Lab 1 - Create a Fabric workspace and data pipeline

Lab Guide

October 2024

Lab overview....3
Task 1: Create a Fabric workspace....3
Task 2: Create Fabric Lakehouse....7
Task 3: Create a data pipeline....12
Option 1: If you have created an Azure SQL database....12
Option 2: If you have downloaded the data from Sharepoint....29

Lab overview

Creating a Fabric workspace and data pipeline is the foundational step for consolidating Contoso's and Litware Inc.'s data, enabling unified data management and streamlined analytics within the Azure ecosystem. This infrastructure lays the groundwork for efficient data ingestion, storage, and processing. It helps the organization's address data silos and fosters a cohesive data environment.

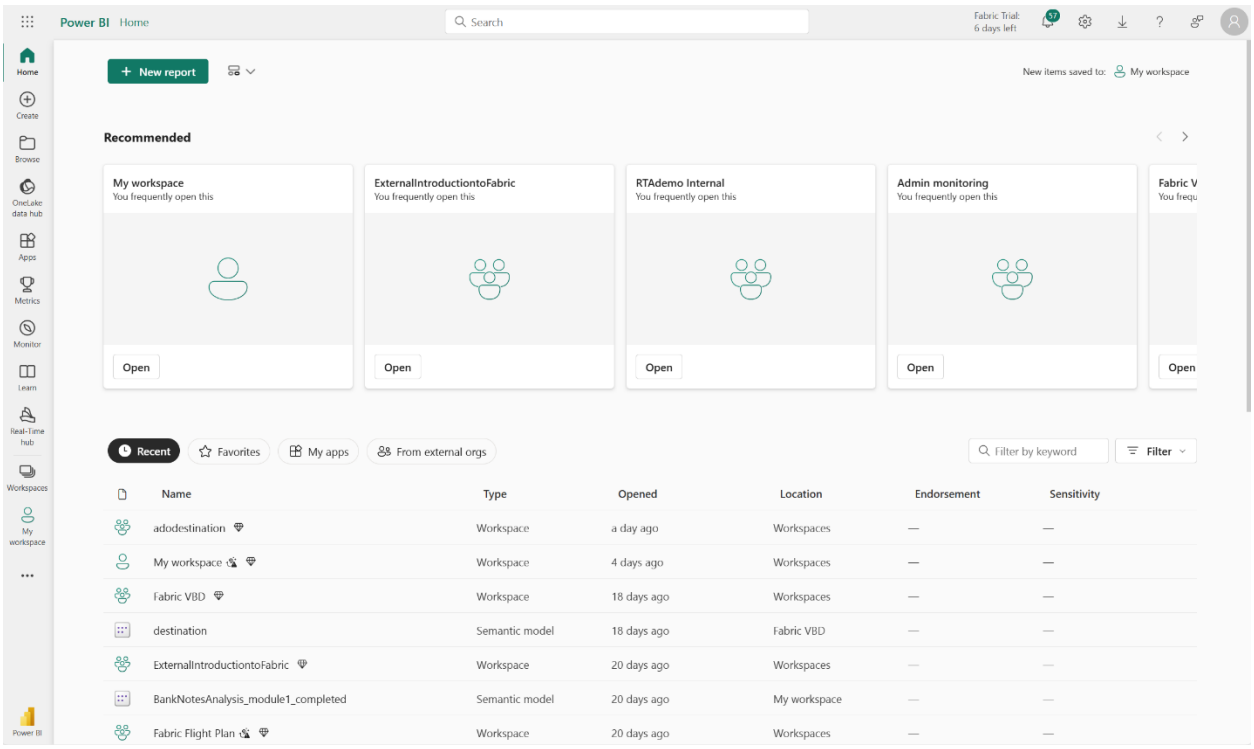
In this lab, you will perform the following tasks:

- **Create a Microsoft Fabric-enabled workspace:** You'll set up a workspace that uses Microsoft Fabric, enabling integration and management of data across various sources and platforms.
- **Create lakehouses:** You will create 3 lakehouses for the medallion architecture where your tables will be stored
- **Use data pipelines for data ingestion:** You'll implement data pipelines to efficiently ingest data into your lakehouse, ensuring a seamless flow of information for processing and analysis in the medallion architecture.

Contoso is setting up a Fabric-enabled workspace as the cornerstone for Contoso's data integration efforts, facilitating a seamless blend of disparate data sources.

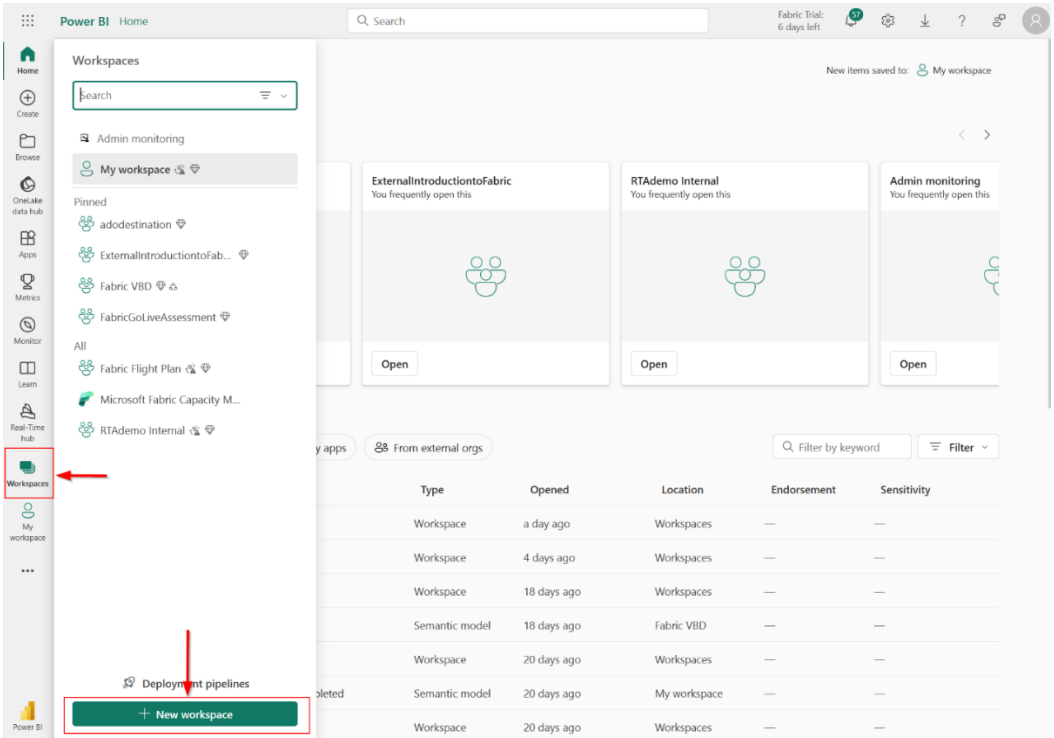
Task 1: Create a Fabric workspace

Open a new browser tab, go to https://app.powerbi.com/ and sign in with your credentials.

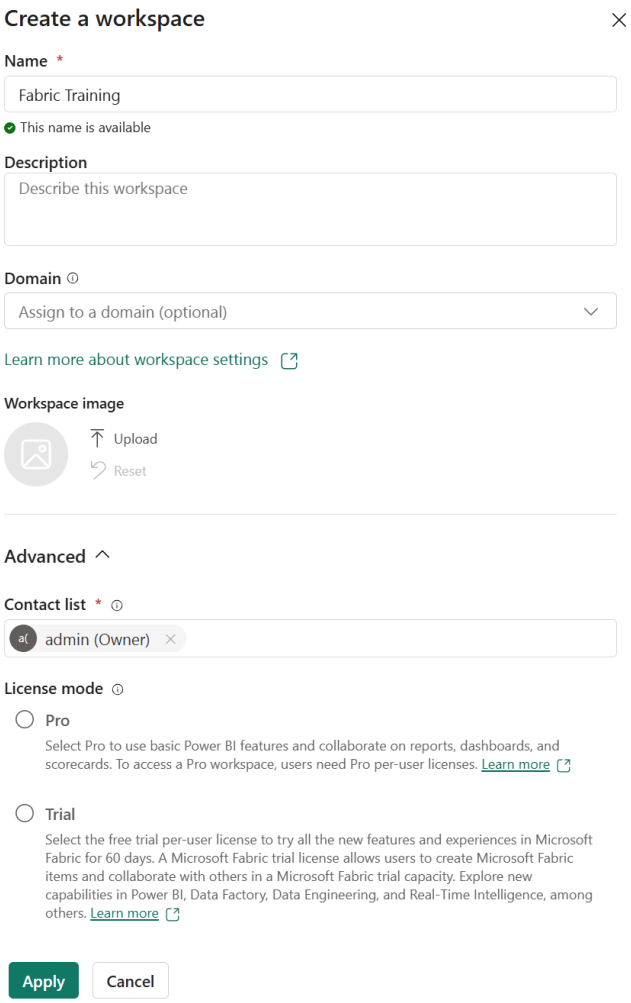


You may see a message alerting you that you have been assigned a Microsoft Fabric (Free) license. Select **OK** to dismiss the message.

1. On the Power BI Home page, in the left navigation pane, select **Workspaces** and then select **+ New workspace**.

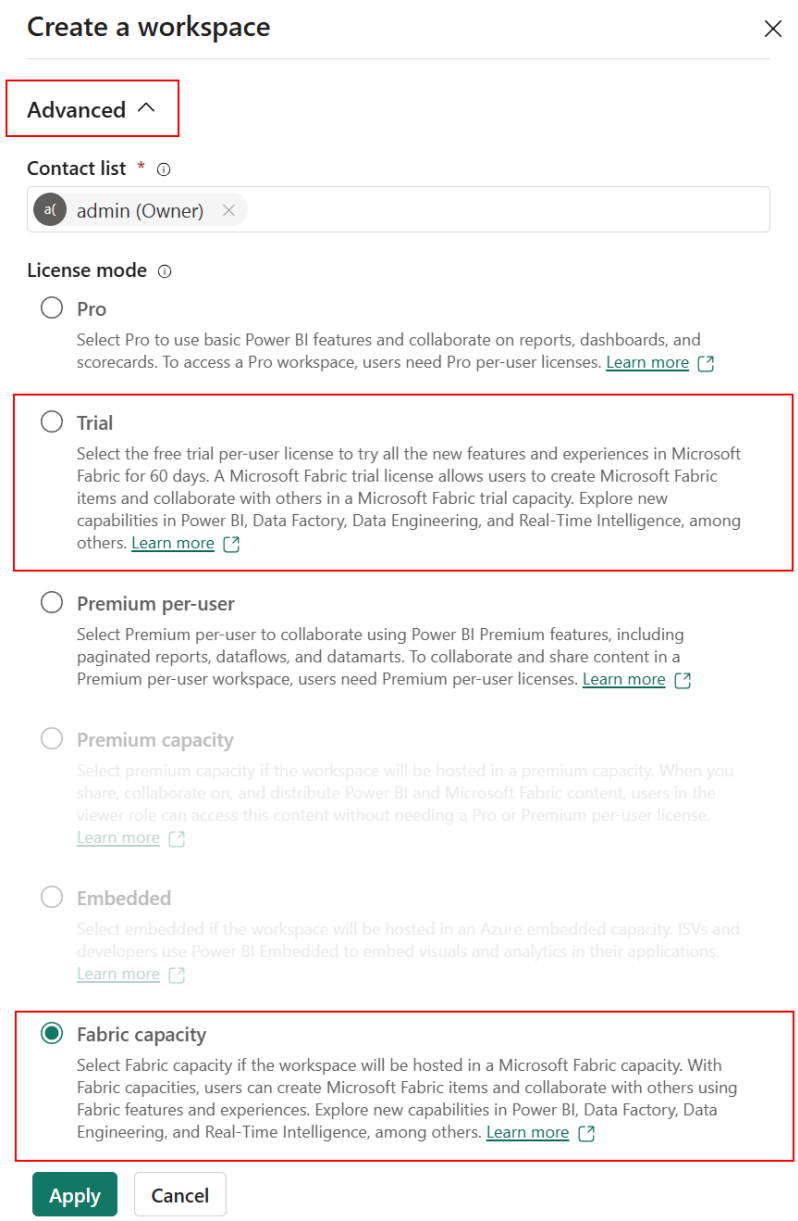


2. In the Name field, enter Fabric Training:

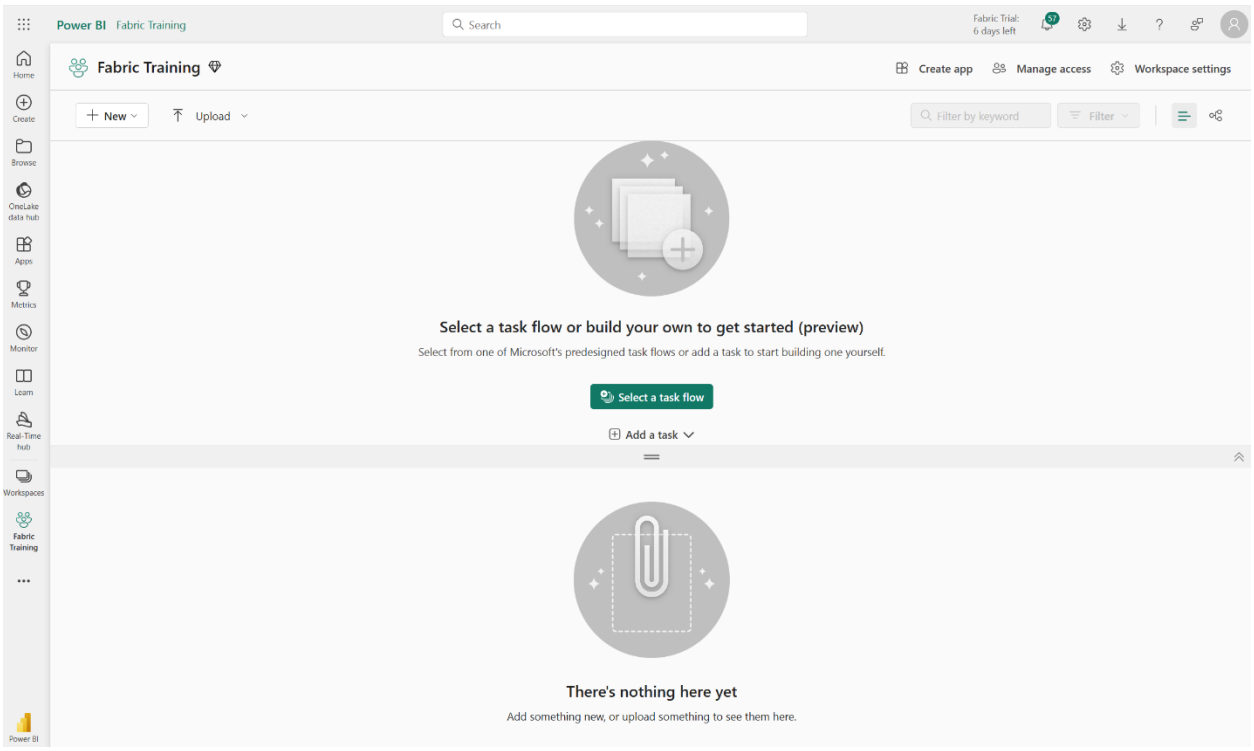


3. After Azure verifies that the name is available, select **Advanced**.

4. In the License mode section, verify that a **Fabric capacity** is selected if you have access to one or choose Trial. Then select **Apply**.



5. On the Power BI page, select **Workspaces** and verify that the Fabric Training workspace was created.

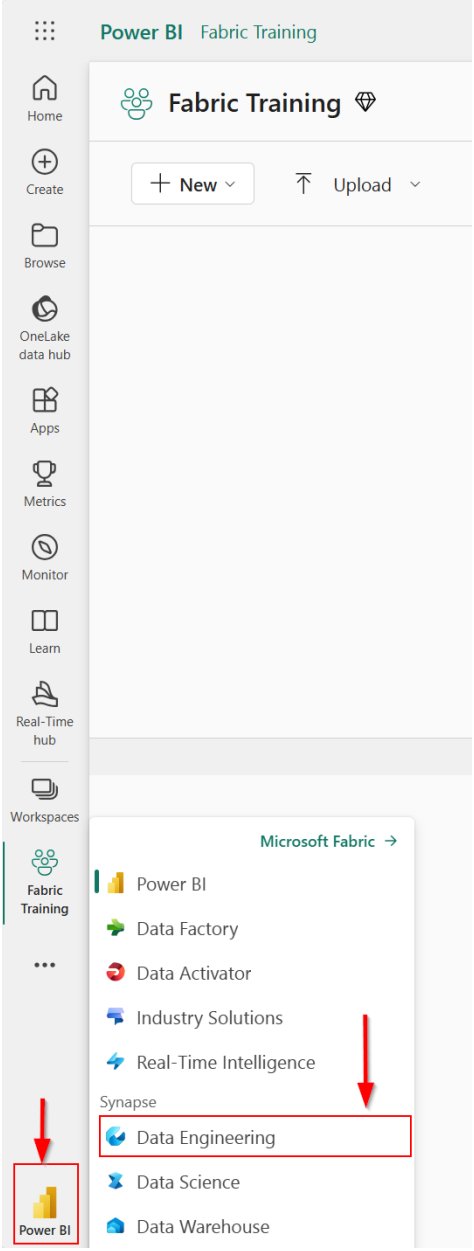


If the workspace was not created, repeat the steps in this task to create the workspace.

Task 2: Create Fabric Lakehouse

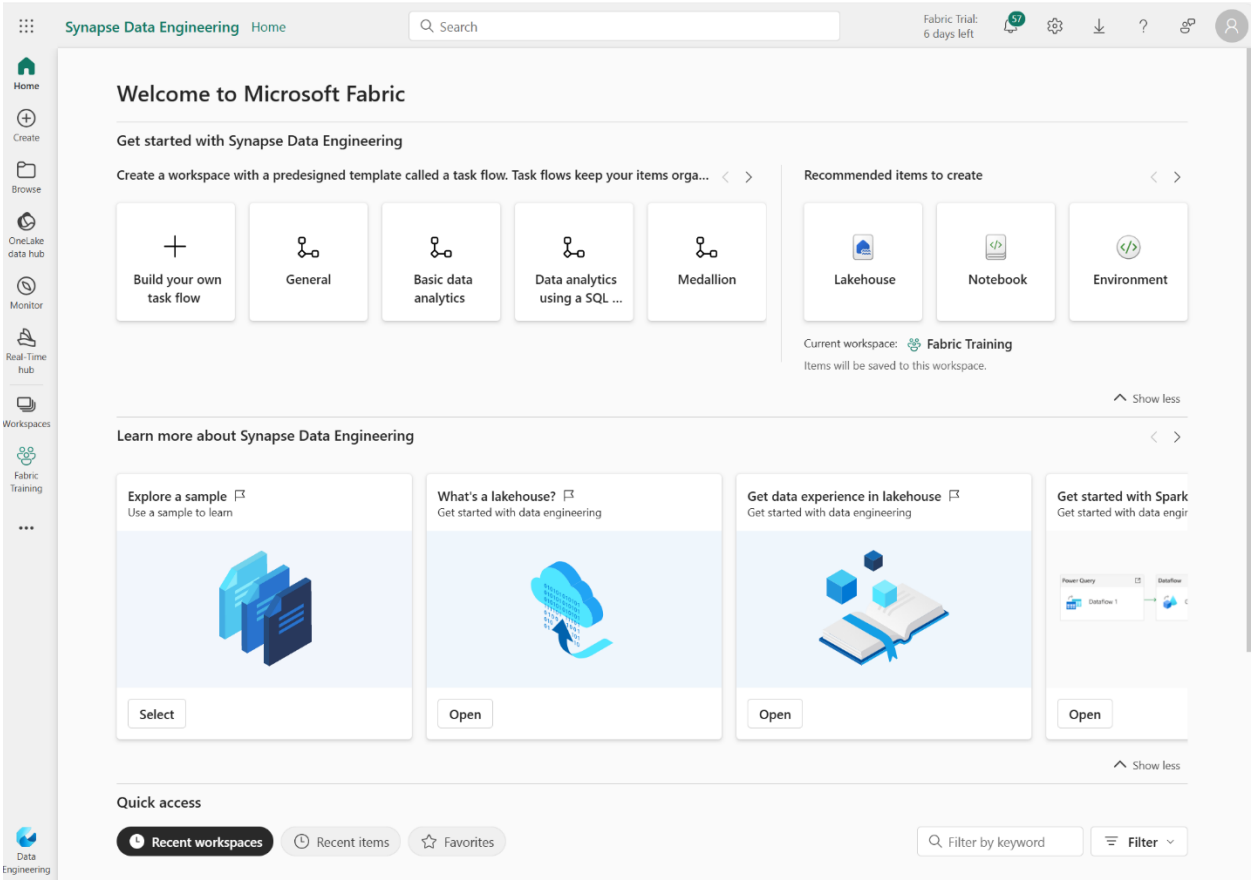
Now, let's see how each department in Contoso can easily create a Lakehouse in their workspace without any provisioning by simply providing the name, given the proper access rights, of course!

1. In the lower left of the navigation pane for the workspace, select **Power BI**. Then, in the Synapse section, select **Data Engineering**.

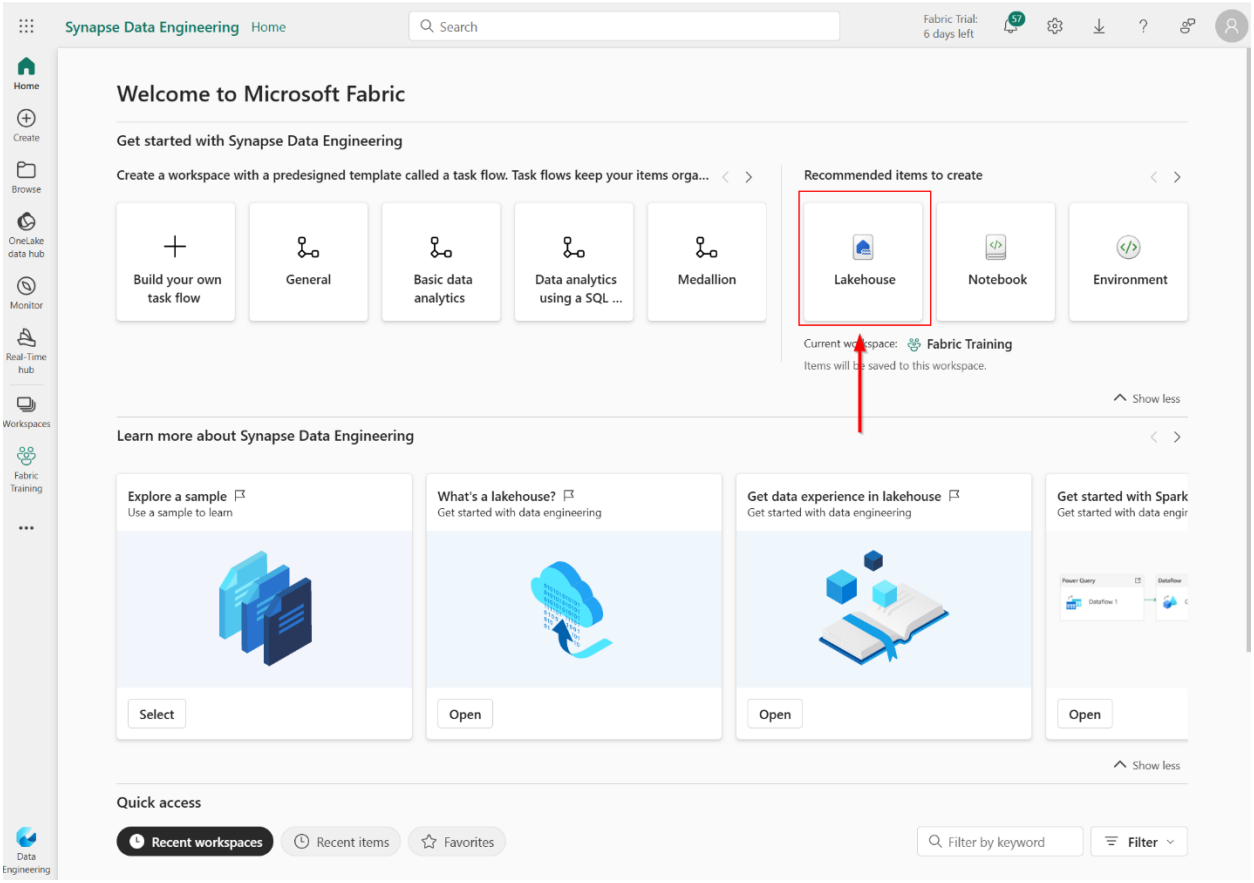


Selecting Data Enginering switches the context from reports and visualizations to data engineering.

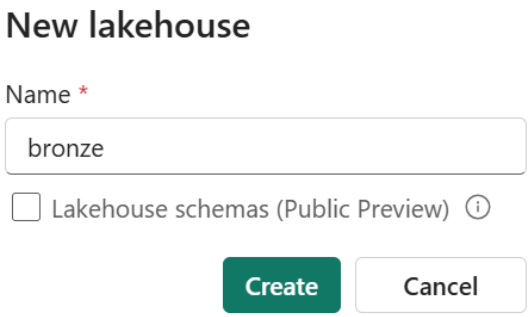
You should see that the options presented in the left navigation pane change to reflect the change in context.

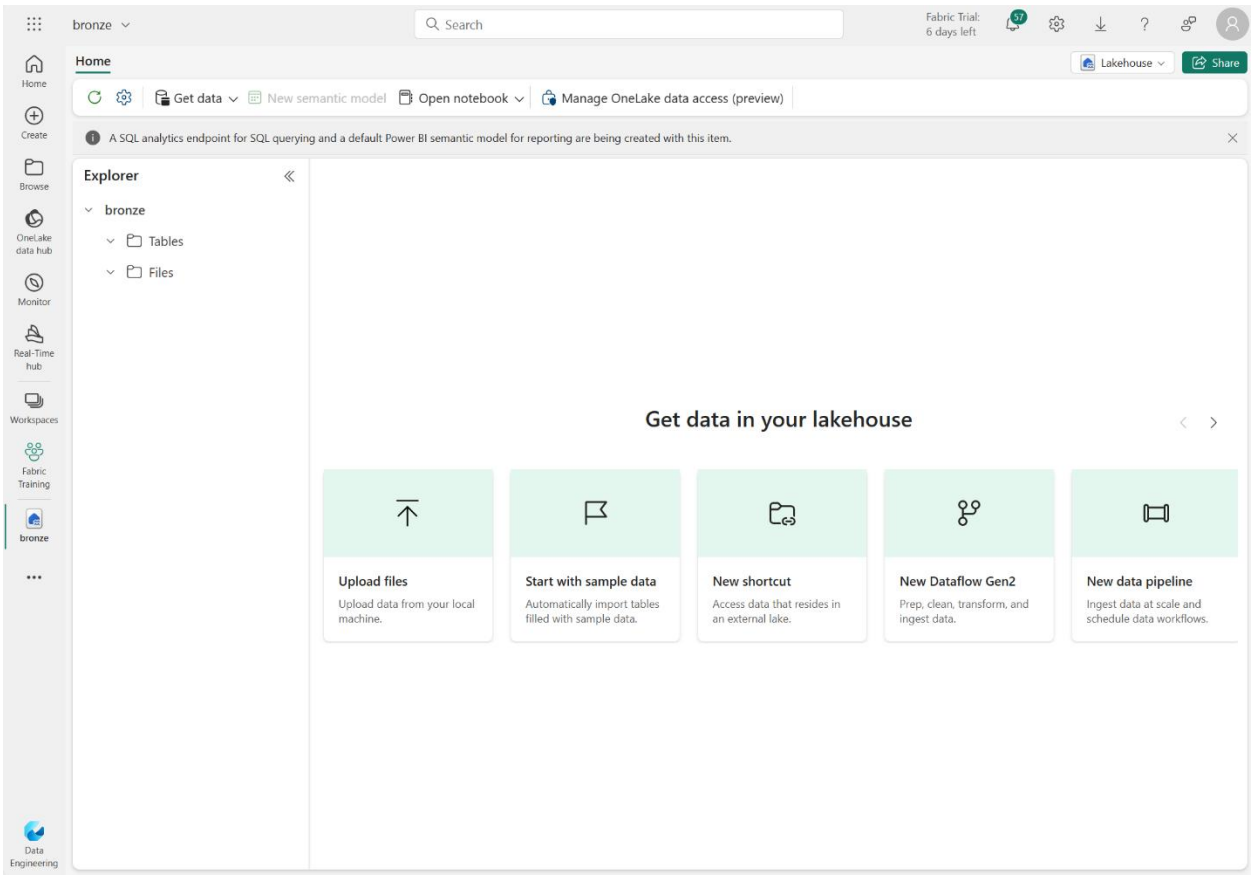


2. On the Synapse Data Engineering Home page, select the **Lakehouse** tile.



3. In the Name field, enter **bronze** and then select **Create**. The context switches to the bronze page.

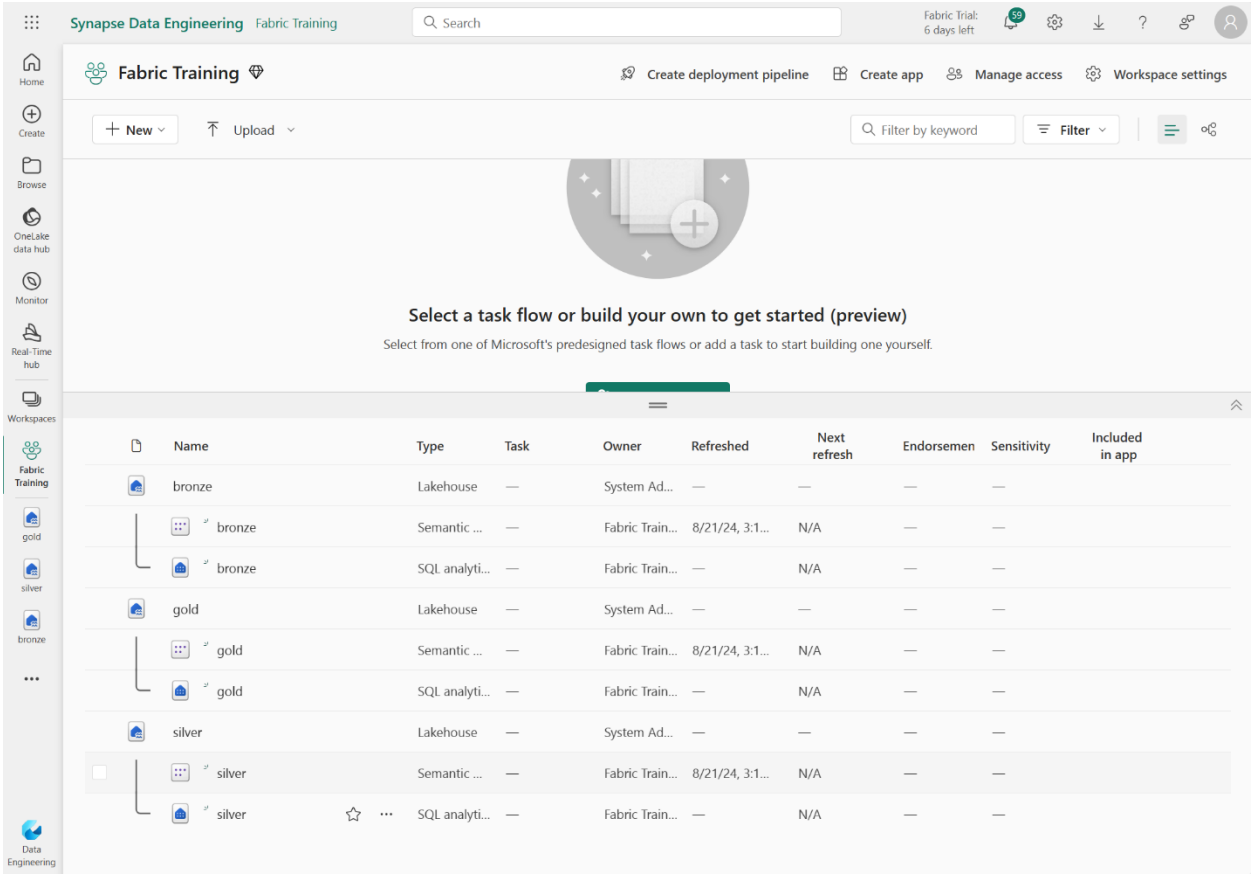




4. Repeat steps 2-3 to create two more lakehouses. You can either go to Home on the left pane and follow the same steps or go to your workspace and then click on New. Use the following names for the new lakehouses:

Lakehouse Name
silver
gold

5. In the left navigation pane for the Synapse Data Engineering Home page, select your FabricTraining workspace. You should see the three lakehouses and related objects listed.



Now you are ready to start data ingestion. You will first ingest the raw data in the bronze layer. After that, you will curate and enrich the data as you move the data to the silver and then gold layer.

Task 3: Create a data pipeline

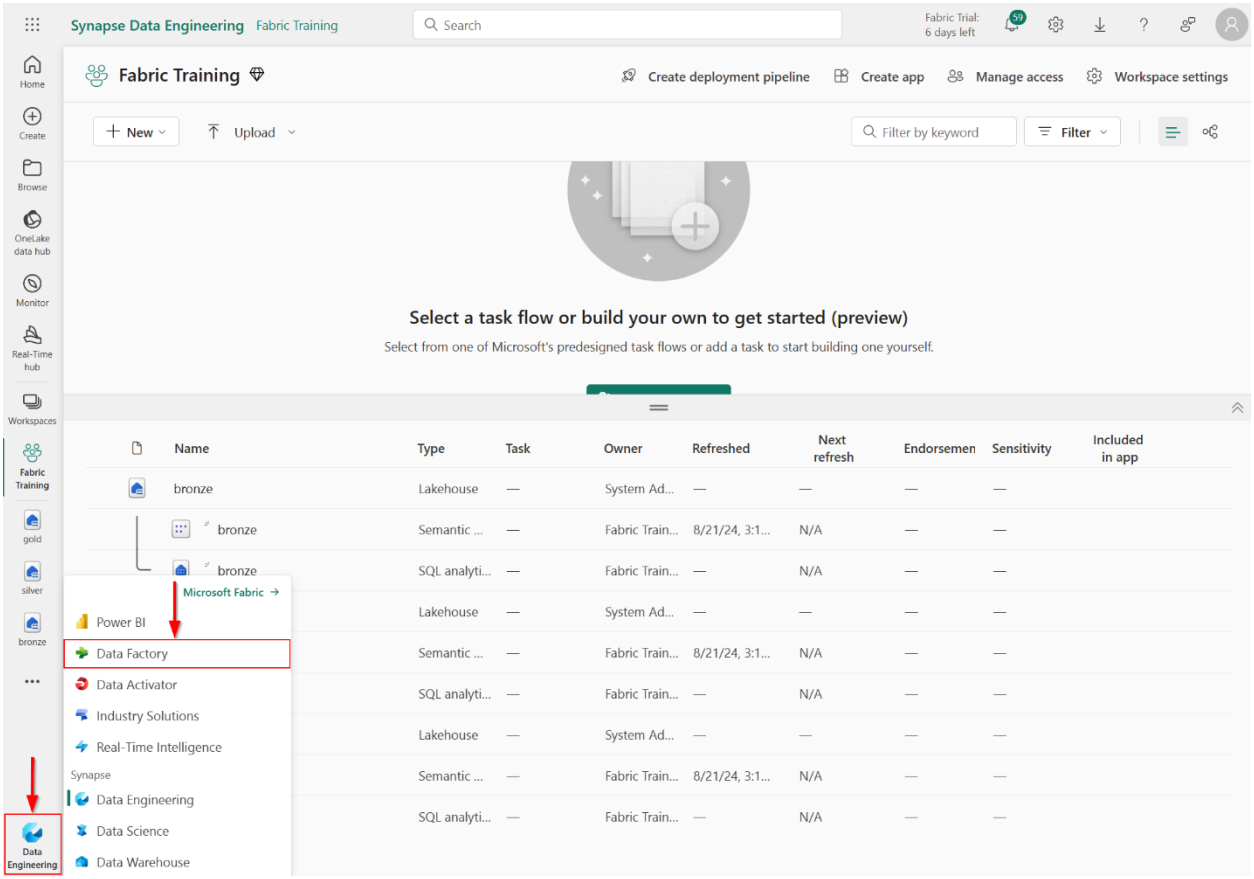
Option 1: If you have created an Azure SQL database

There are multiple ways to ingest data into a Lakehouse, and in this exercise, Contoso focuses on using data pipelines

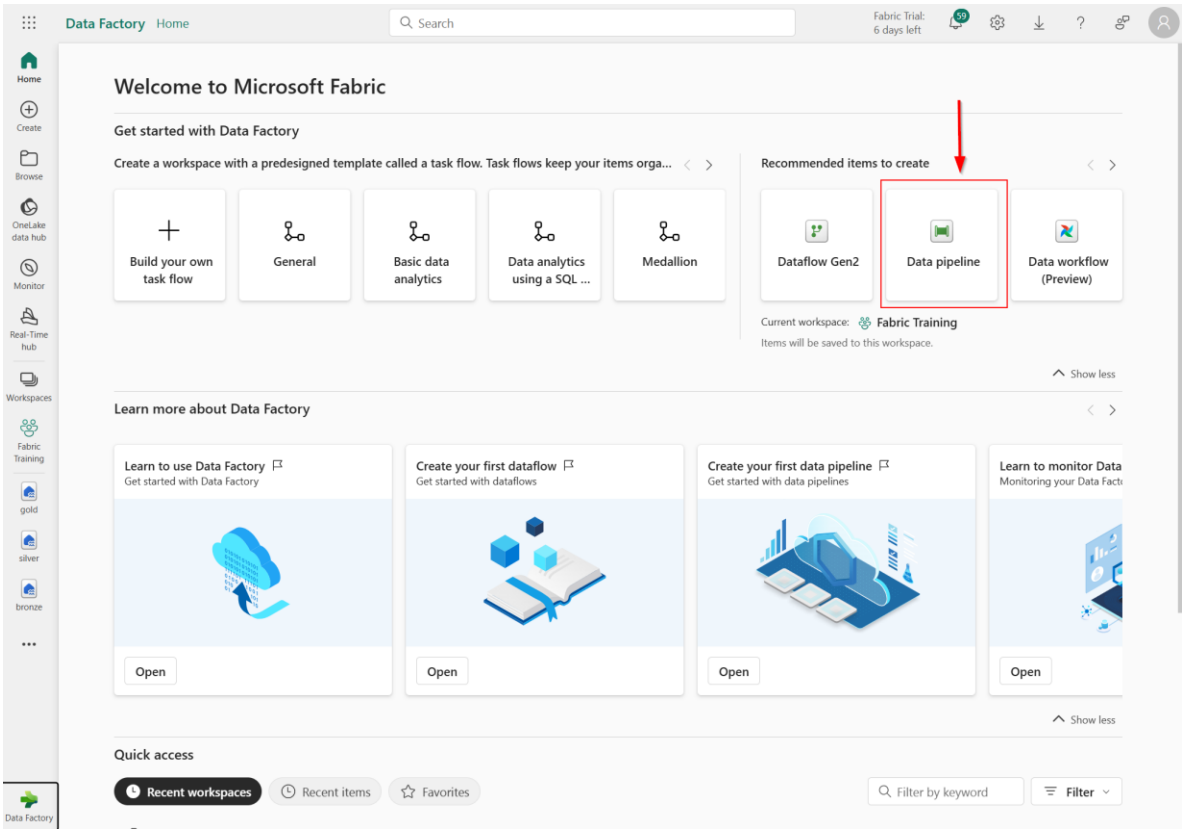
to efficiently funnel diverse datasets into their system, setting the stage for advanced analytics and insights.

Please move to option 2 if you have downloaded the data from Sharepoint and you are not using Azure SQL DB.

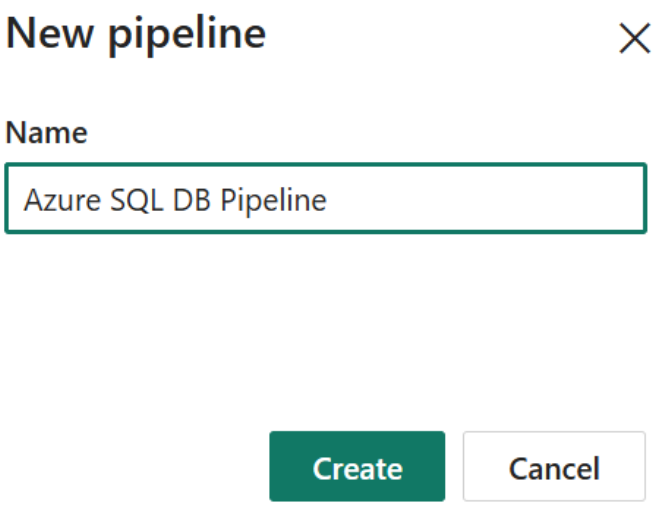
1. In the lower left of the navigation pane for the workspace, select **Data Engineering** and then select **Data Factory**.



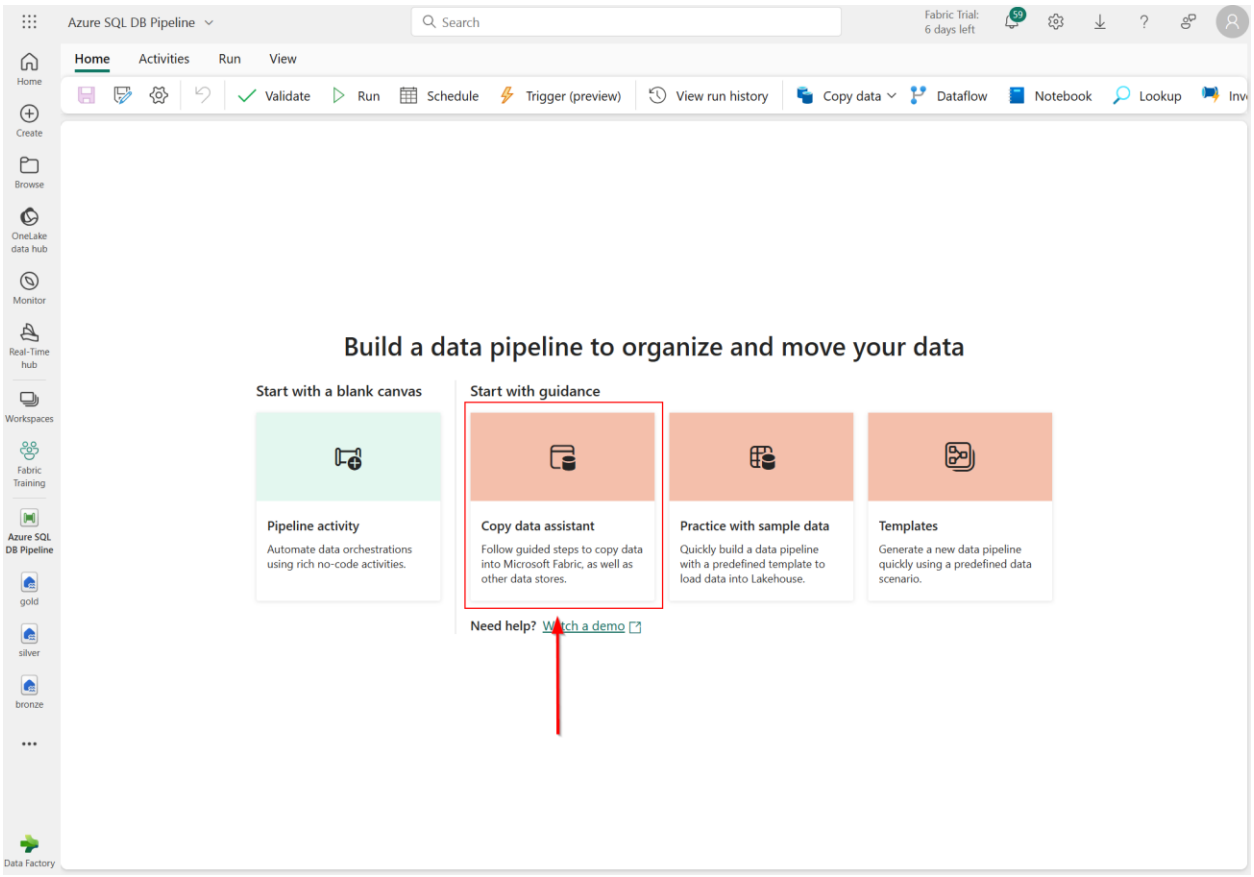
2. Select the **Data pipeline** tile.



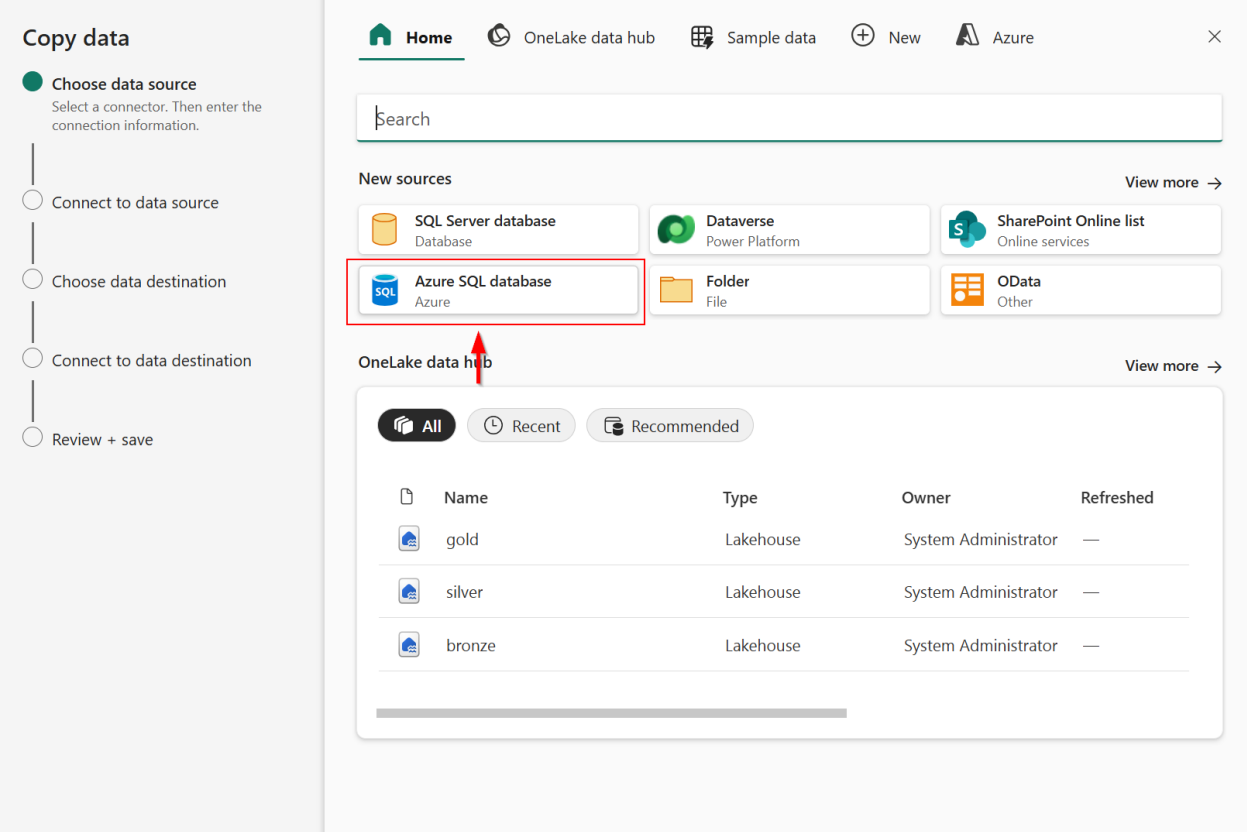
3. Enter **Azure SQL DB Pipeline** for the pipeline name and then select **Create**.



4. Select the **Copy data assistant** tile.

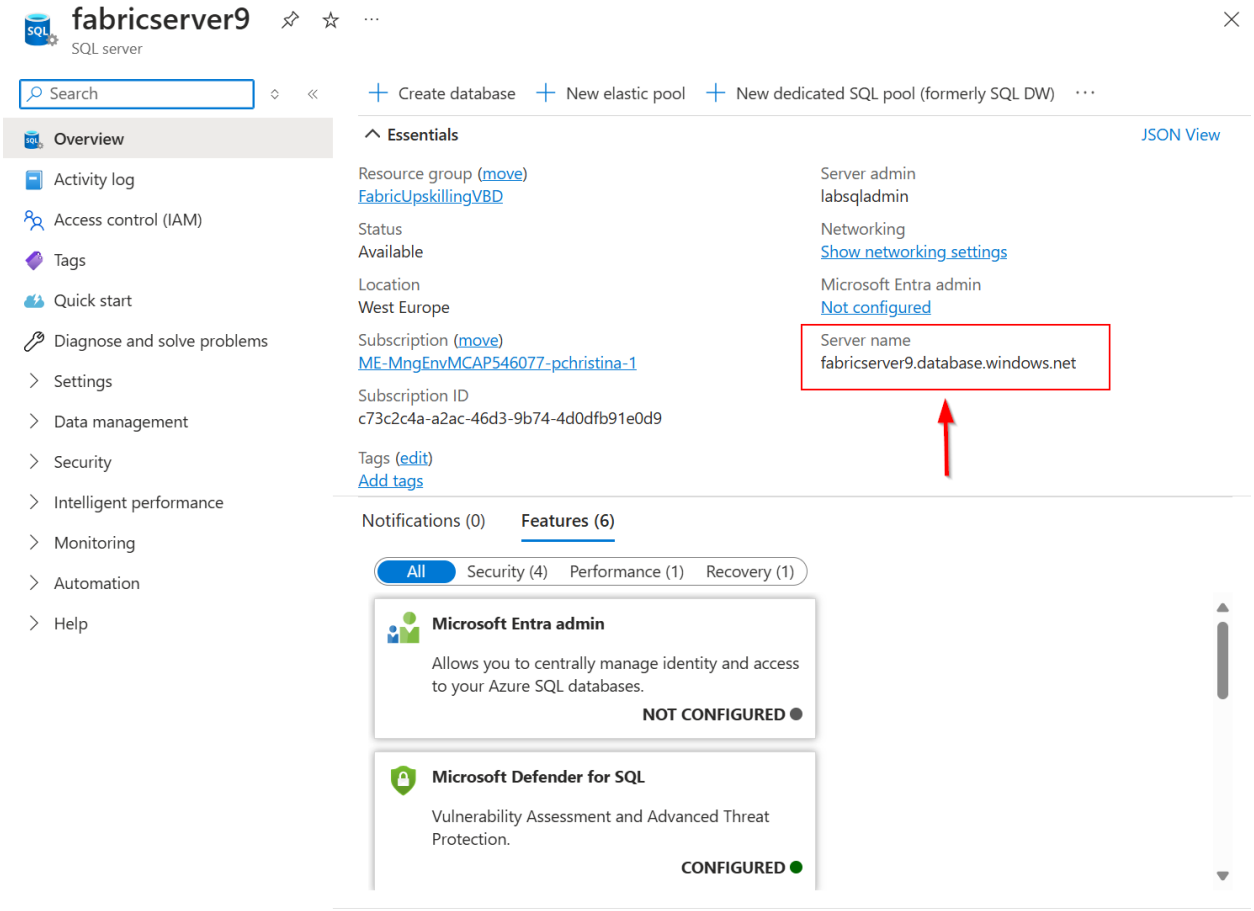


5. On the Choose data source page, select **Azure SQL Database**. You may need to scroll down to see the Azure SQL Database option.

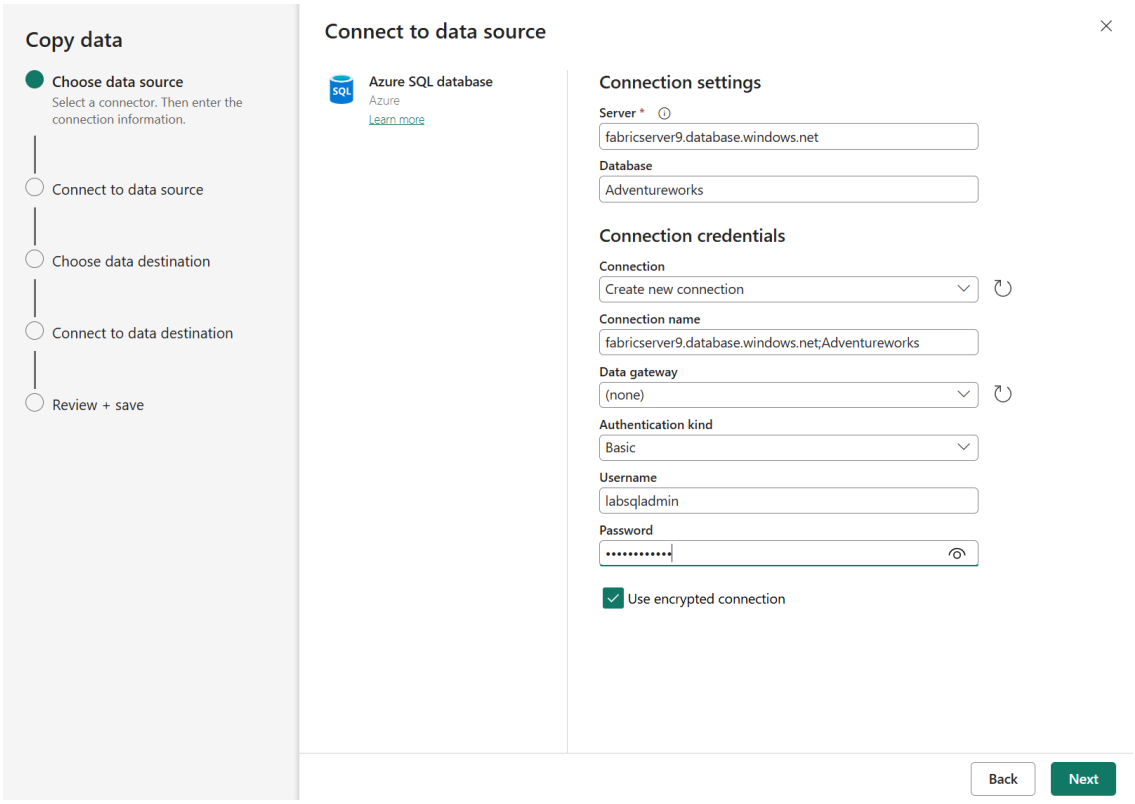


6. Configure the connection by using the values you configured for your Azure SQL db in Lab 0. Leave all other settings at their default values.

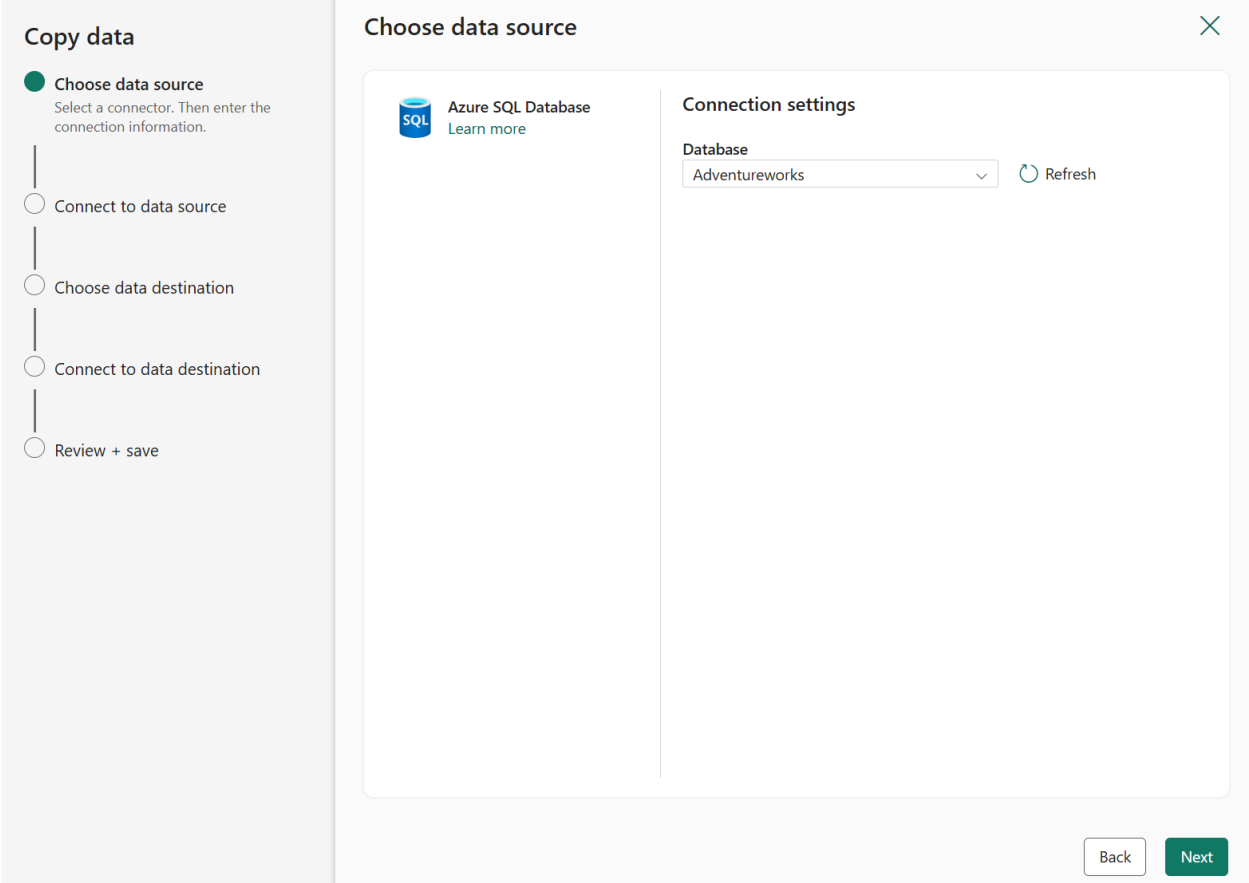
   In case you have used a different server name in Lab 0, please find the server's name in the server overview information in the Azure portal:



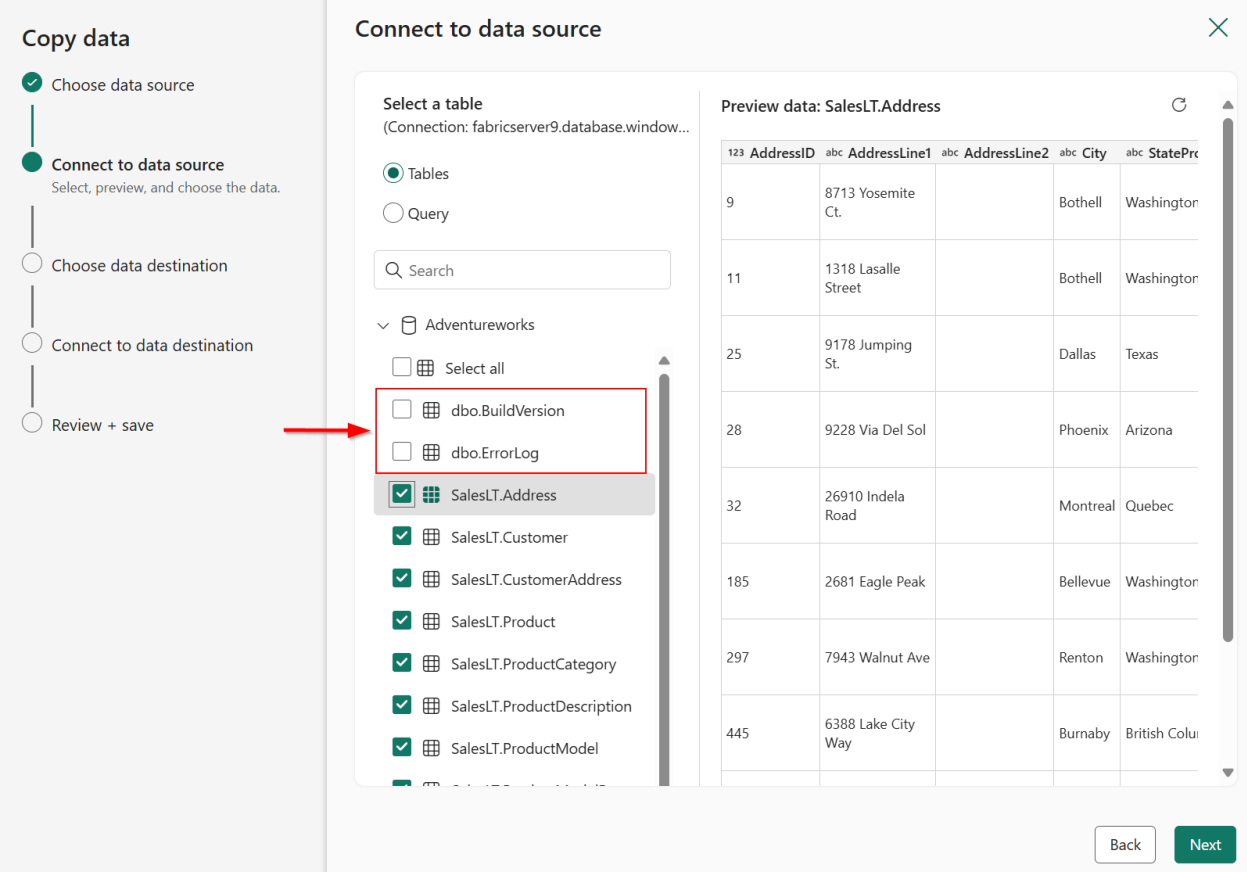
Default	Value
Server	**fabricserver9.database.windows.net** **Or the server name you chose in Lab 0**
Database	**Adventureworks**
Authentication kind	**Basic**
Username	**labsqladmin**
Password	**The password you chose in Lab 0**



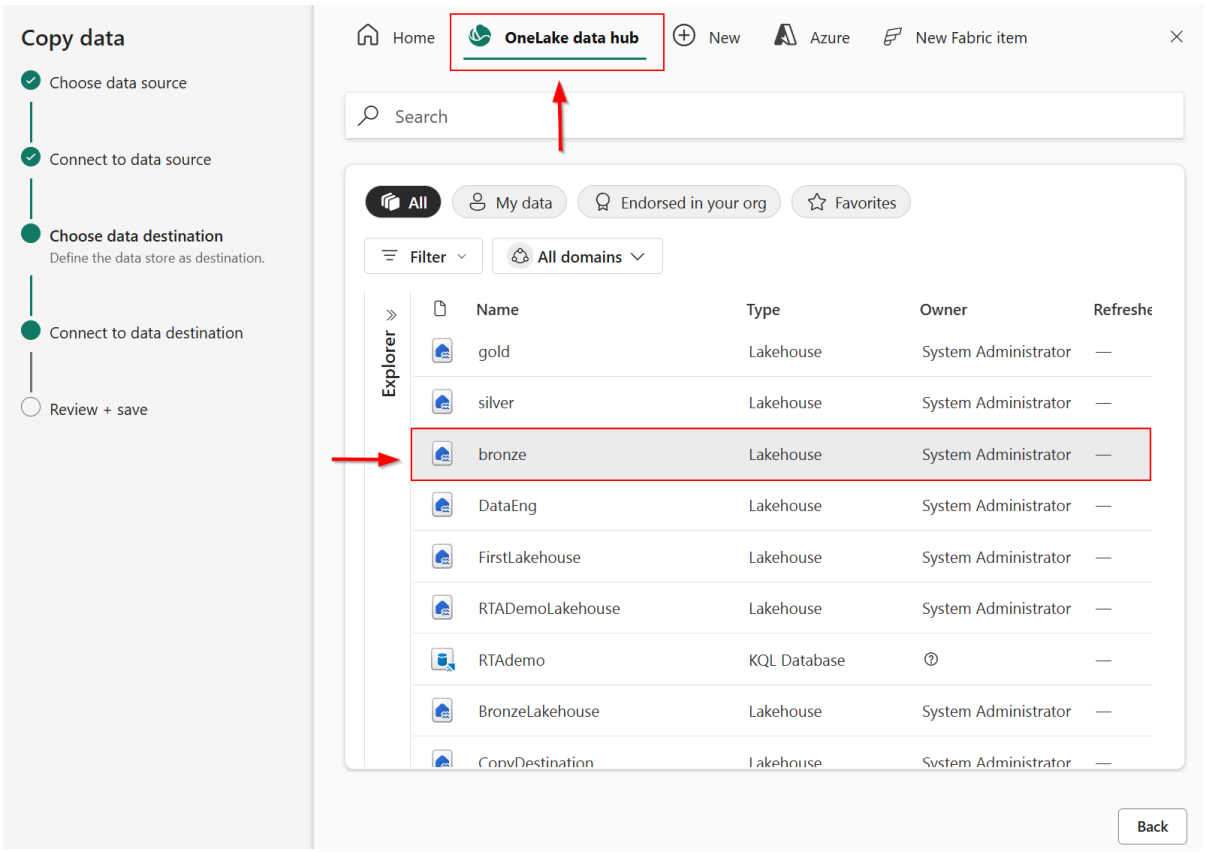
7. Select **Next**. Close any pop-up windows that display and wait for the connection to be created.
8. If you get the following window, choose the Adventureworks database – it may take some seconds for it to show up:



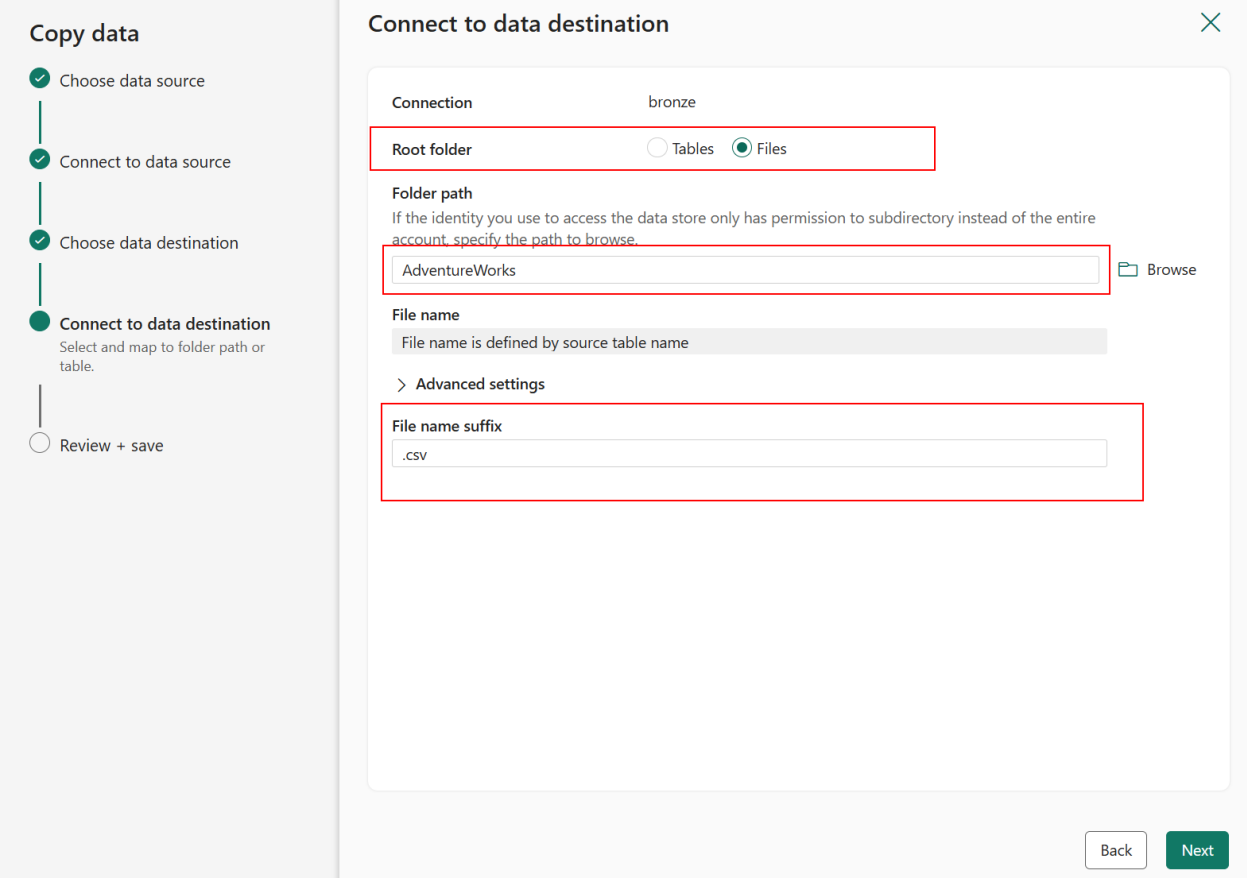
9. On the Connect to data source dialog, in the Select a table section, select **Tables**.
10. Select **Select all**. Clear the **dbo.BuildVersion** and **dbo.ErrorLog** checkboxes and select **Next**.



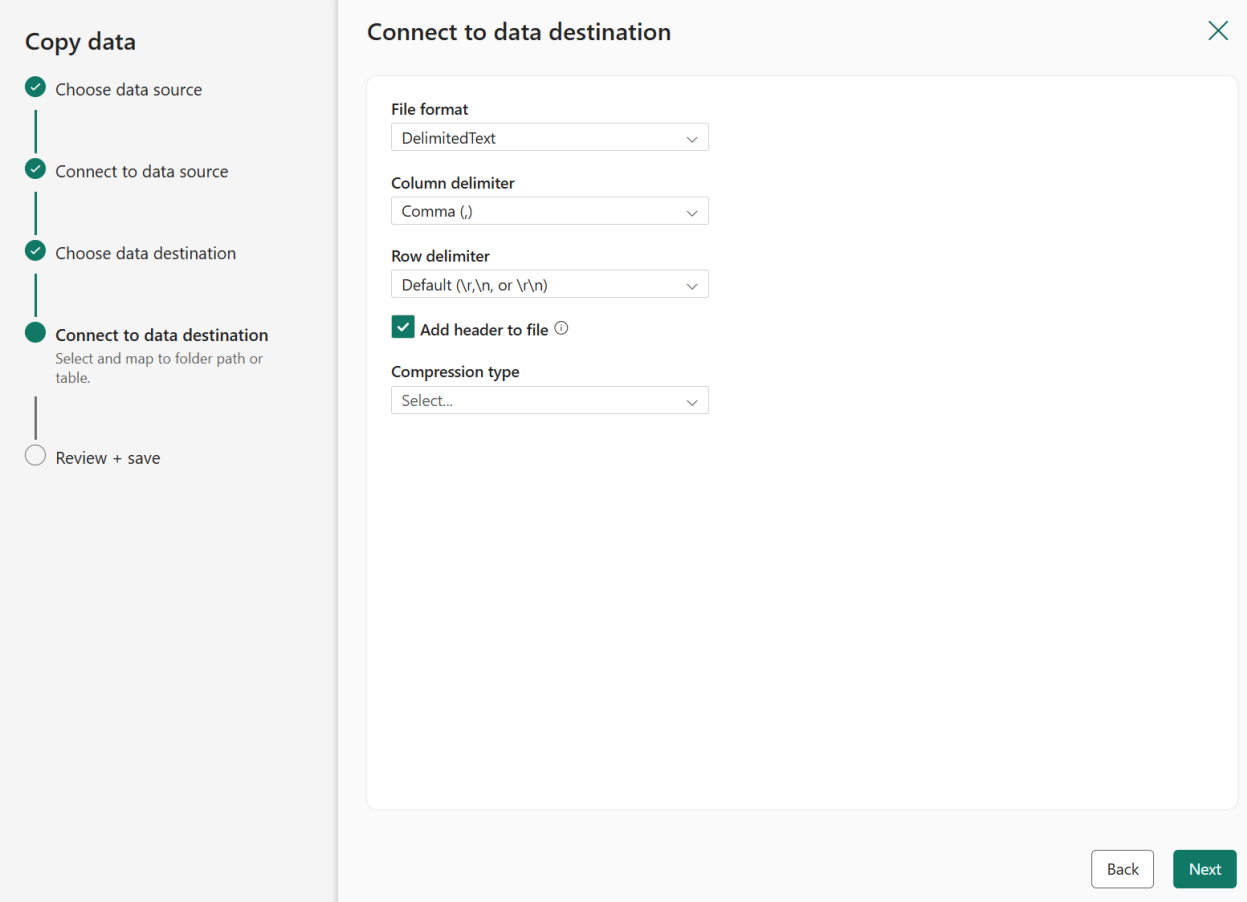
11. On the Choose data destinations page, go to OneLake data hub and choose the bronze lakehouse you created earlier:



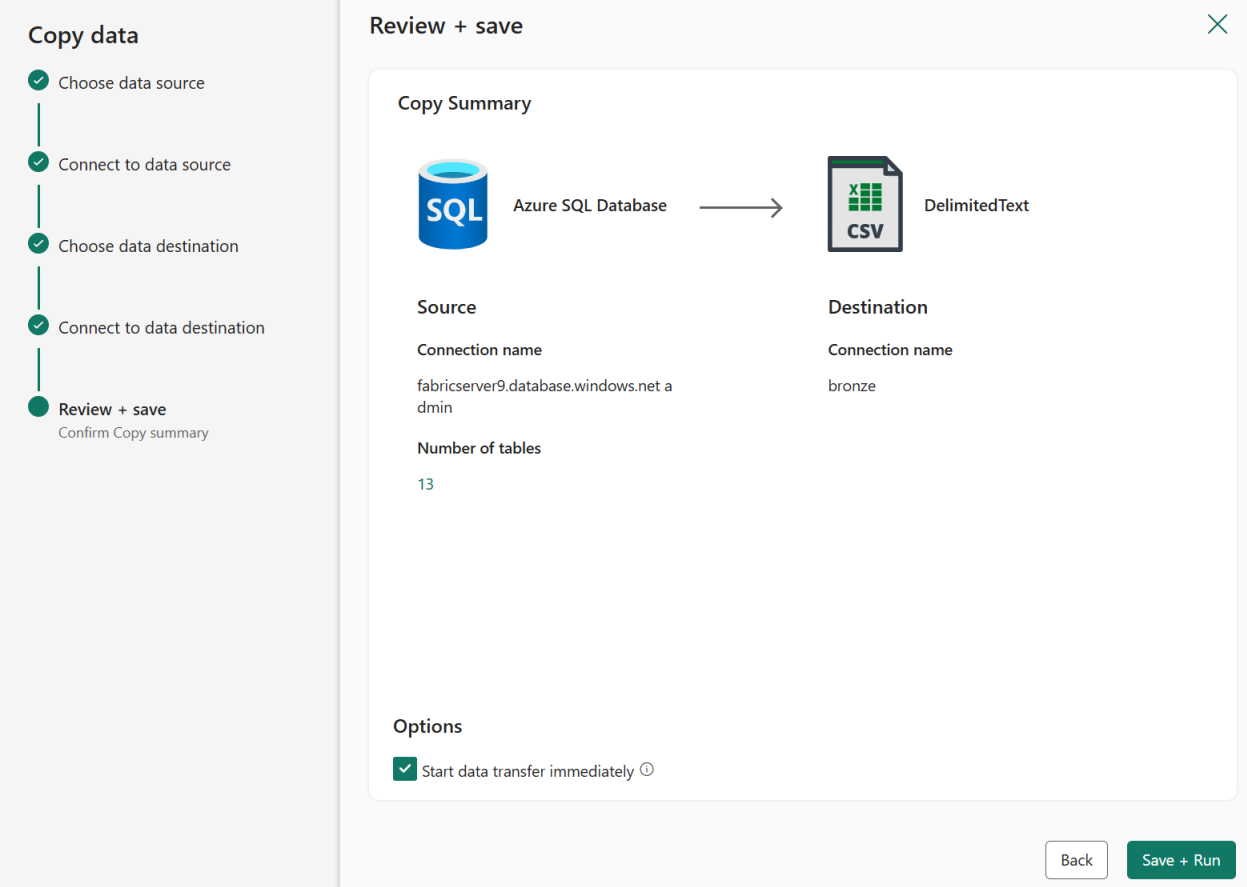
12. After you click on the bronze lakehouse, make sure to choose Files as root Folder and write AdventureWorks as File Path , Filename suffix as .csv. Click Next:



13. Make sure delimited text is selected here and click Next:

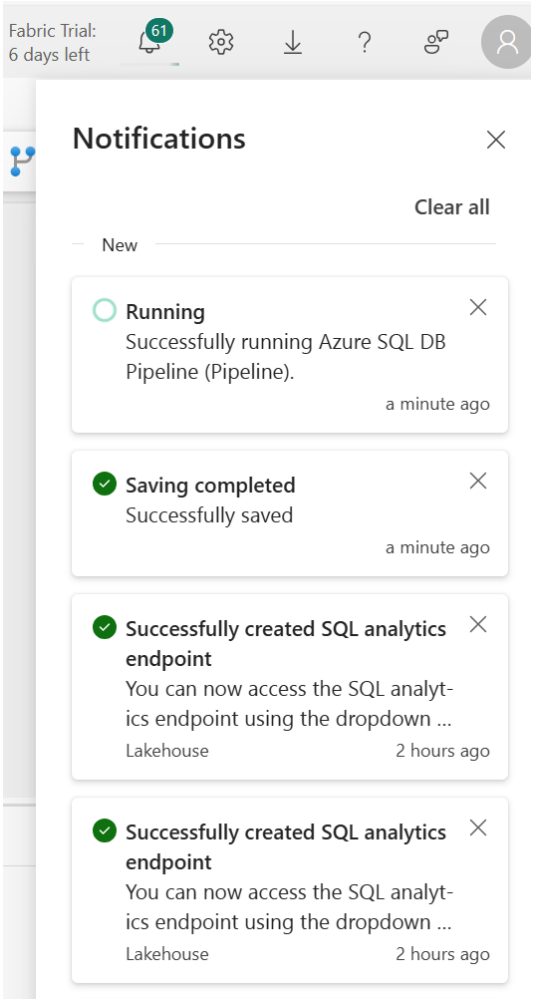


14. Select **Next** and then Select **Save + Run**.

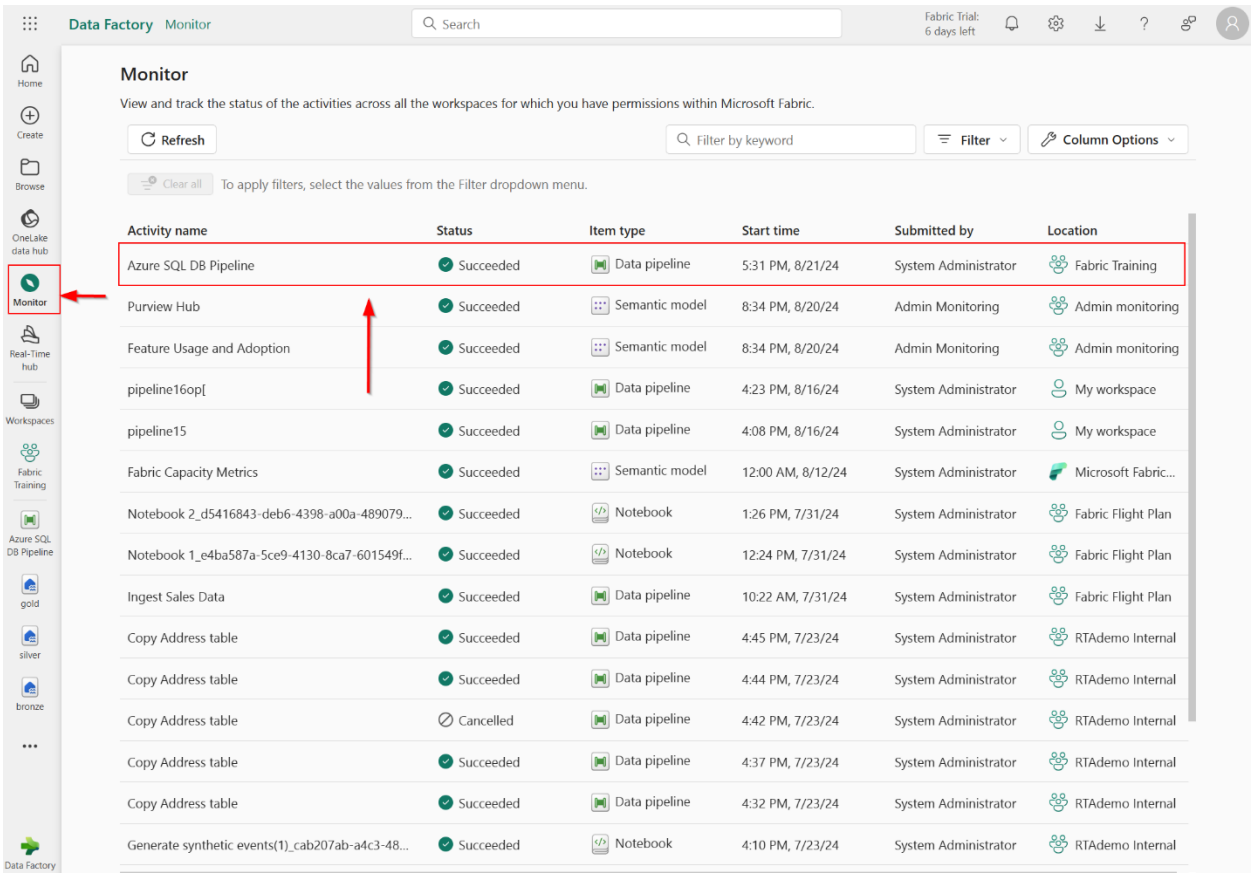


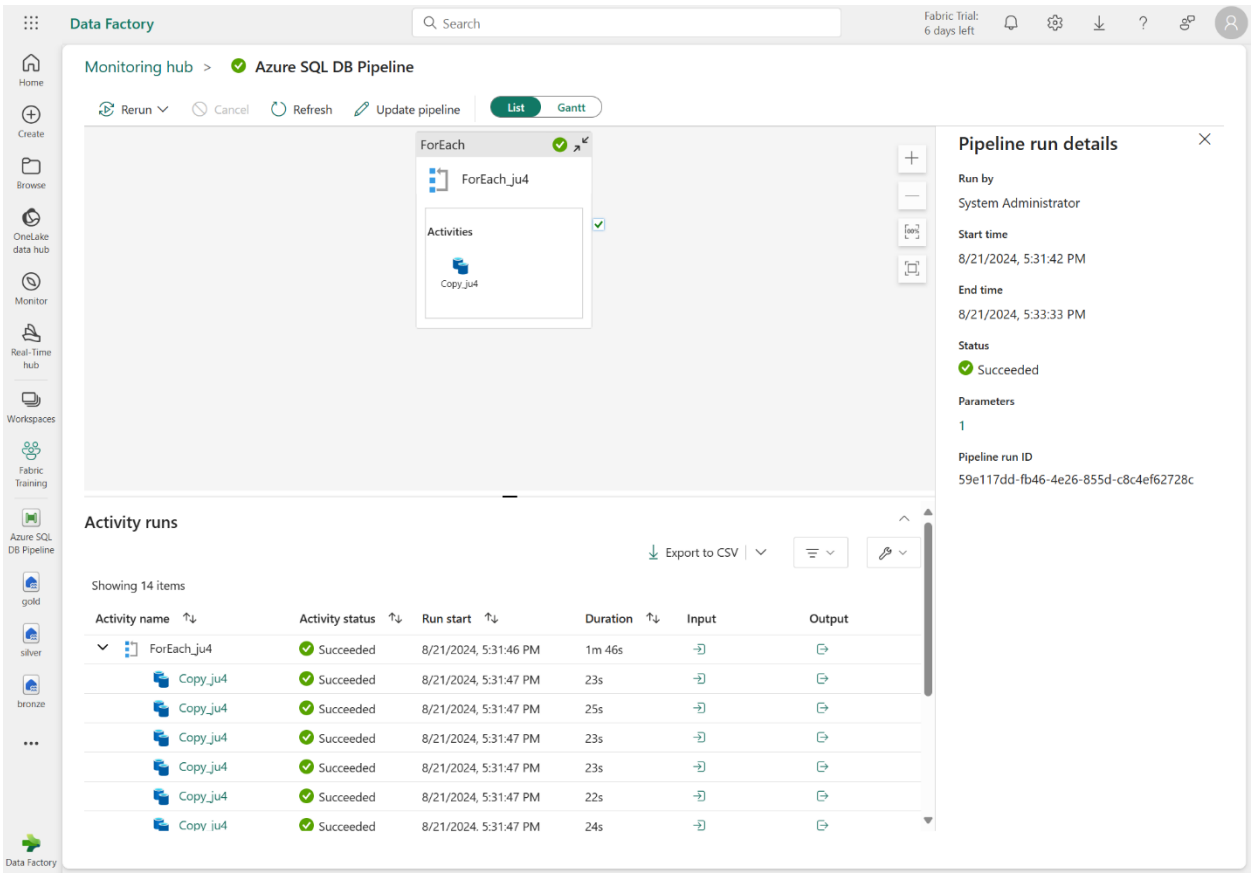
15. After a brief delay, the Pipeline Run window displays. Click **OK.** The pipeline will start processing.

16. On the upper-right of the page, select **Notifications**. You can use the Notifications area to monitor the pipeline. When the pipeline is completed, you will receive a notification there.

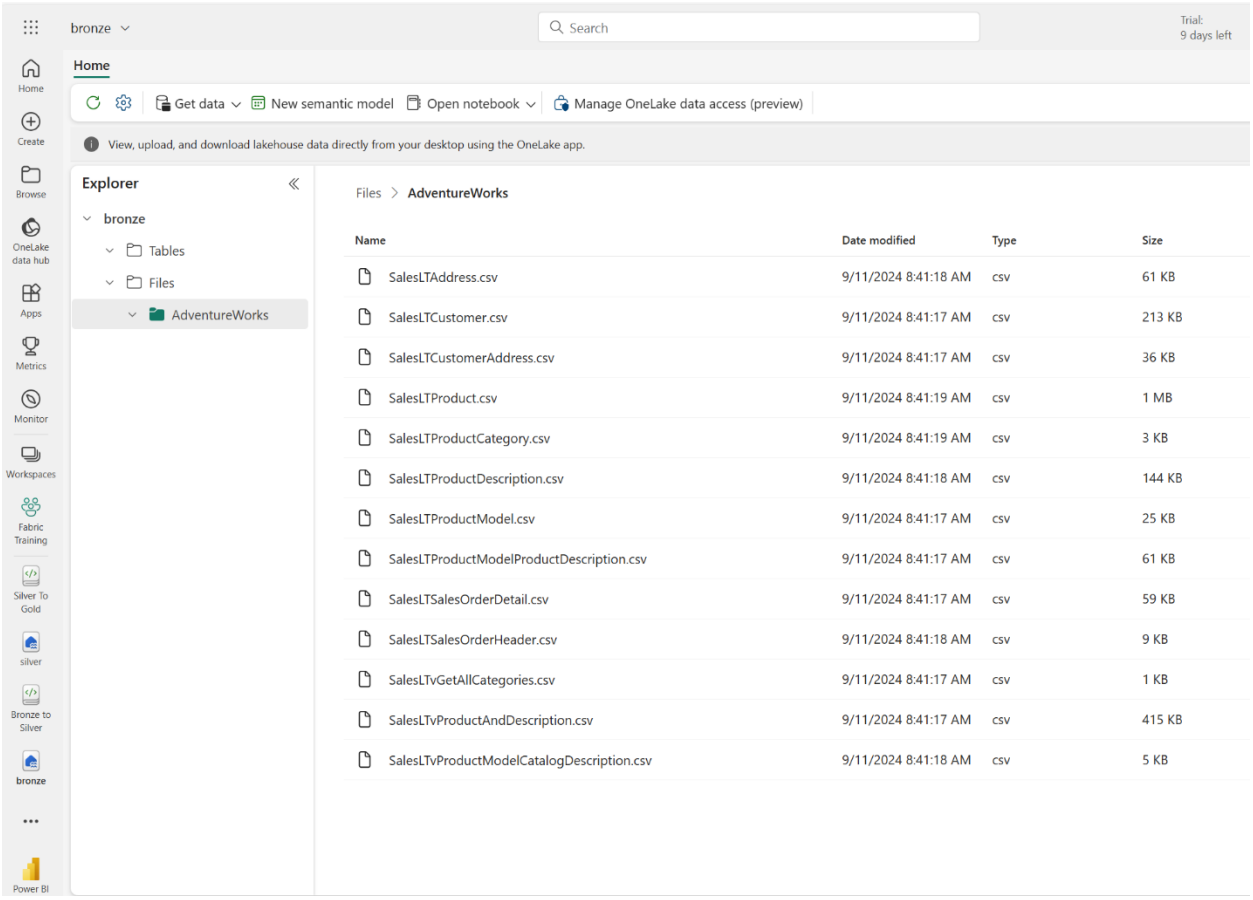


17. You can go to the **Monitor tab** on the left, choose the pipeline and see more information about the pipeline run:


Data Factory Monitor
Search
Fabric Trial: 6 days left
Home
Create
Browse
OneLake data hub
Monitor
Real-Time hub
Workspaces
Fabric Training
Azure SQL DB Pipeline
gold
silver
bronze
Data Factory
Monitor
View and track the status of the activities across all the workspaces for which you have permissions within Microsoft Fabric.
Refresh
Filter by keyword
Filter
Column Options
Clear all
To apply filters, select the values from the Filter dropdown menu.
Activity name | Status | Item type | Start time | Submitted by | Location
Azure SQL DB Pipeline | Succeeded | Data pipeline | 5:31 PM, 8/21/24 | System Administrator | Fabric Training
Purview Hub | Succeeded | Semantic model | 8:34 PM, 8/20/24 | Admin Monitoring | Admin monitoring
Feature Usage and Adoption | Succeeded | Semantic model | 8:34 PM, 8/20/24 | Admin Monitoring | Admin monitoring
pipeline16op[| Succeeded | Data pipeline | 4:23 PM, 8/16/24 | System Administrator | My workspace
pipeline15 | Succeeded | Data pipeline | 4:08 PM, 8/16/24 | System Administrator | My workspace
Fabric Capacity Metrics | Succeeded | Semantic model | 12:00 AM, 8/12/24 | System Administrator | Microsoft Fabric...
Notebook 2_d5416843-deb6-4398-a00a-489079... | Succeeded | Notebook | 1:26 PM, 7/31/24 | System Administrator | Fabric Flight Plan
Notebook 1_e4ba587a-5ce9-4130-8ca7-601549f... | Succeeded | Notebook | 12:24 PM, 7/31/24 | System Administrator | Fabric Flight Plan
Ingest Sales Data | Succeeded | Data pipeline | 10:22 AM, 7/31/24 | System Administrator | Fabric Flight Plan
Copy Address table | Succeeded | Data pipeline | 4:45 PM, 7/23/24 | System Administrator | RTAdemo Internal
Copy Address table | Succeeded | Data pipeline | 4:44 PM, 7/23/24 | System Administrator | RTAdemo Internal
Copy Address table | Cancelled | Data pipeline | 4:42 PM, 7/23/24 | System Administrator | RTAdemo Internal
Copy Address table | Succeeded | Data pipeline | 4:37 PM, 7/23/24 | System Administrator | RTAdemo Internal
Copy Address table | Succeeded | Data pipeline | 4:32 PM, 7/23/24 | System Administrator | RTAdemo Internal
Generate synthetic events(1)_cab207ab-a4c3-48... | Succeeded | Notebook | 4:10 PM, 7/23/24 | System Administrator | RTAdemo Internal



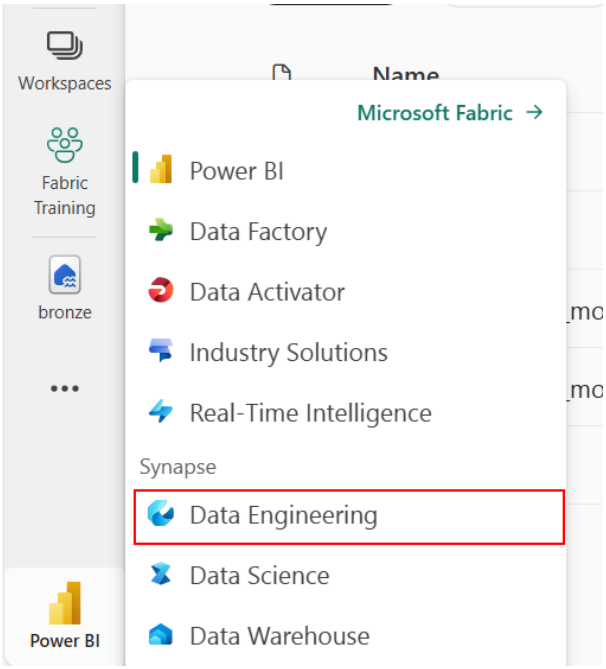
18. After the status shows **Succeeded**, your data has been transferred from Azure SQL Database to bronze lakehouse. You can go to your bronze lakehouse and verify that the files are moved:



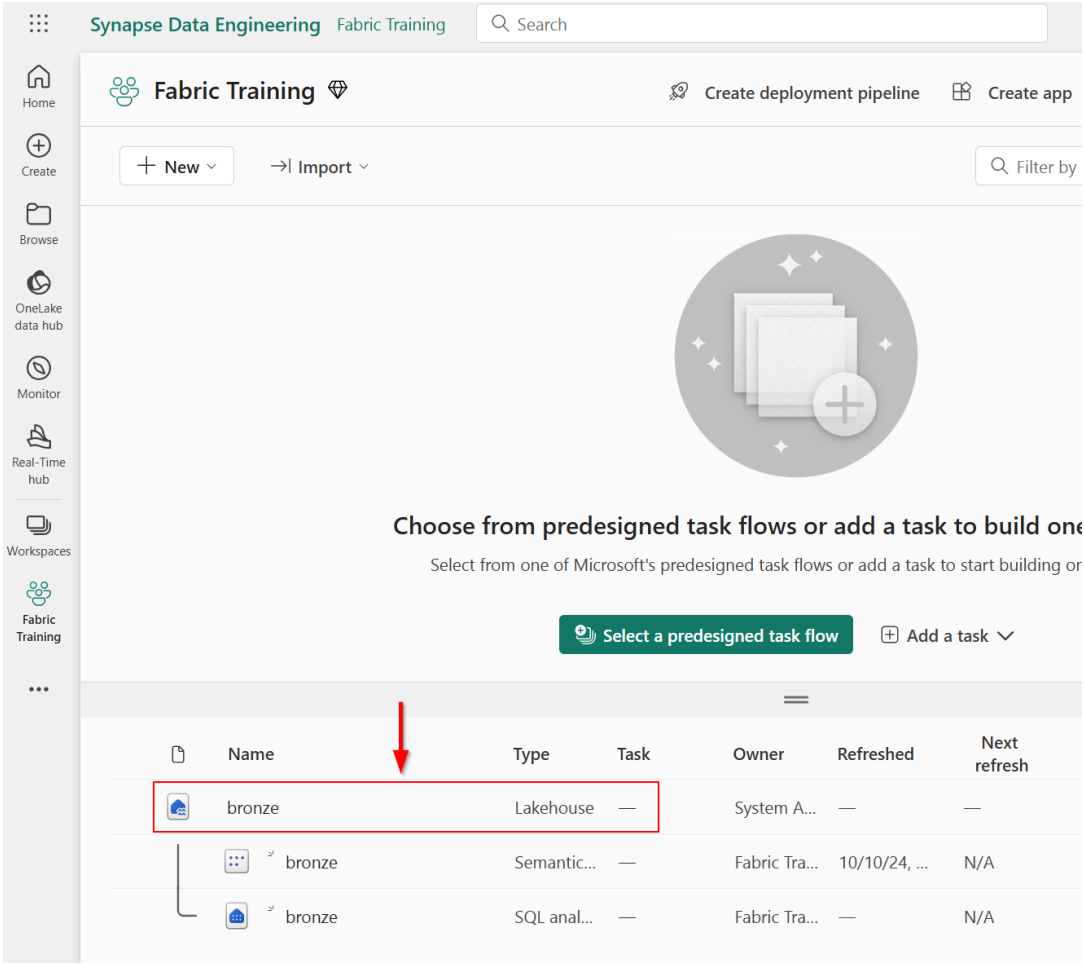
Option 2: If you have downloaded the data from Sharepoint

In case you have followed Lab 0 – Option 2 and you have downloaded the files, please follow the instructions here to upload them directly into the Fabric lakehouse:

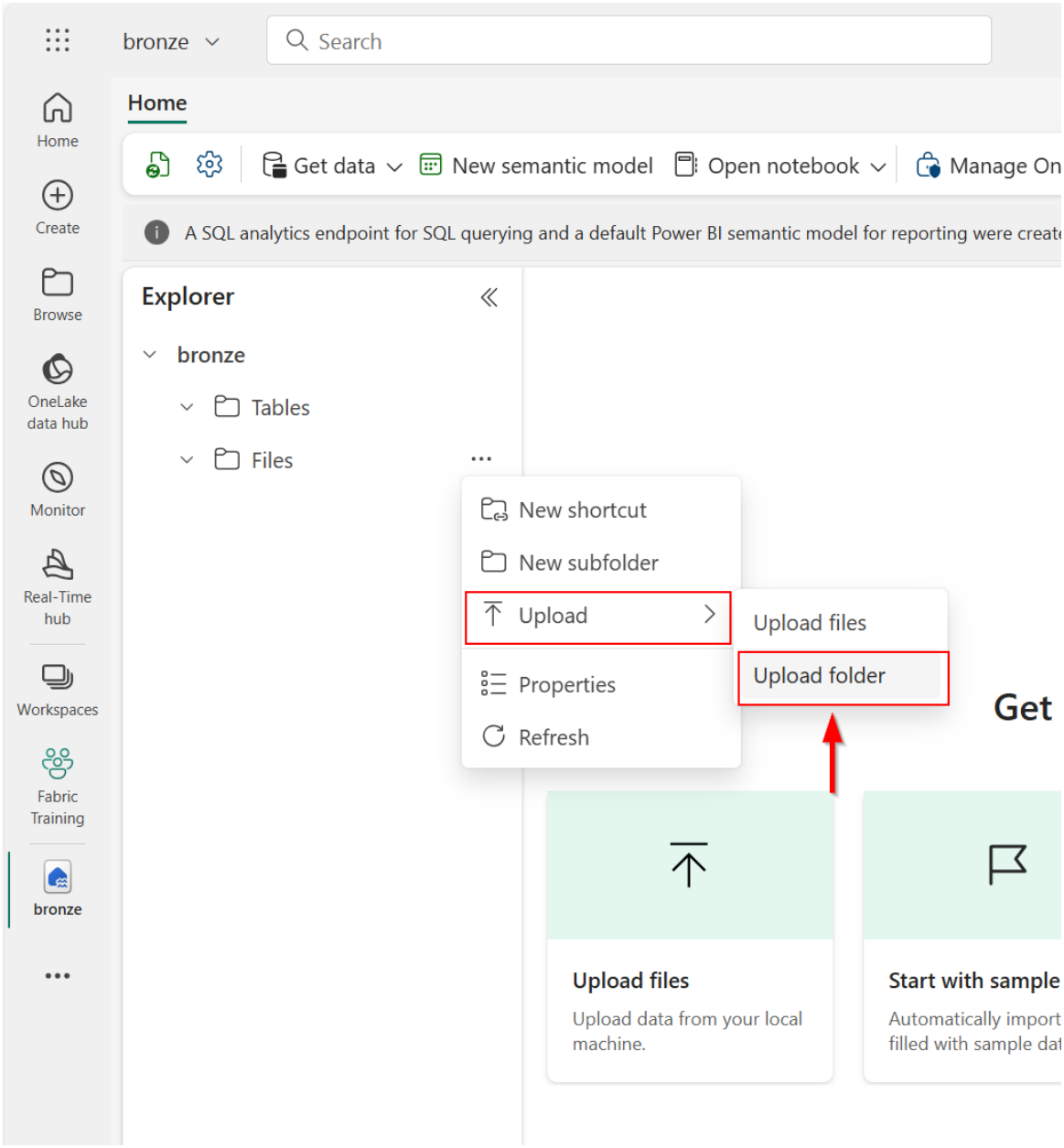
1. In the workspaces on the left bar select the Fabric Training one that you created before
2. Select Power BI at the bottom left and choose Data Engineering:



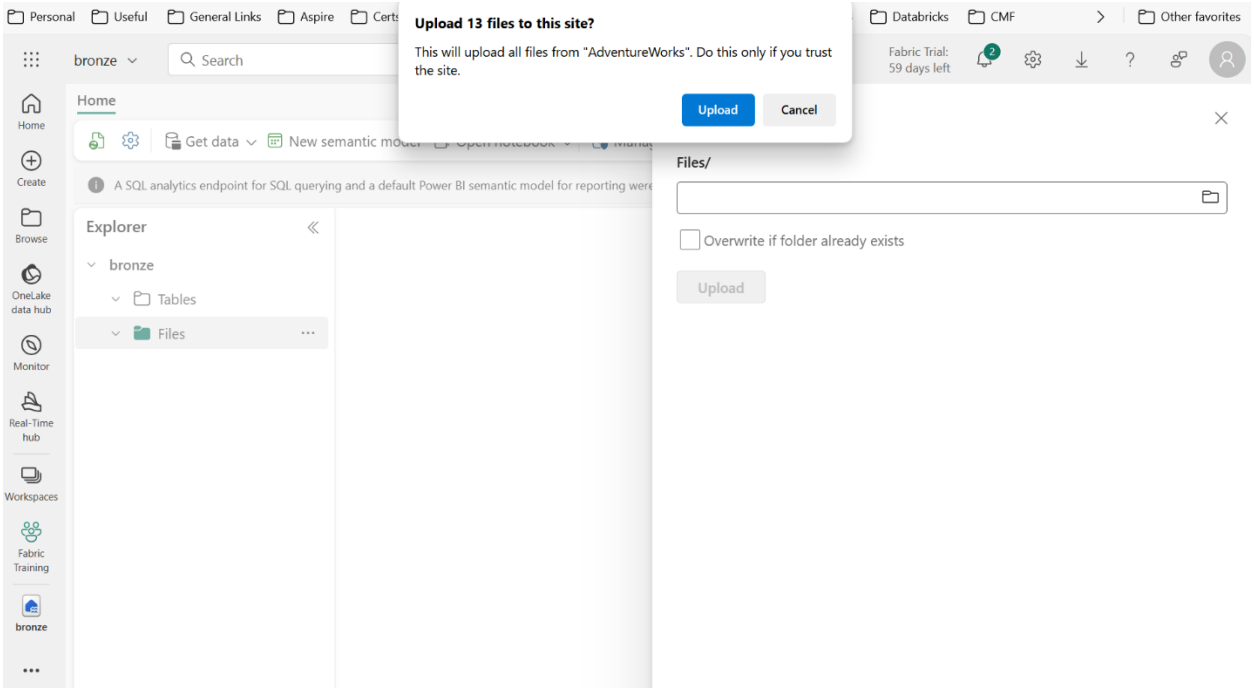
3. Then select the bronze lakehouse:



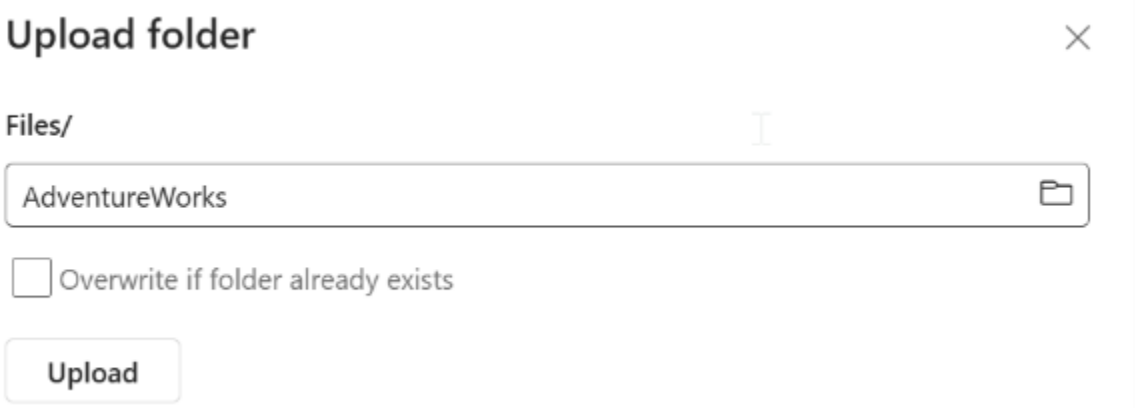
4. On the Bronze Lakehouse go to files, upload and hit upload folder.



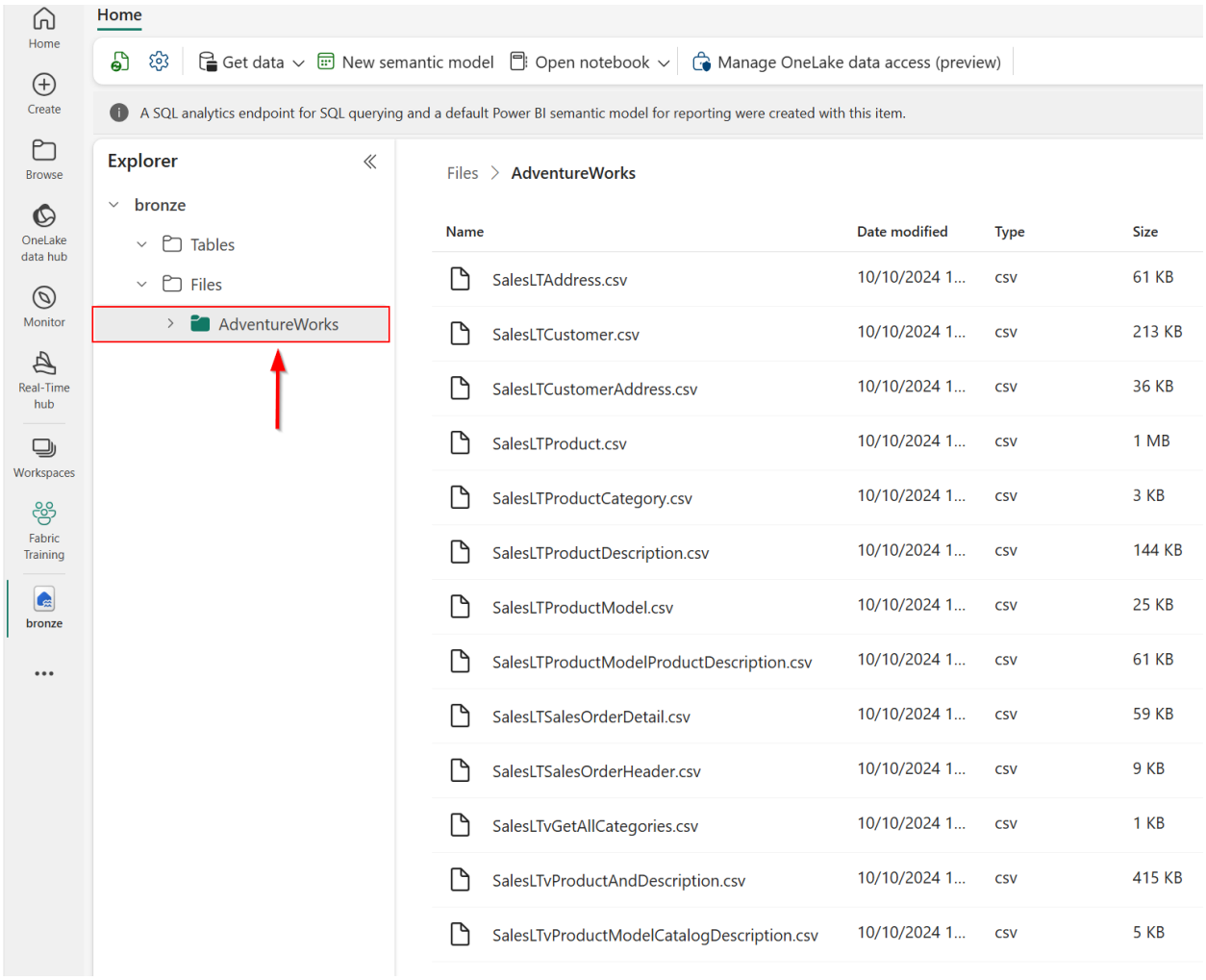
5. Go to the Downloads folder on your laptop and locate the folder you unzipped after downloading it from SharePoint. When prompted, select upload:



When uploading a folder to Bronze lakehouse it should look like this:



6. Under the files you should see the Folder AdventureWorks and the CSV files for every table:


Home
Create
Browse
OneLake data hub
Monitor
Real-Time hub
Workspaces
Fabric Training
bronze
Home
Get data
New semantic model
Open notebook
Manage OneLake data access (preview)
A SQL analytics endpoint for SQL querying and a default Power BI semantic model for reporting were created with this item.
Explorer
bronze
Tables
Files
AdventureWorks
Files > AdventureWorks
Name
Date modified
Type
Size
SalesLTAddress.csv 10/10/2024 1... csv 61 KB
SalesLTCustomer.csv 10/10/2024 1... csv 213 KB
SalesLTCustomerAddress.csv 10/10/2024 1... csv 36 KB
SalesLTProduct.csv 10/10/2024 1... csv 1 MB
SalesLTProductCategory.csv 10/10/2024 1... csv 3 KB
SalesLTProductDescription.csv 10/10/2024 1... csv 144 KB
SalesLTProductModel.csv 10/10/2024 1... csv 25 KB
SalesLTProductModelProductDescription.csv 10/10/2024 1... csv 61 KB
SalesLTSalesOrderDetail.csv 10/10/2024 1... csv 59 KB
SalesLTSalesOrderHeader.csv 10/10/2024 1... csv 9 KB
SalesLTvGetAllCategories.csv 10/10/2024 1... csv 1 KB
SalesLTvProductAndDescription.csv 10/10/2024 1... csv 415 KB
SalesLTvProductModelCatalogDescription.csv 10/10/2024 1... csv 5 KB