



Lab 01: Getting Started with Fabric Data Factory

Introduction

In this lab, you will initiate your lab environment in Fabric by provisioning a Lakehouse, a Data Factory, and copying a set of CSV files from an Azure Blob Storage container.

Objectives

After completing this lab, you will be better able to:

1. Provision a Fabric Lakehouse and a Fabric Data Factory
2. Configure a Data Pipeline to copy a set of CSV files
3. Load Lakehouse tables using the Data Factory

Estimated time to complete this lab

60 minutes

Contents

Lab 01: Getting Started with Fabric Data Factory	1
Introduction	1
Objectives	1
Task 1: Provision the Lakehouse and the Data Factory	3
Task 2: Use the Copy Data wizard	5

Lab Prerequisites

- Workspace: Fabric, Power Premium or Fabric trial
- Individual license: Power Pro or Premium Per User account

Information provided by your training provider

- Trial tenant (if applicable): login & password, workspace to use for the lab.
- Azure Data Lake Gen2 (containing data sources): account name & shared access signature.

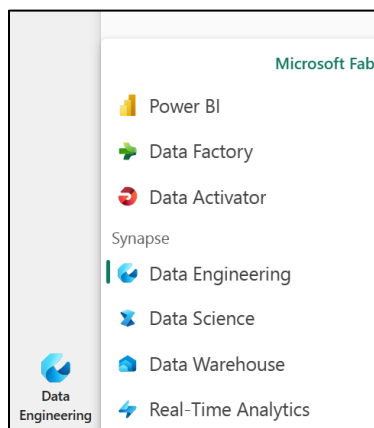
Task 1: Provision the Lakehouse and the Data Factory

In this task, you will provide the Lakehouse and the Data Factory environment to be used in this training.

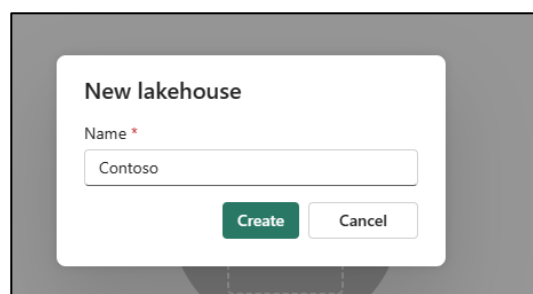
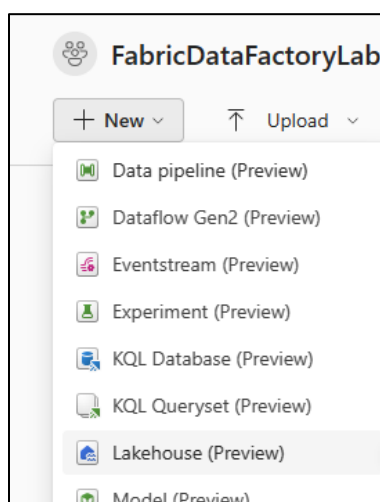
Connect to the Microsoft Fabric environment and go to your assigned workspace, priorly indicated by your trainer. If you don't already have an assigned workspace, create a new one using either:

- Fabric Capacity (F sku)
- Fabric Capacity Trial
- PBI Premium Capacity (P sku)

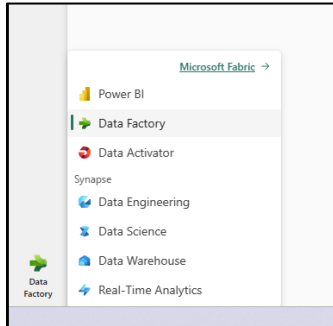
Using the Menu on the bottom left corner, switch to the **Data Engineering** mode.



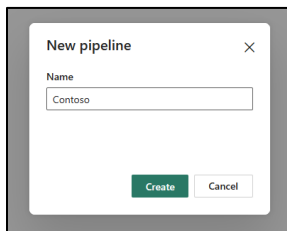
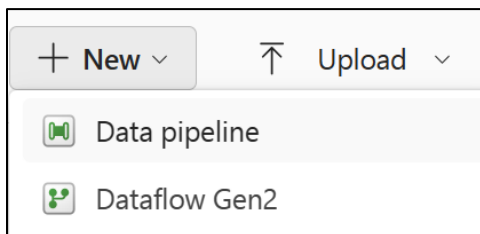
Create a new **Lakehouse** et define the name as **Contoso**.







Using the Menu on the bottom left corner, switch to the **Data Factory** mode.



Create a new **Data pipeline** and define the name as **Contoso**.



Your workspace should contain the following artifacts:

	Name	Type
	Contoso	Lakehouse
	Contoso	Semantic model (default)
	Contoso	SQL analytics endpoint
	Contoso	Data pipeline

Task 2: Use the Copy Data wizard












In this task, you will copy a set of CSV files from an Azure Data Lake Gen 2 account to the "Bronze" zone of your Lakehouse using the Copy Data wizard.

These files are stored in the following path:

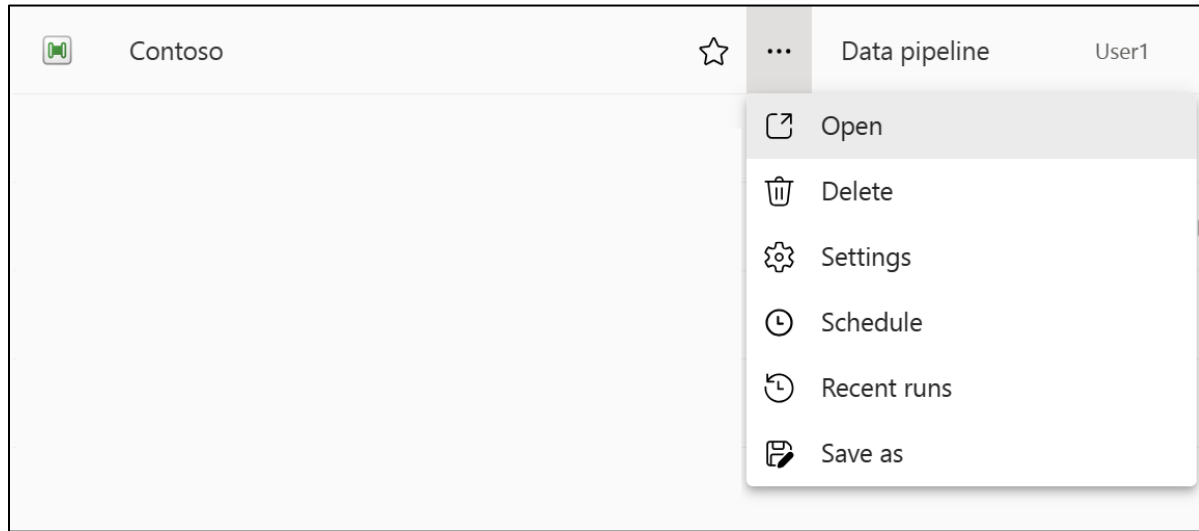
Authentication method: Access key ([Switch to Azure AD](#))

Location: [labdata](#) / [csv](#) / Dimensions

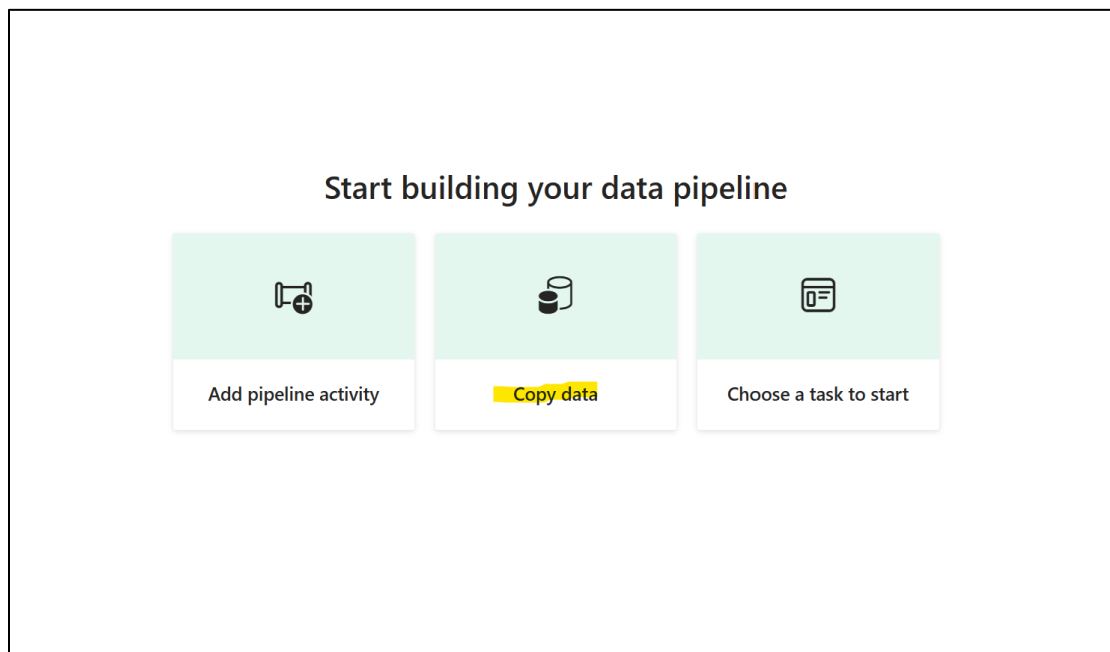
Search blobs by prefix (case-sensitive)

Name	
<input type="checkbox"/>	 Calendar.csv
<input type="checkbox"/>	 CustomerGroupMapping.csv
<input type="checkbox"/>	 CustomerGroups.csv
<input type="checkbox"/>	 Customers.csv
<input type="checkbox"/>	 Geography.csv
<input type="checkbox"/>	 ProductCategories.csv
<input type="checkbox"/>	 Products.csv
<input type="checkbox"/>	 ProductSubCategories.csv
<input type="checkbox"/>	 Promotions.csv
<input type="checkbox"/>	 Stores.csv
<input type="checkbox"/>	 StrategicCustomers.csv

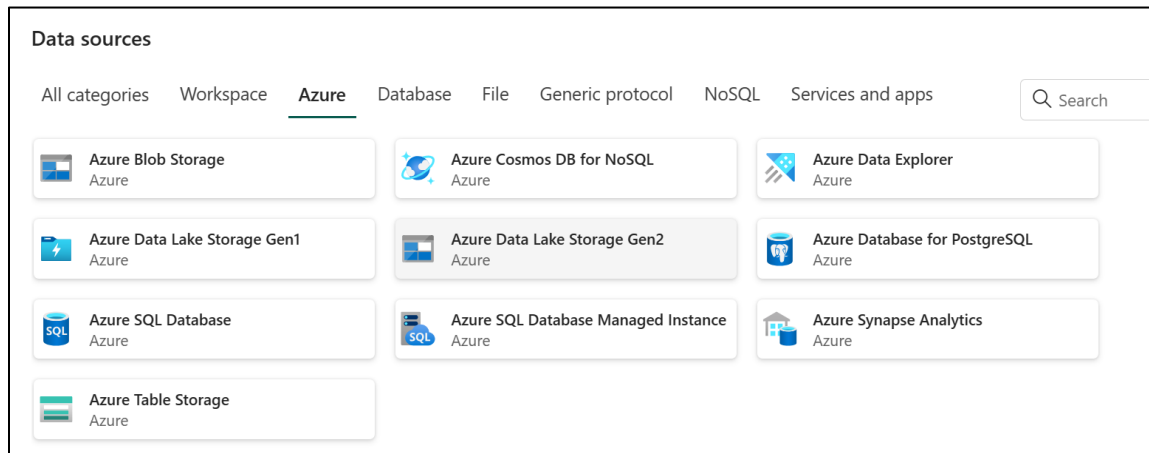
Open the **Data Pipeline** named Contoso created in the previous lab task.



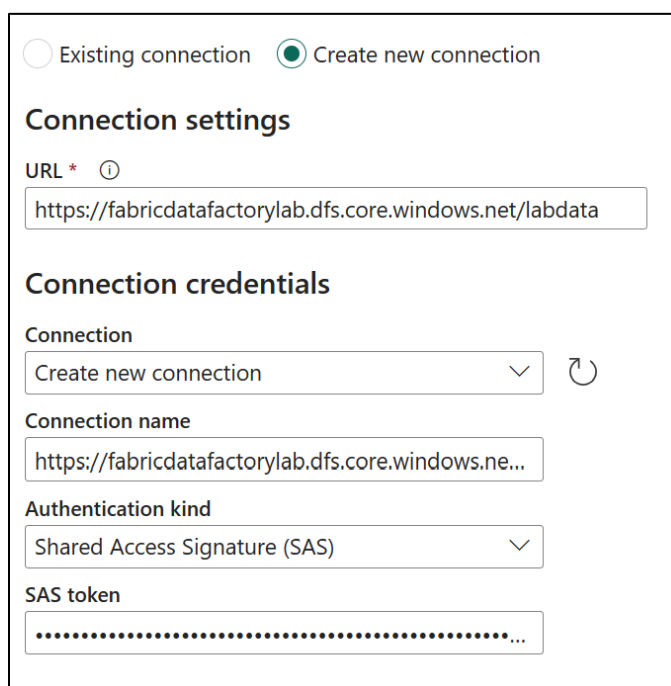
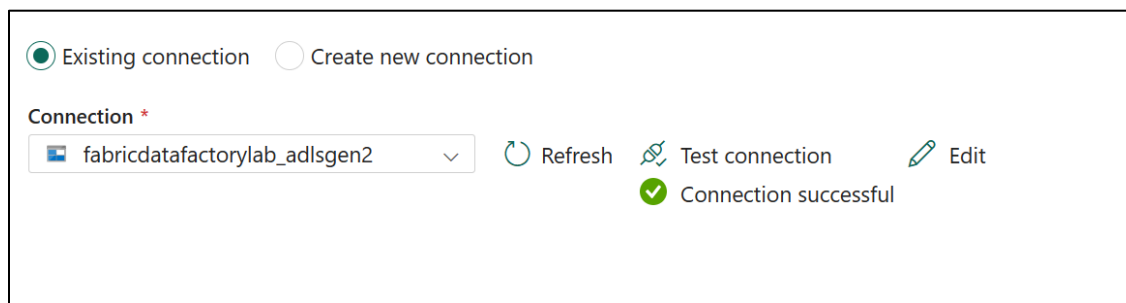
Click on **Copy data** to start the Copy data wizard.



Select **Azure Data Lake Storage Gen2** and click **Next**.



Use the existing connection or create a new connection with the URL and a SAS token.



Select the **Dimensions** folder located in the **CSV** and enable the **Binary copy**. Click on **Next**.

The screenshot shows the 'Connect to data source' step in the Microsoft Fabric Data Factory interface. On the left, a vertical progress bar indicates the current step is 'Connect to data source'. The main area is titled 'Connect to data source' and contains a search bar and a list of file systems. The 'Dimensions' folder under the 'csv' file system is selected. On the right, the 'Preview data: Dimensions' section shows the 'Schema agnostic (binary copy)' option is checked, and an 'Advanced' link is visible.

Select **Lakehouse** as the destination. Click on **Next**.

The screenshot shows the 'Choose data destination' step in the Microsoft Fabric Data Factory interface. On the left, the progress bar indicates the current step is 'Choose data destination'. The main area is titled 'Choose data destination' and contains a 'Data destinations' section. The 'Workspace' tab is selected, and the 'Lakehouse Workspace' is highlighted as the chosen destination.

Use the **Contoso** Lakehouse created previously and click **Next**.

The screenshot shows the 'Choose data destination' step in the Microsoft Fabric Data Factory interface. On the left, the progress bar indicates the current step is 'Choose data destination'. The main area is titled 'Choose data destination' and contains a 'Lakehouse' section. The 'Existing Lakehouse' radio button is selected, and the 'Contoso' Lakehouse is chosen from the dropdown menu. A 'Refresh' button is also visible.

Type **raw** as the folder path for the destination and select **Preserve hierarchy** as the Copy behavior. Click on **Next**.

The screenshot shows the 'Copy data' wizard in Microsoft Fabric Data Factory. The left sidebar indicates the current step is 'Connect to data destination'. The main panel is titled 'Connect to data destination' and contains the following fields:

- Root folder:** Radio buttons for 'Tables' and 'Files'. 'Files' is selected.
- Folder path:** A text input field containing 'raw'. A 'Browse' button is to the right. Below the input is a note: 'If the identity you use to access the data store only has permission to subdirectory instead of the entire account, specify the path to browse.'
- File name:** A text input field containing 'File names are defined by source'.
- Copy behavior:** A dropdown menu with 'Preserve hierarchy' selected.

Select the **Binary** file format. Click on **Next**.

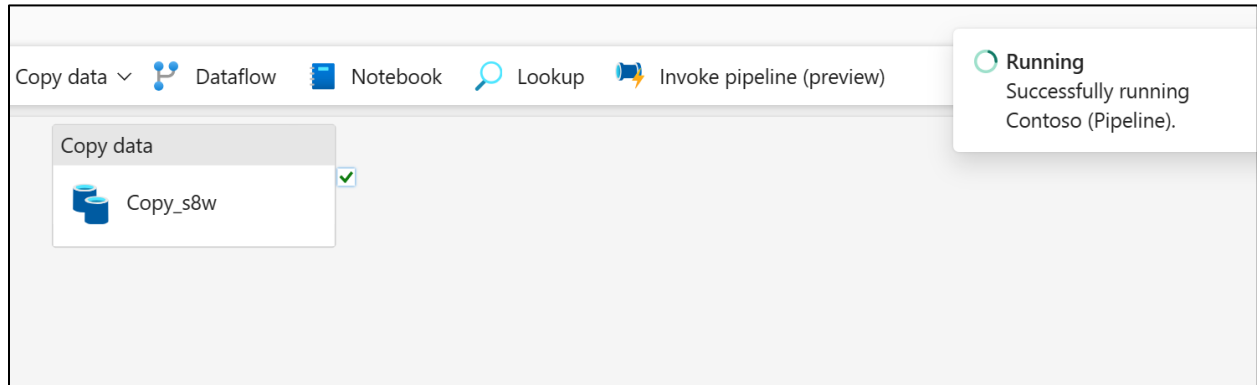
The screenshot shows a 'File format' dropdown menu. The word 'Binary' is selected and displayed in the dropdown box.

Click on **Save + Run**.

The screenshot shows the 'Copy data' wizard in Microsoft Fabric Data Factory, at the 'Review + save' step. The left sidebar indicates the current step is 'Review + save'. The main panel is titled 'Copy Summary' and contains the following information:

- Source:** Azure Data Lake Storage Gen2. Connection name: fabricdatafactory/lab_adisgen2. Folder path: csv/Dimensions.
- Destination:** Lakehouse. Connection name: Contoso.
- Options:** A checkbox labeled 'Start data transfer immediately' is checked.
- Buttons:** 'Back', 'Save + Run', and 'Cancel' buttons are at the bottom.

The **Copy data** component will be executed automatically.



A new activity appears, representing the current pipeline execution. Wait until having the end of the execution with the **Succeeded** Activity status.

Pipeline run ID: f05bf68b-6a3e-4d19-8911-30df972bfc0f				Pipeline status Succeeded
Showing 1 - 1 items				
Activity name	Activity status	Run start	Duration	
Copy_s8w	Succeeded	10/14/2023, 8:01:23 PM	28s	

Click on the activity name to see the execution details, including the duration, the number of files processed, and the quantity of data read and written.

Copy data details

Copy_5zf

Source

Azure Data Lake Storage Gen2

Destination

Lakehouse

Data read: 2,101 MB

Files read: 11

Data written: 2,101 MB

Files written: 11

Status Succeeded

Start time 10/14/2023, 8:38:09 PM

Pipeline run activity ID 0b2912fa-98f1-489d-a740-45c41b3227c9

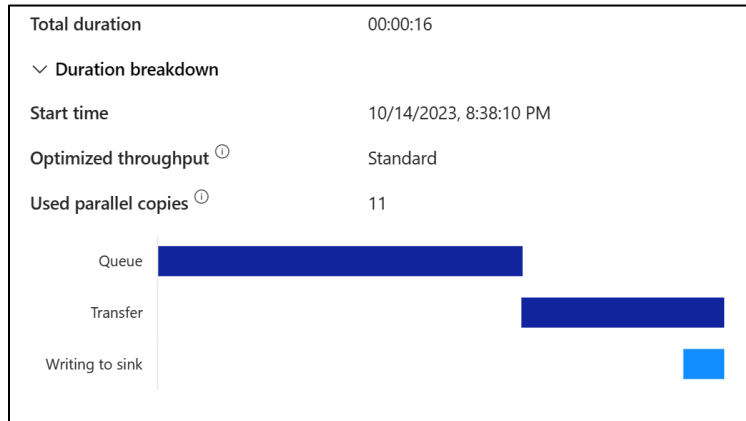
Throughput 420,133 KB/s

Total duration 00:00:16

> Duration breakdown

> Advanced

Click on **Duration breakdown** to see the detailed sequence of the activity and the level of parallelism.



Rename the **Copy data** activity as **Copy_CSV**.

The screenshot shows the configuration panel for the 'Copy data' activity, which has been renamed to 'Copy_CSV'. The 'General' tab is selected, showing the following settings:

- Name:** Copy_CSV
- Description:** (Empty text box)
- Activity state (preview):** Active (selected), Inactive
- Timeout:** 01:12:00:00
- Retry:** 0
- Advanced:** (Expandable section)

Save the pipeline and click on the Workspace in the toolbar to see the list of the workspace artifacts. Click on the **Contoso Lakehouse** created in the same workspace.

Name	Type
Contoso	Lakehouse
Contoso	Semantic model (default)
Contoso	SQL analytics endpoint
Contoso	Data pipeline

On the Lakehouse explorer, verify the 11 CSV files copied in the **raw** folder.

