflows Gen2 with CI/CD and GIT integration

- Open the Dataflow Gen2 you would like to migrate

- Copy the queries or export a PQT template of the Dataflow Gen2

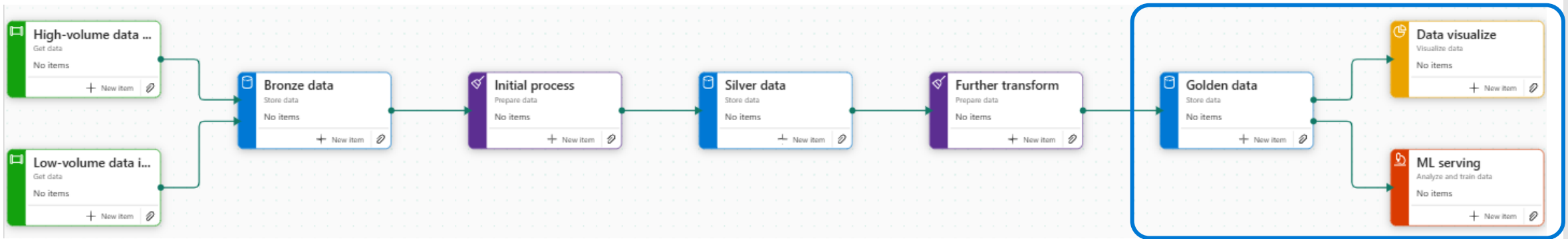
- Create a new Dataflow Gen2 with CI/CD and GIT integration enabled

- Import the queries or PQT template into the new Dataflow Gen2

# Monitor your Dataflows



# Medallion Architecture in Microsoft Fabric Data Factory



# Refresh history

- Provides list of all your data refreshes
- Shows you up to 50 refresh histories or up to 6 months back
- Stores up to 250 refresh histories or 6 months back in the onelake

The screenshot displays the Databricks interface. At the top, a table lists data assets. The first row, 'Dataflow 1', is highlighted, and its context menu is open. The 'Refresh history' option in this menu is highlighted with a red box. Below the table, the 'Refresh history' section for 'Dataflow 1' is shown, featuring a table of refresh events.

Name	Type	Owner
Dataflow 1	Dataflow Gen2	Alex
DataflowsStagingLakehouse		
DataflowsStagingLakehouse		
DataflowsStagingLakehouse		
DataflowsStagingWarehouse		
DataflowsStagingWarehouse		

Start time	Status	Duration	Type
11/3/2023, 3:27:50 AM	✓ Succeeded	00:00:37	On demand
11/3/2023, 2:50:10 AM	✓ Succeeded	00:00:29	On demand



# Reviewing your dataflow refresh from the UI

Drill down into one of the refreshes

← Details

×

Dataflow 1 > 11/3/2023, 2:50:10 AM

Status

✔ Succeeded

Start time

11/3/2023, 2:50:10 AM

Duration

00:00:41

Request ID

XXXXXXXX-XXXX-XXXX-XXXXXXXXXXXX

Dataflow ID

XXXXXXXX-XXXX-XXXX-XXXXXXXXXXXX

Type

On demand

End time

11/3/2023, 2:50:52 AM

Dataflow type

-

Session ID

XXXXXXXX-XXXX-XXXX-XXXXXXXXXXXX

Tables

Name	Status
Customers	✔ Succeeded
Orders	✔ Succeeded

Activities

Name	Status
WriteToDatabaseTableFrom_TransformForOutputToDatabaseTableFrom_Customers	✔ Succeeded
WriteToDatabaseTableFrom_TransformForOutputToDatabaseTableFrom_Orders	✔ Succeeded

✎ Edit dataflow

← Details

×

Dataflow 1 > 11/3/2023, 2:50:10 AM > Customers

Name

Customers

Start time

11/3/2023, 2:50:16 AM

Duration

00:00:26

Status

✔ Succeeded

End time

11/3/2023, 2:50:42 AM

Engine

-

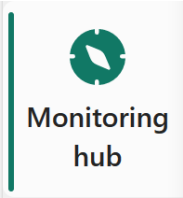
Activity statistics

Endpoint	Volume processed
Lakehouse	42,343 bytes
OData	51,654 bytes

✎ Edit dataflow

# Monitoring hub

Provides you with a dashboard that gives you an overview of the status of your dataflows



## Monitoring hub

Monitoring hub is a station to view and track active activities across different products.

Refresh

Filter by keyword

Filter

Column Options

Reset to default    To apply filters, select the values from the Filter dropdown menu.

Activity name	Status	Item type	Start time	Submitter	Location
Dataflow 1	Succeeded	Dataflow Gen2	3:27 AM, 11/3/23	Henry Cooper	11_02_Tests
PL_pipeline	Succeeded	Data pipeline	3:24 AM, 11/3/23	Henry Cooper	11_02_Tests

Showing all available data

# Copilot for Data Factory in Microsoft Fabric



# Data integration + AI

## Copilot in Data Factory



- Easily integrate generative AI into your dataflows and pipelines using Copilot

Chat with **Copilot** to describe data transformations in natural language

Tap into generative AI capabilities from **Azure Open AI** as data transformation steps

Use **Copilot** to schedule and run and manage dataflows

The screenshot displays the Azure Data Factory Copilot interface. The main workspace shows a dataflow with a 'Source' step connected to an 'Open AI' step. The 'Open AI' step is configured with the prompt: 'Table.AddColumn(OpenAI.Classification, "DissatisfactionReason", ("CustomerReview"))'. Below the dataflow, a table of customer reviews is displayed, with columns for Order ID, City, Country, Customer Review, Dissatisfaction Reason, Order Date, and Due Date. The table contains 8 rows of data.

Order ID	City	Country	Customer Review	Dissatisfaction Reason	Order Date	Due Date	
1	71774	Kansas City	USA	Product was of great quality, but it arrived later than expected. Never got any update from the shipping company as for the reasons for this delay.	Product shipment arrived late	6/1/2022, 12:23:01 PM	6/8/2020
2	71776	Casper	USA	Product packaging was damaged. Either a dog tried to eat it or the shipping service didn't treat it with care.	Product packaging was damaged	2/24/2023, 08:34:25 PM	3/01/2020
3	71780	Houston	USA	This package just showed up on my neighbor's house front porch. I am thankful she knows my name and gave it to me.	Delivered to incorrect address	1/19/2023, 04:20:19 PM	1/26/2020
4	71782	Saint Ann	USA	Product doesn't work - Tried replacing batteries but no luck. Do we have any warranty?	Product was defective	1/29/2022, 07:03:16 PM	2/6/2020
5	71783	Escondido	USA	Product never arrived to me, although it says delivery was completed on time.	Delivered to incorrect address	2/6/2023, 12:00:43 PM	2/13/2020
6	71784	Trabuco Canyon	USA	Product is smashed - Need to return it, can someone from your company pick it up at my home address?	Product was defective	3/5/2023, 06:19:09 PM	3/11/2020
7	71796	Dallas	USA	I recently purchased a new laptop through your company, and it works well, but battery doesn't last for more than a couple of hours. This is a deal breaker for me as I am usually on the road and need to use without power.	Product was defective	11/30/2022, 04:40:12 PM	12/7/2020
8	71797	Seattle	USA	Does your shipping company understand how to provide a good customer service? Very unhappy with the delivery service as the package	Delivered to incorrect address	1/20/2023, 1:42:32 PM	1/27/2020

The interface also includes a 'Query settings' panel on the right, a 'Copilot' chat window, and a 'Data destination' section at the bottom.

# Supported capabilities in Dataflow Gen2

- Generate new transformation steps for an existing query
- Provide a summary of the query and the applied steps
- Generate a new query

The screenshot displays the Dataflow Gen2 interface, which is divided into several panels:

- Queries [1]:** A list on the left shows a single query named "Copilot-Query".
- Query Editor:** The central area shows a table with 13 rows of Windows OS versions and their release years. The formula bar at the top indicates a transformation: `Table.TransformColumnTypes(Source, {{"OS", type text},`.
- Query settings:** A panel on the right containing:
  - Properties:** Name "Copilot-Query".
  - Entity type:** Custom.
  - Applied steps:** A list showing "Source" and "Changed Type".
  - Data destination:** No data destination.
- Copilot Preview:** A panel on the far right showing AI-generated suggestions:
  - A "Get data" button.
  - A message: "Create a new query with sample data that lists all the Microsoft OS versions and the year they were released".
  - A suggested query input field containing "Copilot-Query" and an "Undo" button.
  - A red-bordered box containing the same suggestion text: "Create a new query with sample data that lists all the Microsoft OS versions and the year they were released".
  - A disclaimer: "AI-generated content can have mistakes. Make sure it's accurate and appropriate before using it. [Review terms](#)".

# Supported capabilities in Data pipelines

- Pipeline Generation
- Error message assistant
- Summarize Pipeline

The screenshot displays the Azure Data Pipeline Designer interface. The main canvas shows a pipeline with the following steps: a 'Lookup' step named 'Get Purchase Order', followed by another 'Lookup' step named 'Check Sailing Schedule', then an 'If Condition' step named 'Process Shipping Order'. The 'If Condition' step has two outgoing paths: one leading to a 'Copy data' step named 'Copy data2', and another leading to a 'Teams (Preview)' step named 'Teams1'. The interface includes a sidebar on the right with the 'Copilot' tab selected, showing a 'Summarize this pipeline' button and a list of steps with their descriptions. Below the list is a text input field for a query or question, and a disclaimer at the bottom states: 'AI-generated content can have mistakes. Make sure it's accurate and appropriate before using it. [Review Terms](#)'.

**Parameters** Variables Settings Output

+ New

**Copilot** PREVIEW

22:50

Summarize this pipeline

- 1. Get Purchase Order:** Looks up data from a delimited text file stored in an Azure Blob Storage container named "purchaseorder".
- 2. Check Sailing Schedule:** Queries an Azure SQL table named "SailingSchedule" in the "Shipper" database to find a matching sailing

Describe the summary you want or ask a question.  
Type '/' to list all existing connection references that you'd like to provide.

AI-generated content can have mistakes. Make sure it's accurate and appropriate before using it. [Review Terms](#)

# Limitations of Copilot for Data Factory

- Copilot can't perform transformations or explanations across multiple queries in a single input

- Copilot doesn't understand previous inputs and can't undo changes

- Copilot can't make layout changes to queries in your session

- Copilot doesn't produce a message for the skills that it doesn't support

- Copilot may produce inaccurate results

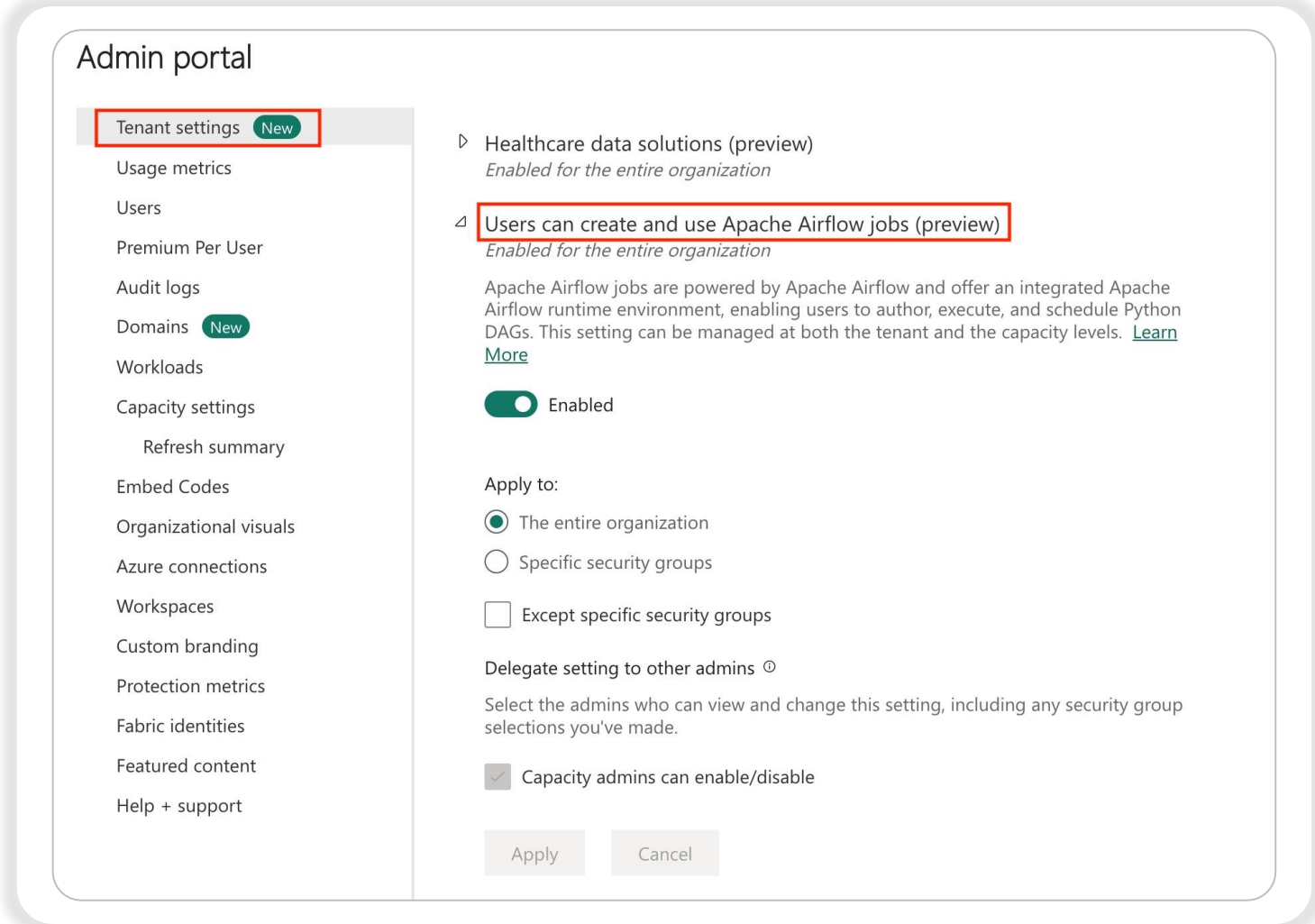
# Workflow Orchestration with Apache Airflow job





# Introduction to Apache Airflow job

- Next generation of Azure Data Factory's Workflow Orchestration Manager
- Brings a SaaS-like experience to running DAGs in a fully managed Apache Airflow environment



The screenshot displays the 'Admin portal' interface. On the left is a navigation menu with items: 'Tenant settings' (highlighted with a red box and a 'New' badge), 'Usage metrics', 'Users', 'Premium Per User', 'Audit logs', 'Domains' (with a 'New' badge), 'Workloads', 'Capacity settings', 'Refresh summary', 'Embed Codes', 'Organizational visuals', 'Azure connections', 'Workspaces', 'Custom branding', 'Protection metrics', 'Fabric identities', 'Featured content', and 'Help + support'. The main content area shows the 'Users can create and use Apache Airflow jobs (preview)' setting (also highlighted with a red box). This setting is 'Enabled for the entire organization' and is currently 'Enabled' (indicated by a green toggle switch). Below this, the 'Apply to:' section has radio buttons for 'The entire organization' (selected) and 'Specific security groups', and a checkbox for 'Except specific security groups'. A 'Delegate setting to other admins' section includes a checkbox for 'Capacity admins can enable/disable', which is checked. At the bottom are 'Apply' and 'Cancel' buttons.

Admin portal

Tenant settings **New**

- Usage metrics
- Users
- Premium Per User
- Audit logs
- Domains **New**
- Workloads
- Capacity settings
- Refresh summary
- Embed Codes
- Organizational visuals
- Azure connections
- Workspaces
- Custom branding
- Protection metrics
- Fabric identities
- Featured content
- Help + support

Healthcare data solutions (preview)  
*Enabled for the entire organization*

Users can create and use Apache Airflow jobs (preview)  
*Enabled for the entire organization*

Apache Airflow jobs are powered by Apache Airflow and offer an integrated Apache Airflow runtime environment, enabling users to author, execute, and schedule Python DAGs. This setting can be managed at both the tenant and the capacity levels. [Learn More](#)

☒ Enabled

Apply to:

- ☒ The entire organization
- ☐ Specific security groups
- ☐ Except specific security groups

Delegate setting to other admins ⓘ

Select the admins who can view and change this setting, including any security group selections you've made.

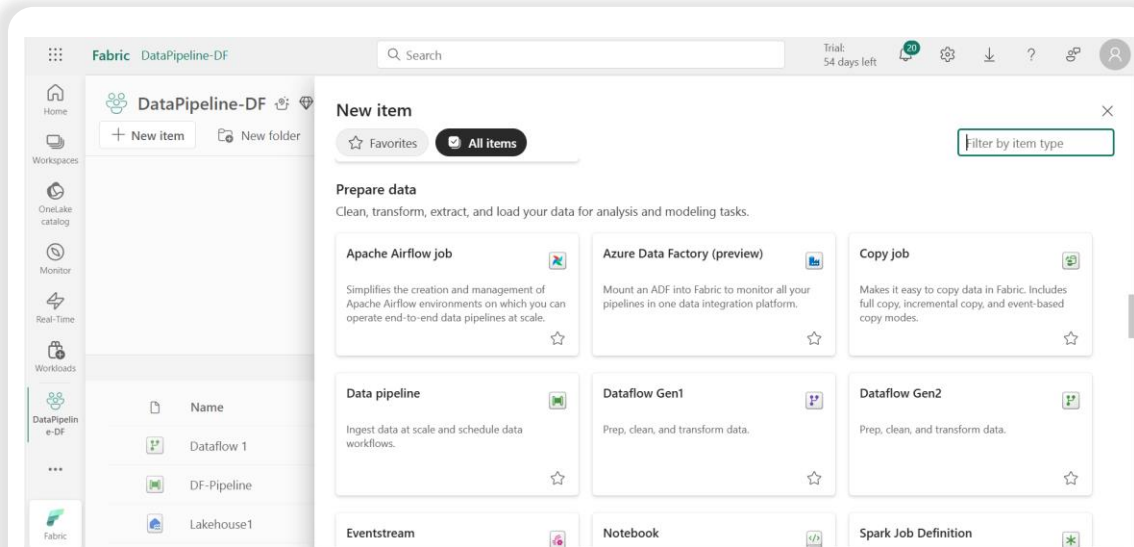
- ☒ Capacity admins can enable/disable

Apply Cancel

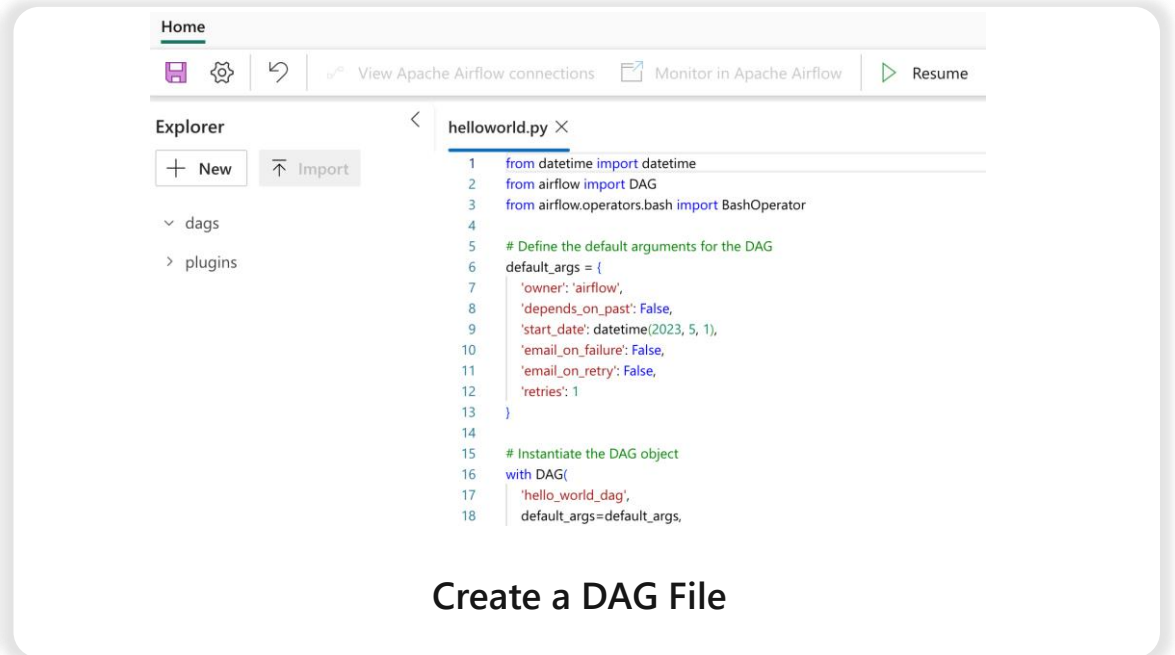
# Key Features

Key Features	Apache Airflow Job in Fabric	Workflow Orchestration Manager in Azure Data Factory
Git sync	Yes	Yes
Enable AKV (Azure Key Vault) as backend	Yes	Yes
Install private package as requirement	Yes	Yes
Diagnostic logs and metrics	No	Yes
Blob Storage	No	Yes
Apache Airflow cluster IP address	Yes	Yes
Autoscale for managing production workload execution spikes	Yes	Partial
High Availability for mitigating outage/downtime	Yes	No
Deferrable Operators for suspending idle operators and free up workers	Yes	No
Pause and Resume TTL (Time to live)	Yes	No
SaaSified Experience - 10 secs to get started - Authoring DAGs - Fabric Free Trial	Yes	No

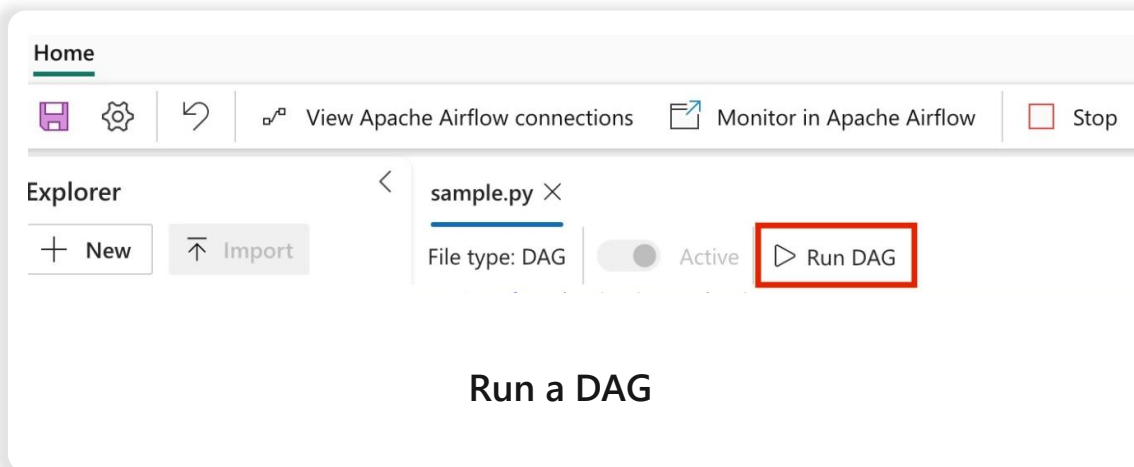
# Create an Apache Airflow Job



Create an Apache Airflow Job in your workspace



Create a DAG File



Run a DAG

# Product comparison



# Dataflows vs. Dataflows Gen2

Feature	Dataflows	Dataflows Gen2
Output file format	CSV	Parquet (V-order)
Fast copy	No	Yes
Output data destinations	No	Yes
Premium only	No	*Yes
AI Insights	Yes	*No
AutoML	Yes	**No (???)
Attach Common Data Model (CDM) folder	Yes	**No (???)
Linked Tables	Yes	**No (? / use Shortcuts)
* To be determined		
** Unknown		
Important note, AI Insights and AutoML are not owned by the data integration team but by the product host (Power BI AI PMs)		

# Coming up tomorrow.....



## 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

# Thank you