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 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.

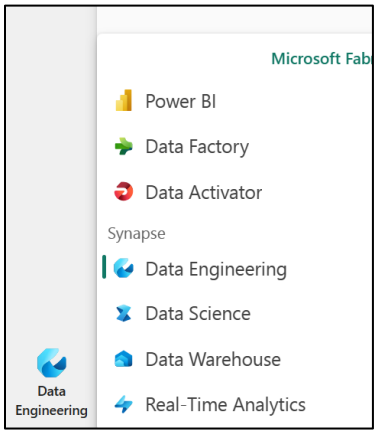
Task 1: Provision the Lakehouse and the Data Factory

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

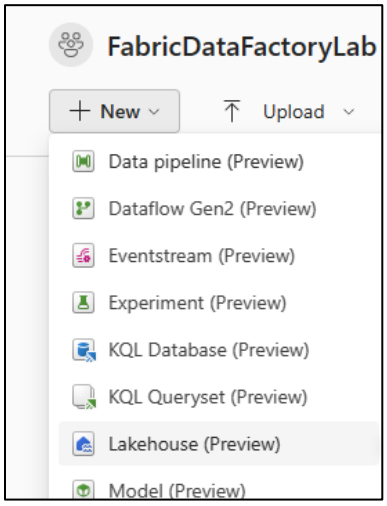
Connect to the Microsoft Fabric environment and go to your assigned workspace, priorly indicated by your trainer. If you don't already have an assigned workspace, create a new one using either:

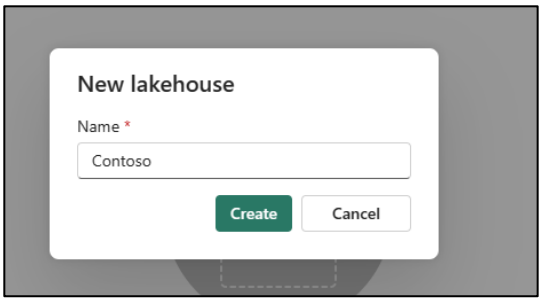
- Fabric Capacity (F sku)
- Fabric Capacity Trial
- PBI Premium Capacity (P sku)

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

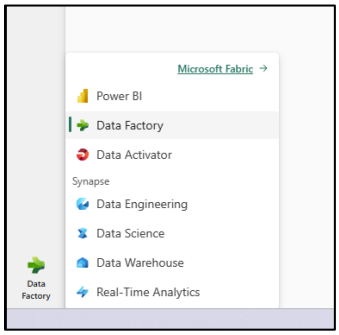


Create a new **Lakehouse** et define the name as **Contoso**.

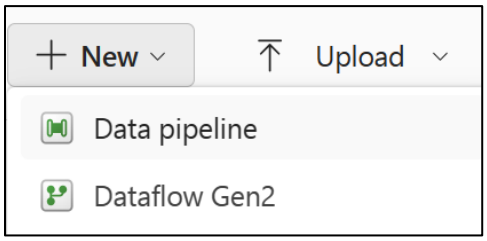


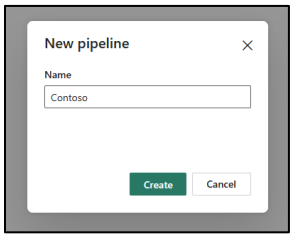


Using the Menu on the bottom left corner, switch to the **Data Factory** mode.

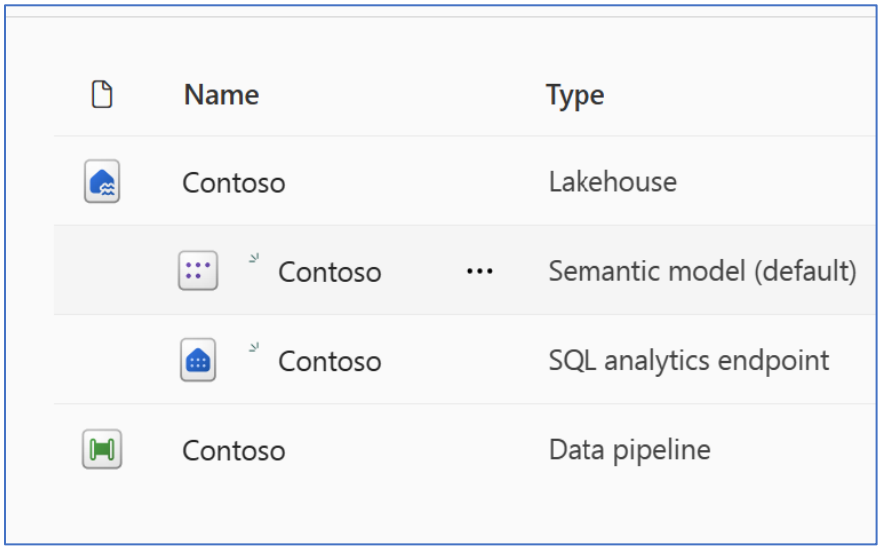


Create a new **Data pipeline** and define the name as **Contoso.**





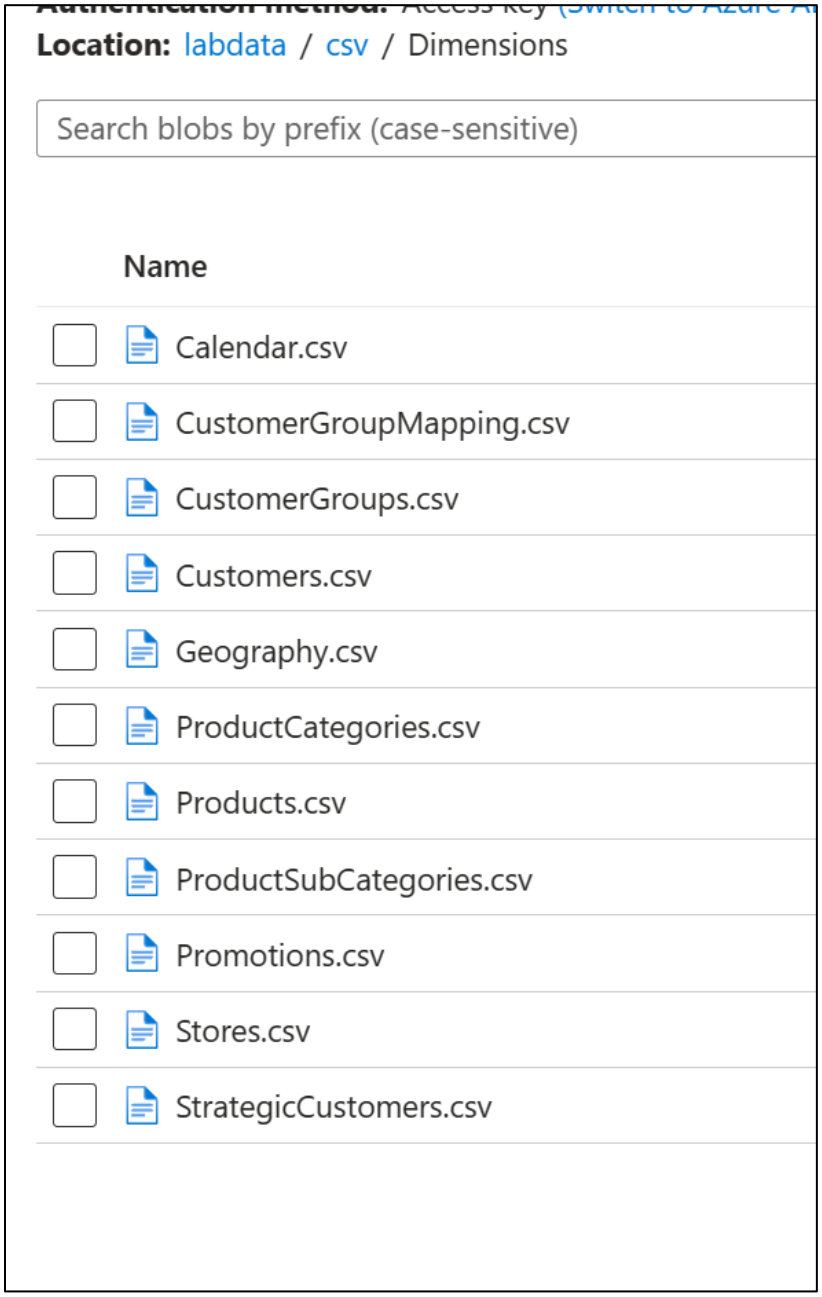
Your workspace should contain the following artifacts:



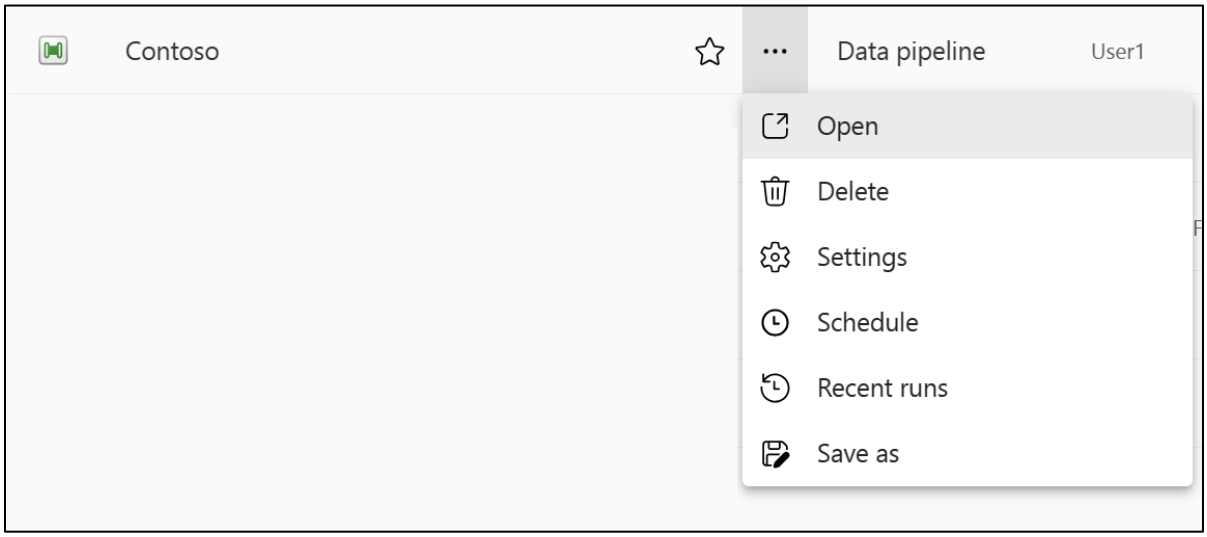
Task 2: Use the Copy Data wizard

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.

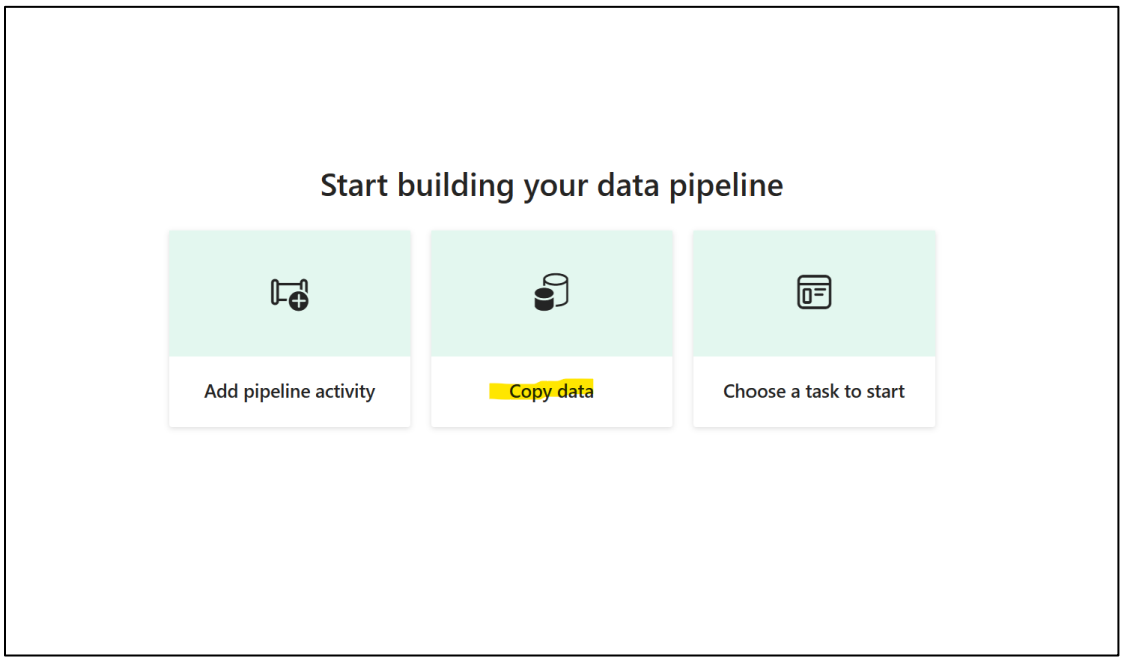
These files are stored in the following path:



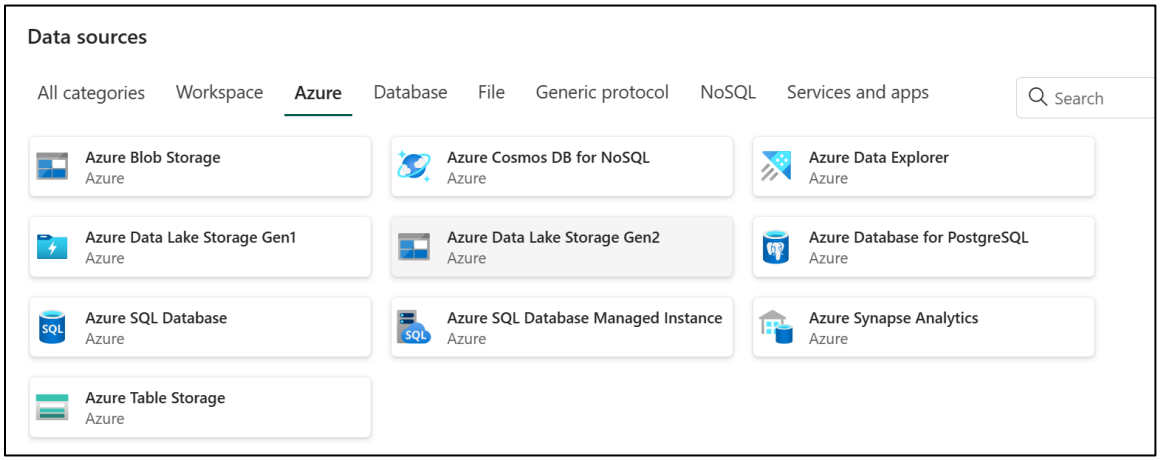
Open the **Data Pipeline** named Contoso created in the previous lab task.



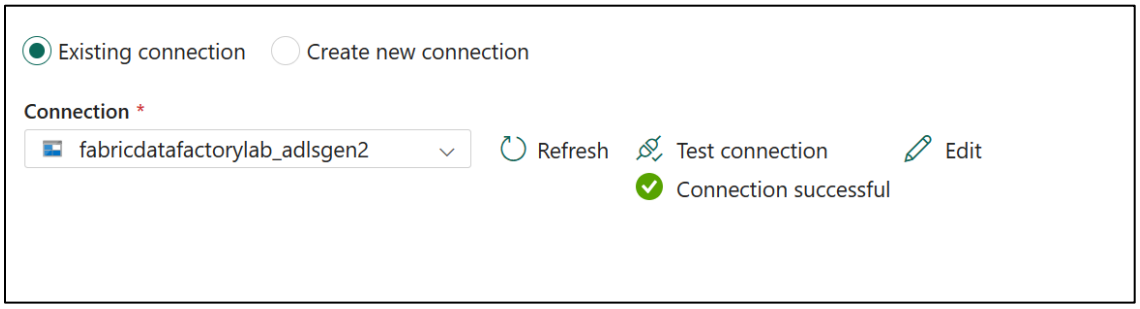
Click on **Copy data** to start the Copy data wizard.

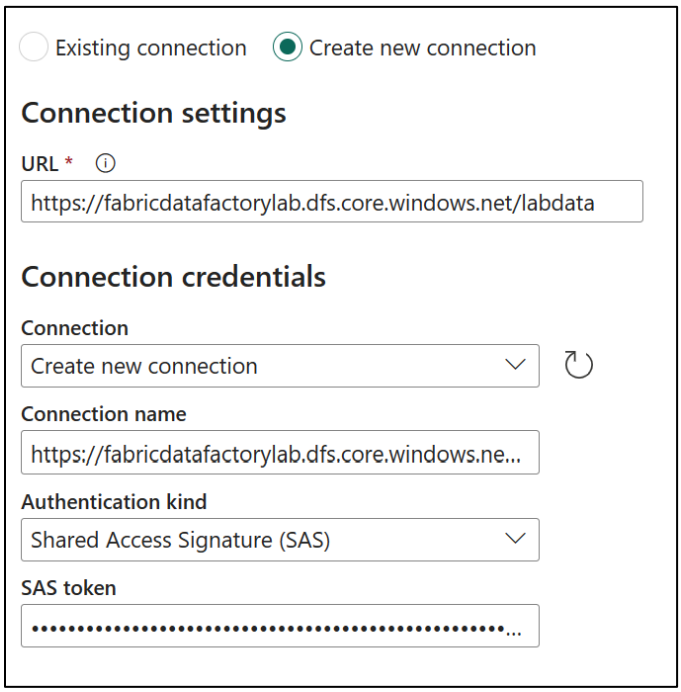


Select **Azure Data Lake Storage Gen2** and click **Next**.

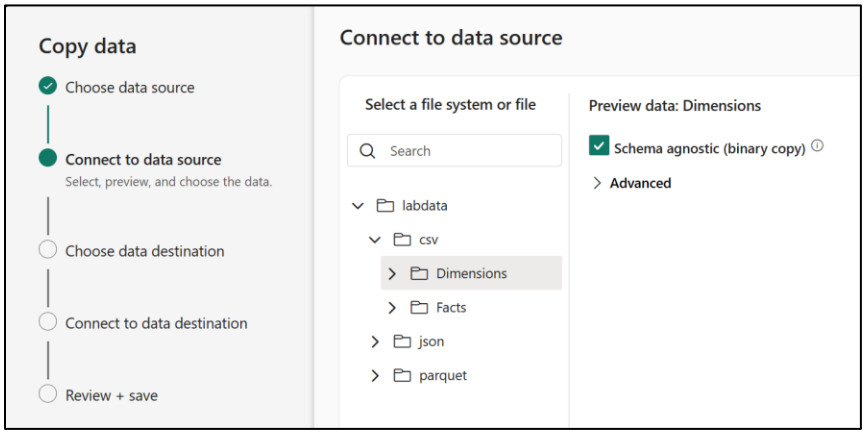


Use the existing connection or create a new connection with the URL and a SAS token.

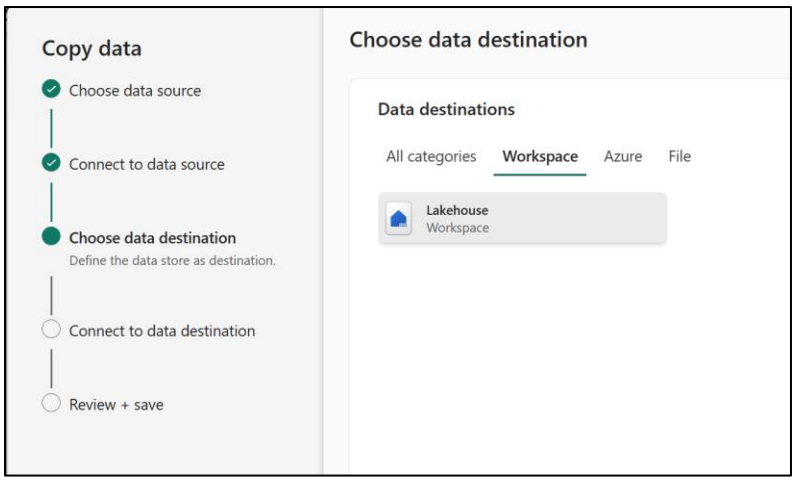




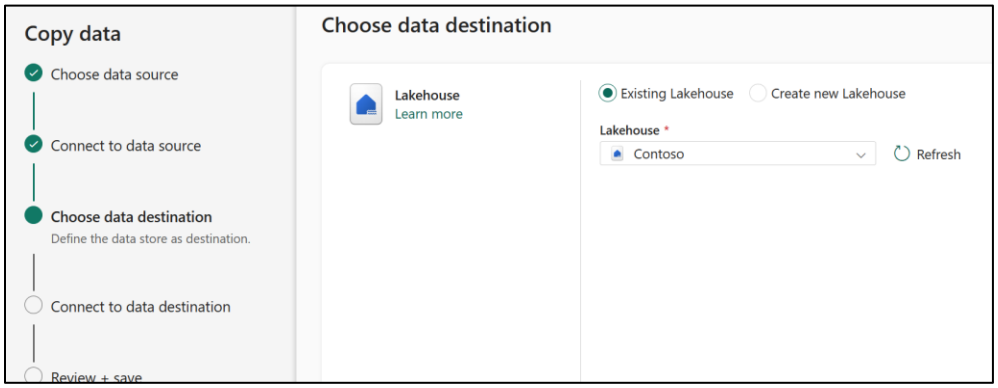
Select the **Dimensions** folder located in the **CSV** and enable the **Binary copy**. Click on **Next**.



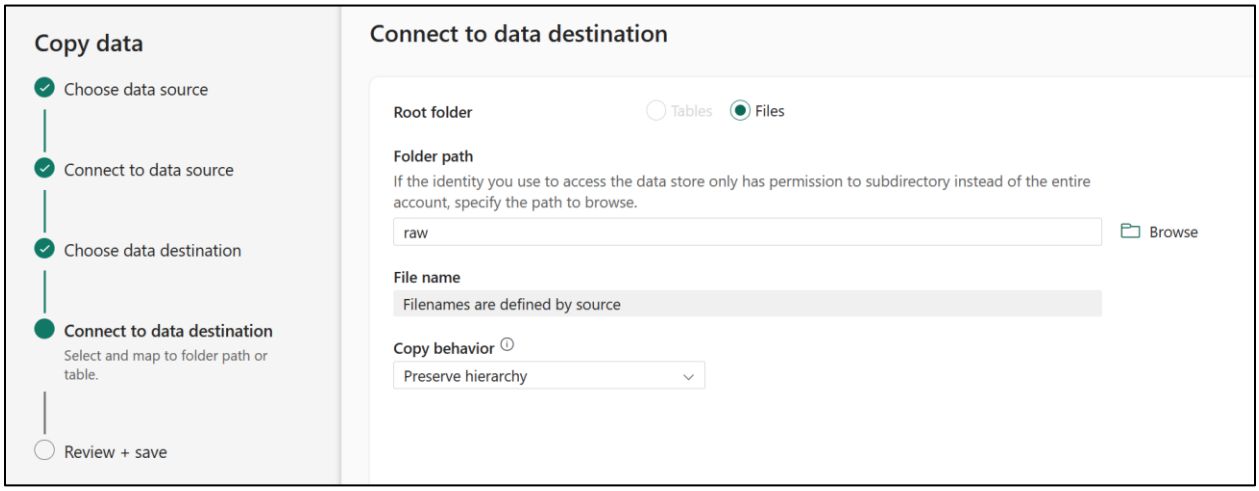
Select **Lakehouse** as the destination. Click on **Next**.



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

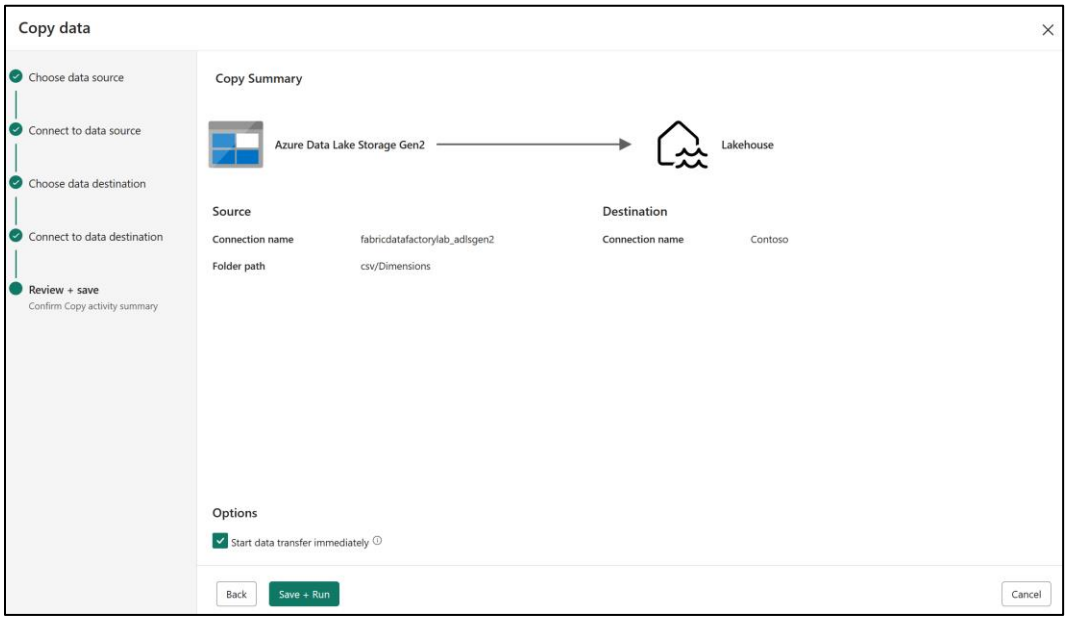


Type **raw** as the folder path for the destination and select **Preserve hierarchy** as the Copy behavior. Click on **Next**.

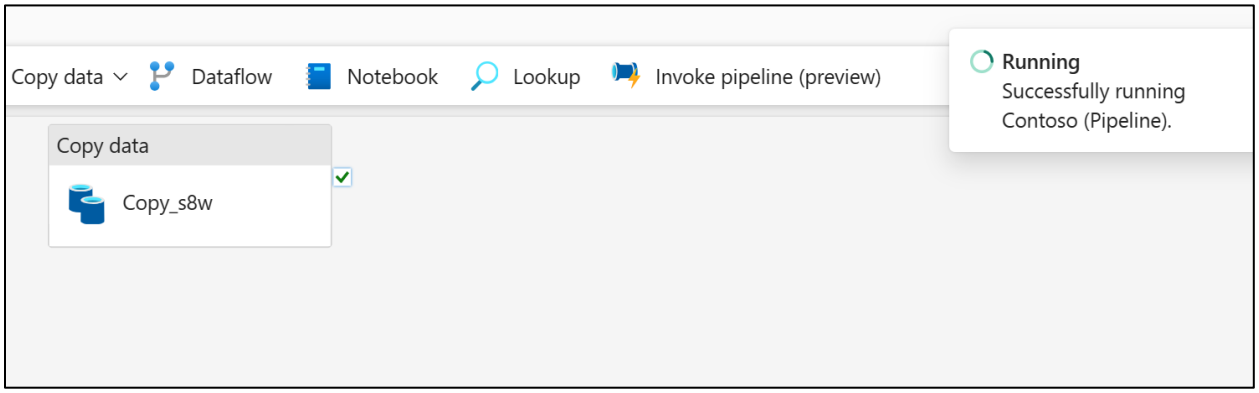


Select the **Binary** file format. Click on **Next**.

Click on **Save + Run**.

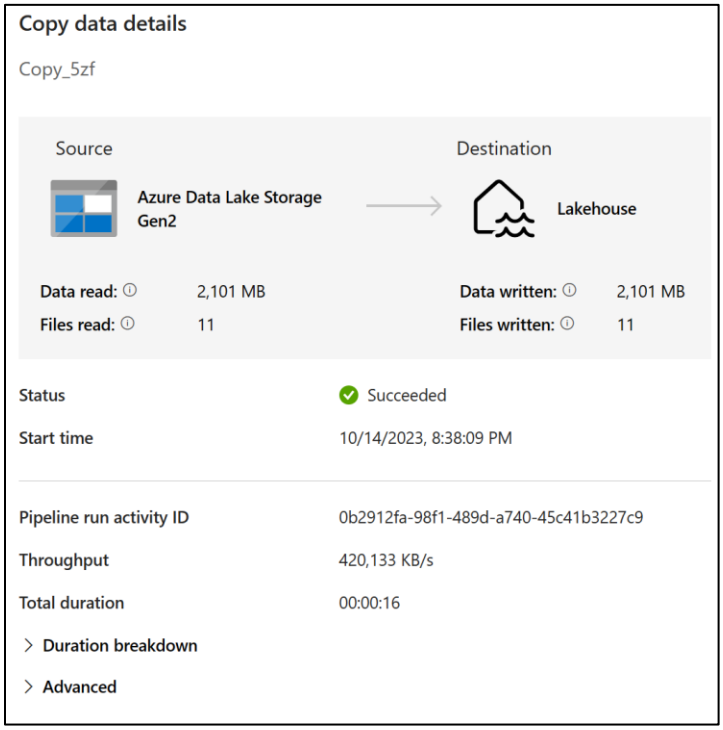


The **Copy data** component will be executed automatically.

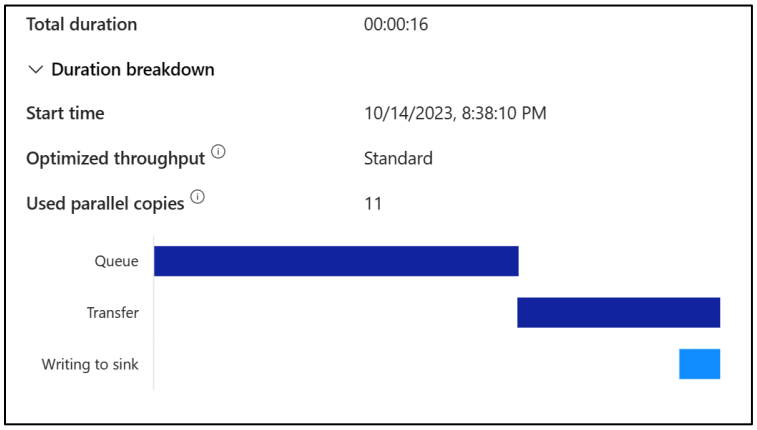


A new activity appears, representing the current pipeline execution. Wait until having the end of the execution with the **Succeeded** Activity status.

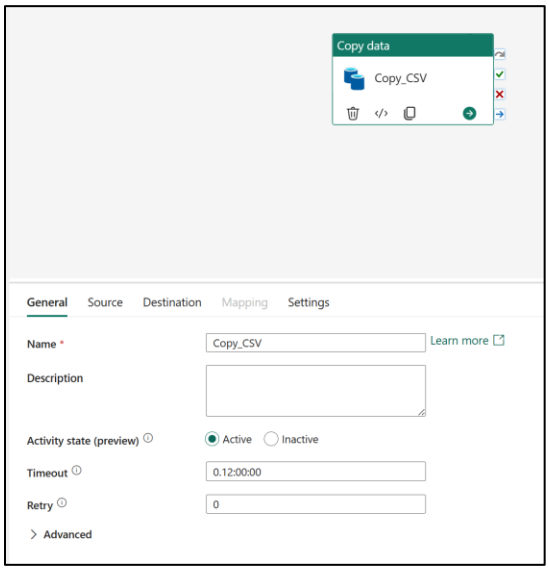
Click on the activity name to see the execution details, including the duration, the number of files processed, and the quantity of data read and written.



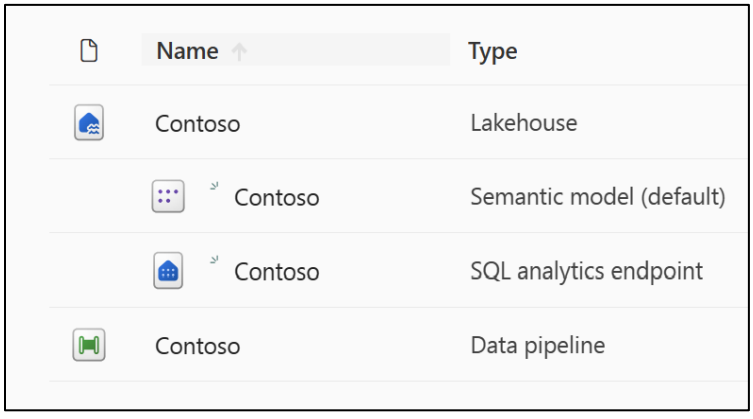
Click on **Duration breakdown** to see the detailed sequence of the activity and the level of parallelism.



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



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.



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