Lab 7 – Explore Fabric pipelines

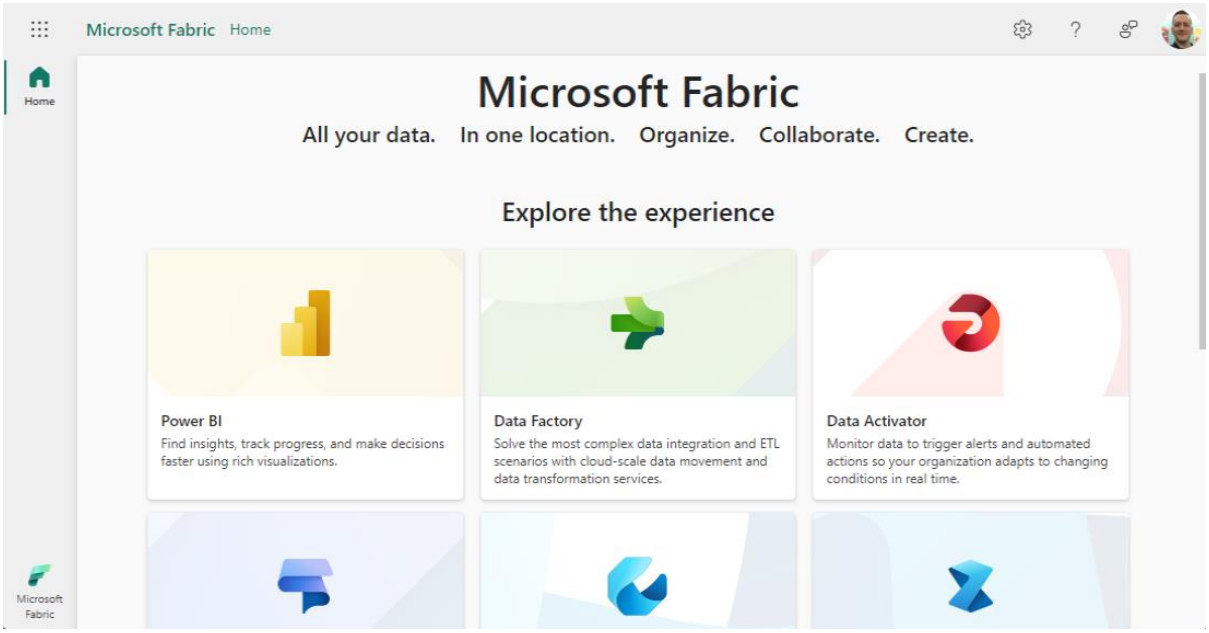
An Azure Synapse Analytics workspace gathers various data engineering and related tools together – including integration pipelines like those of ADF – but the experience is not seamless. Details of component services (for example the external data lake) remain visible and must be provisioned and managed by you, making Synapse a "platform as a service" (PaaS) offering.

In contrast, Microsoft Fabric is a *serverless* service, encapsulating all data processing capabilities and data storage – known as OneLake – as always-on, immediately available resources. Ingestion of data into the lake, and orchestration of transformation activities can be handled in Fabric's Data Factory experience. The purpose of this lab is to explore Data Factory pipelines in Fabric.

The lab requires access to a Microsoft Fabric tenant. If you do not have access to a tenant, you will be unable to complete the lab. If you are unsure whether you have access to a tenant, attempt Lab 7.1.

Lab 7.1 – Connect to a Fabric tenant

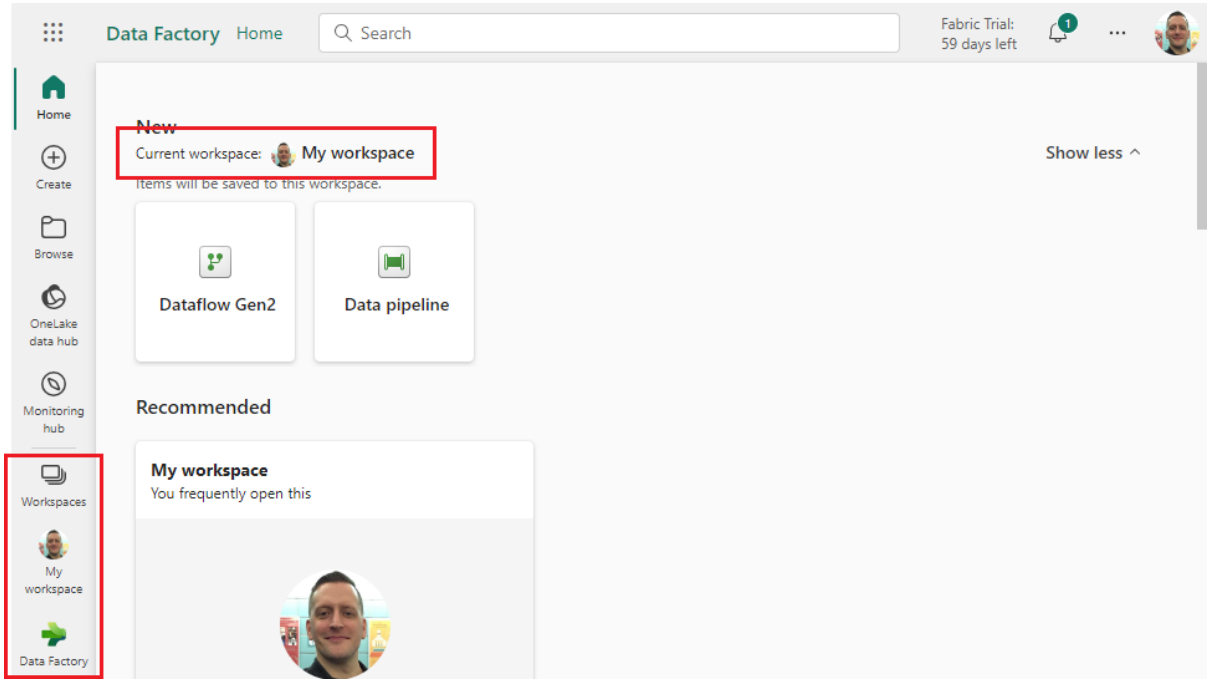
1. Connect to your Fabric tenant by browsing to https://app.fabric.microsoft.com/home.



   If your organisation has disabled its Fabric tenant, or if it has no Power BI subscription, you can connect using a personal developer account with an active Visual Studio Enterprise subscription. Unlike Azure resources, Fabric tenants do not need to be provisioned – you can simply connect to the service and go.

2. Select the "Data Factory" tile to access Fabric's Data Factory experience.

3. Microsoft Fabric resources are organised into workspaces. Ensure that your current workspace is your personal workspace (called "My workspace" as indicated in the screenshot below).

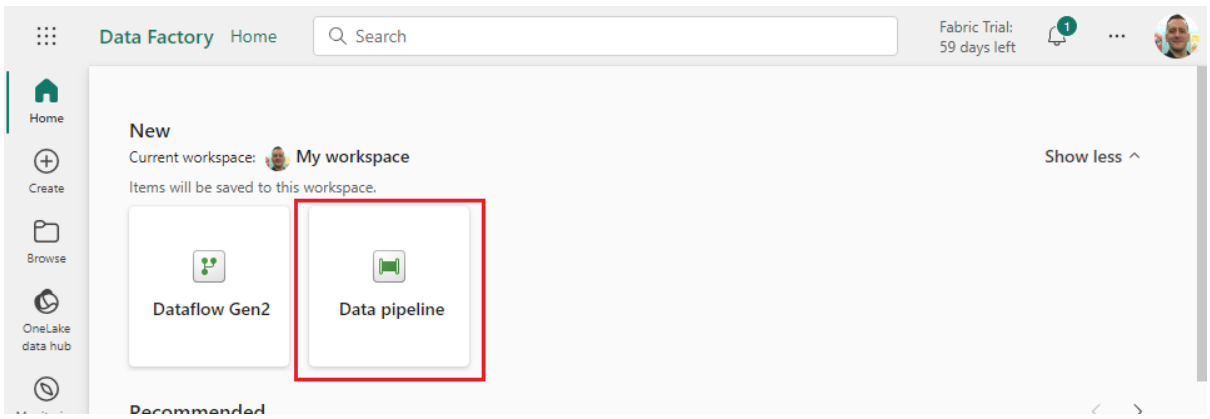


4. If your current workspace is something different, select your personal workspace from the list of workspaces accessed from the left-hand sidebar, then re-select the "Data Factory" experience using the button in the bottom left.

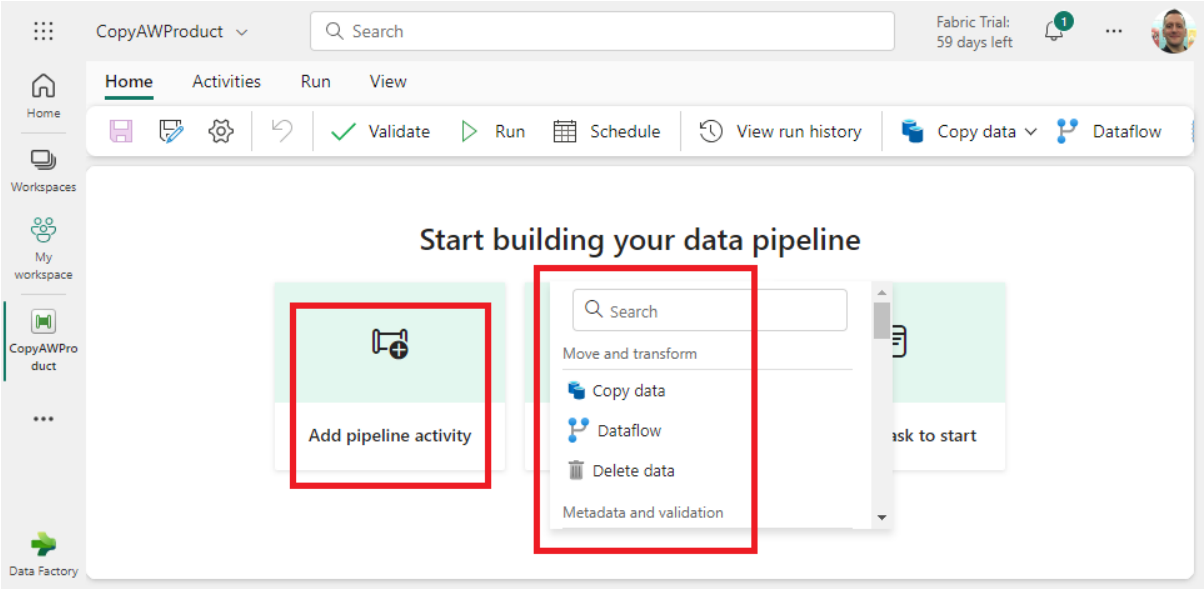
Lab 7.2 – Create a copy pipeline

In this lab you will replicate the ADF copy pipeline you created in Lab 2, this time as a Microsoft Fabric data pipeline.

- Click the "Data pipeline" button visible in the Data Factory experience home page. Name the new pipeline when prompted, then click "Create".



- Like ADF, Fabric provides several approaches to pipeline authoring – select "Add pipeline activity" to start a new pipeline with an activity of your choice. A dropdown menu offers a list of activities from which to choose. Many of these activities will be familiar from Azure Data Factory.



- Choose "Copy data" to create a Copy activity on the new pipeline's design surface. The pipeline design surface is similar to that provided by ADF Studio. Notice that:
  - Like ADF, activities are configured using the configuration pane in the lower part of the design surface. Like ADF, the Copy activity has General, Source, Sink, Mapping and Settings tabs, but Fabric refers to sinks as "destinations".
  - Other authoring tools are available in the tabbed ribbon above the design surface; for example, the "Activities" tab provides a toolbox from which new activities can be added to the design surface.

Lab 7.3 – Configure the Copy activity source

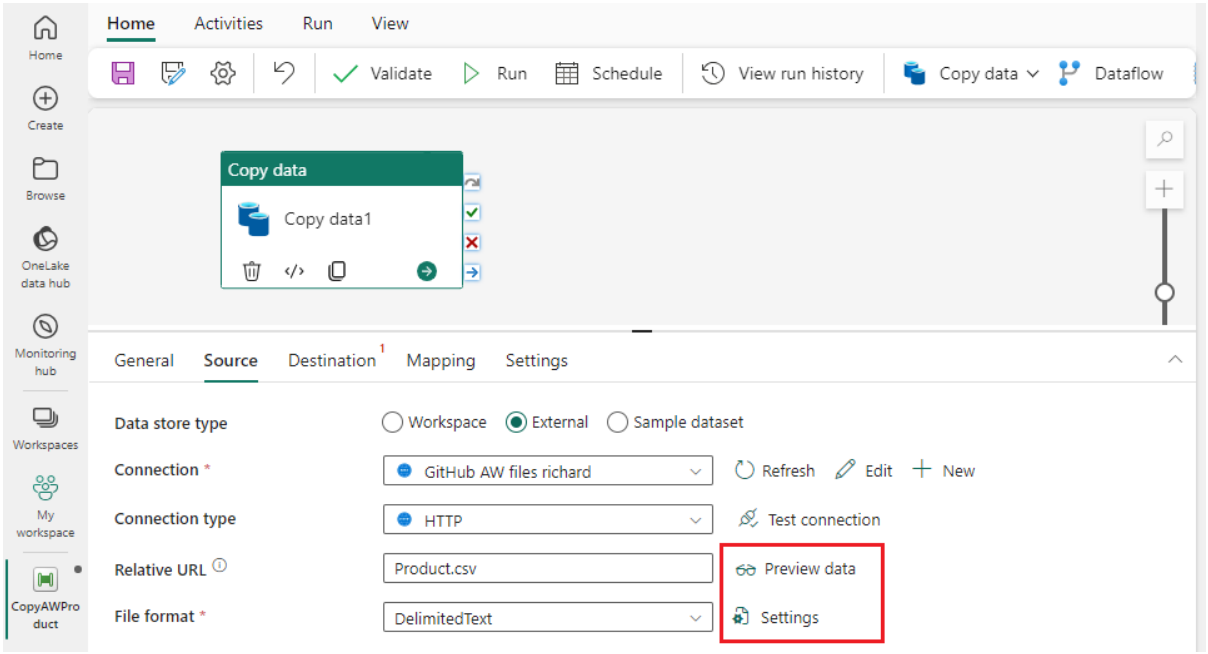
Configure the Copy activity's source using the "Source" tab on the activity's configuration pane.

1. Select a data store type of "External", then select "+ New" to create a new "Connection". Connections behave like linked services in ADF or Synapse.

2. Choose the "HTTP" connection type, then click Continue. Configure connection settings as follows:
   - **URL:** https://raw.githubusercontent.com/microsoft/sql-server-samples/master/samples/databases/adventure-works/oltp-install-script/ (including the final forward slash)
   - **Connection name:** (enter a connection name of your choice)
   - **Authentication kind:** Anonymous

   Click "Create" to create the new connection.

3. Back on the Source configuration tab, set the following options:
   - **Connection type:** HTTP
   - **Relative URL:** Product.csv
   - **File format:** DelimitedText



4. Use the "Preview data" button indicated in the screenshot to preview the data import. Check that the column header and delimiter settings are appropriate – if not, use the "Settings" button to the right of the "File format" dropdown to correct them.

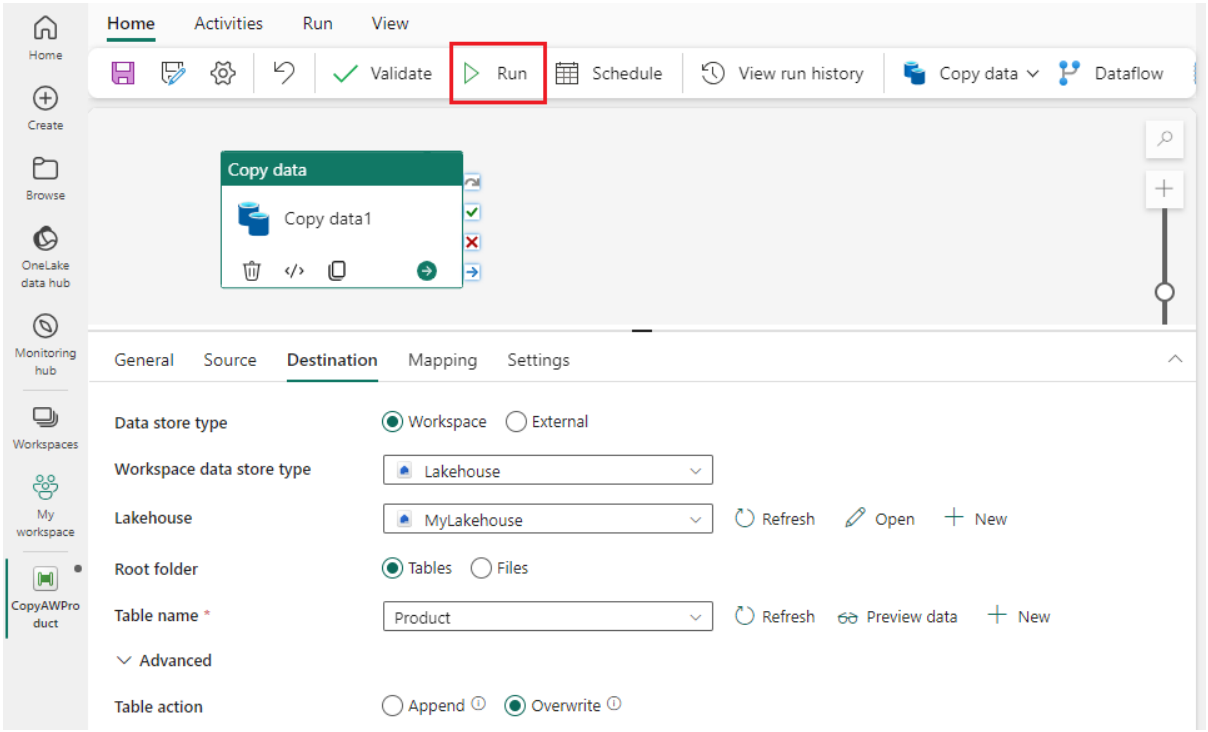
Lab 7.4 – Configure the Copy activity destination

In this lab you will configure a lakehouse as the Copy activity's destination. Fabric lakehouses are simply data lake storage managed for you in your Fabric tenant's OneLake and divided between "files" and "tables". All lakehouse storage is file-based – table storage refers to data stored in the Delta format you encountered in Lab 4.3.

Configure the Copy activity's destination as follows.

1. Using the "Destination" tab on the activity's configuration pane, set Data store type to "Workspace" then choose "Lakehouse" from the Workspace data store type dropdown. Create a new Lakehouse with a name of your choice. Notice that explicit connections or linked services are not required for workspace data stores.

2. Choose the "Tables" root folder and name the table "Product". Notice that no file format information is required – table data is stored in Delta format. Expand the "Advanced" section and set "Table action" to "Overwrite".

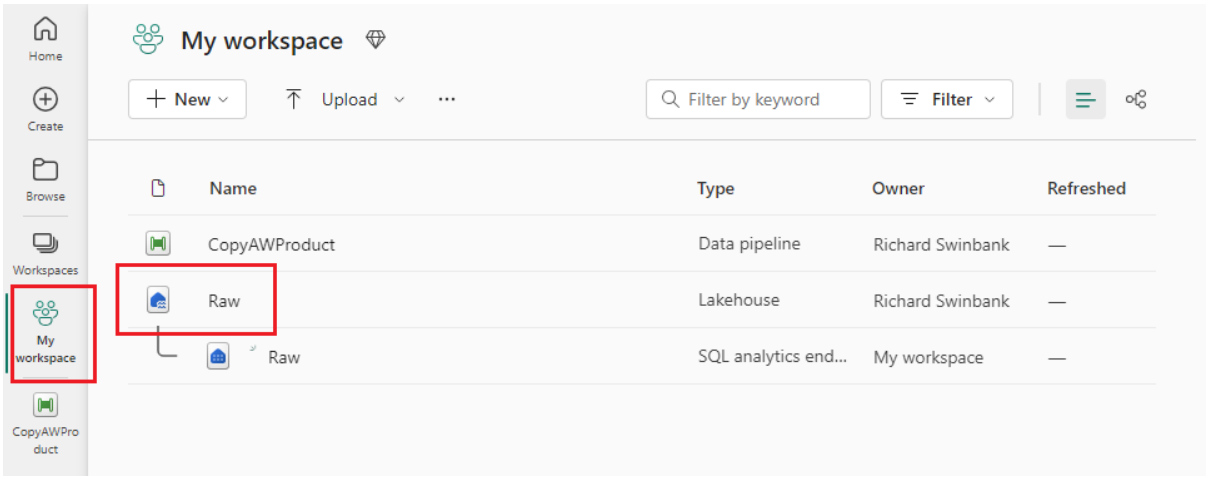


3. Run the pipeline using the "Run" button found on the "Home" or "Run" ribbons.

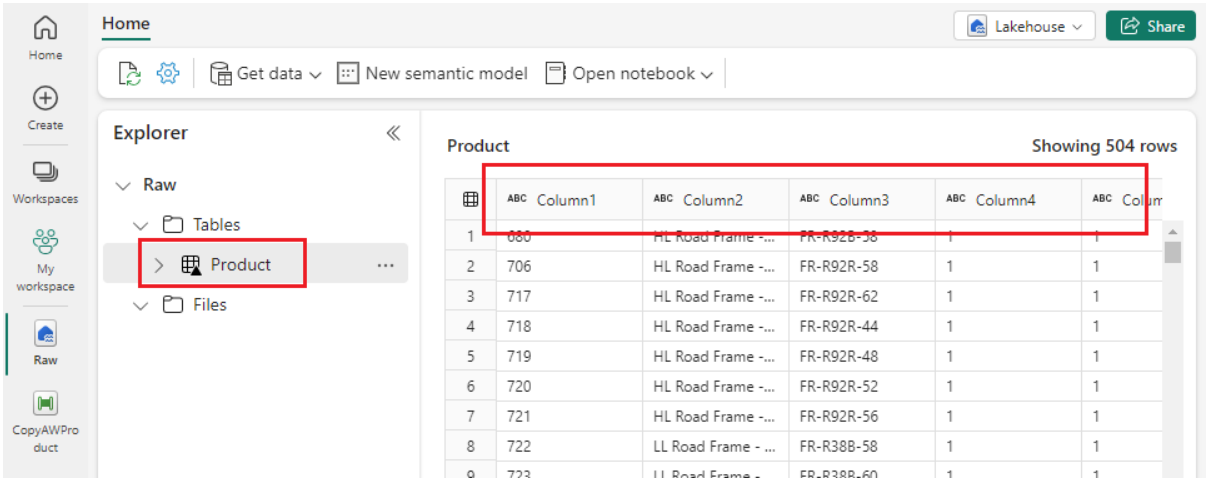
Lab 7.5 – Inspect the results

In this lab you will inspect the results of the Copy operation.

1. Browse to your personal workspace, then select the lakehouse you created in Lab 7.4.



2. Expand the "Tables" root in the lakehouse explorer and select the "Product" table. Notice that all columns have names like "Column1", "Column2", etc and have string data types (indicated by the "ABC" symbol preceding column names).



Recap

In Lab 7 you:

- Connected to a Fabric tenant
- Created and configured a data pipeline
- Used the pipeline to copy data into a lakehouse.

The purpose of this lab was to introduce data pipelines in Microsoft Fabric, demonstrating some similarities and differences. Features we have not covered include:

- Git configuration. While it is possible to connect a Fabric workspace to a Git repository, saving data pipelines to version control is not yet supported.
- Copy activity column mapping. The Delta table format (column selection, names and types) can be controlled via the Mapping configuration tab on the Copy activity. This is also true in ADF for column-oriented formats like delimited text or Parquet; Delta output is not supported by the ADF Copy activity.
- Gen2 Dataflows. ADF-style data flows are not supported in Microsoft Fabric; Gen2 dataflows provide a Power Query-like data transformation experience which you may wish to explore further.