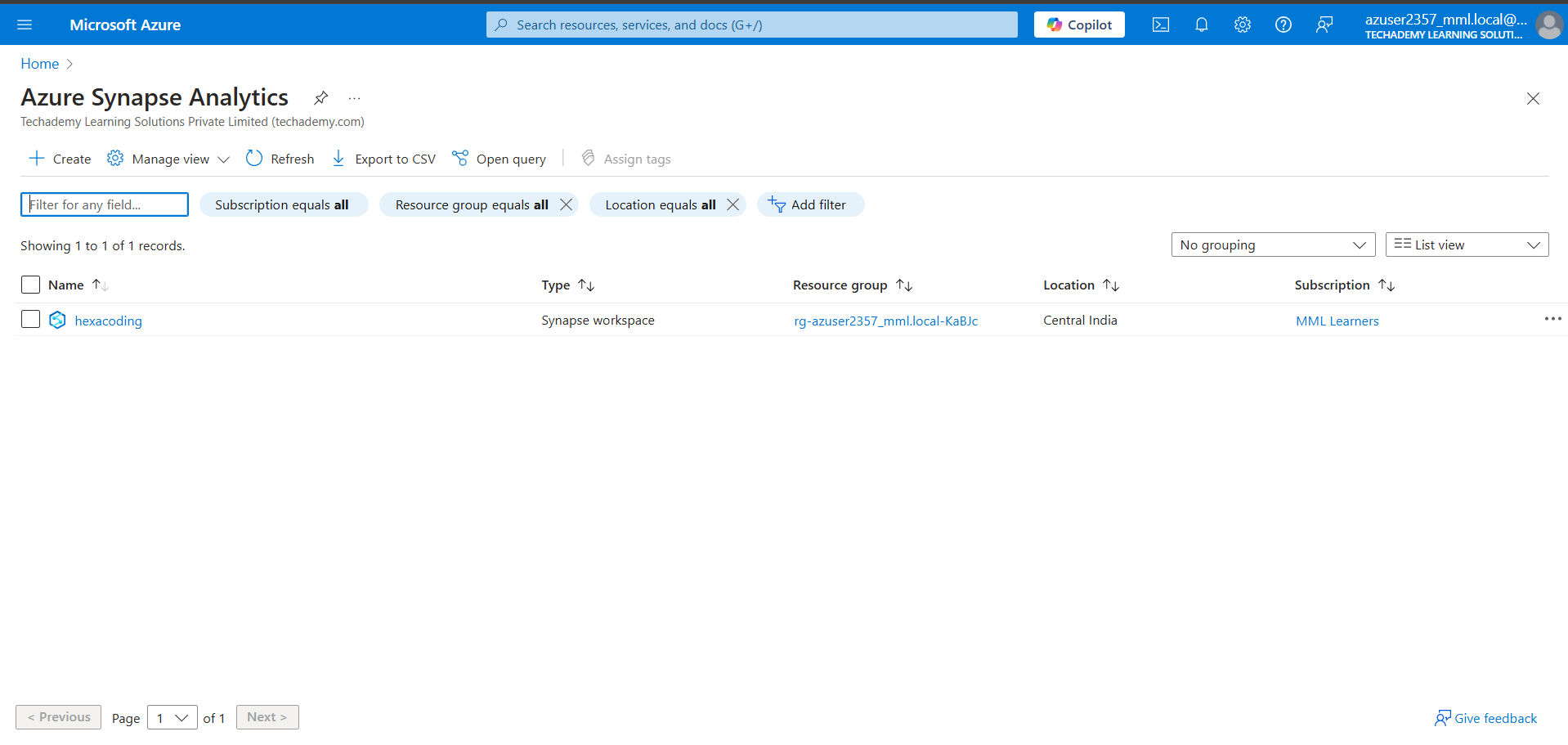
**Hexaware Technologies**

Azure Synapse – Coding Challenge

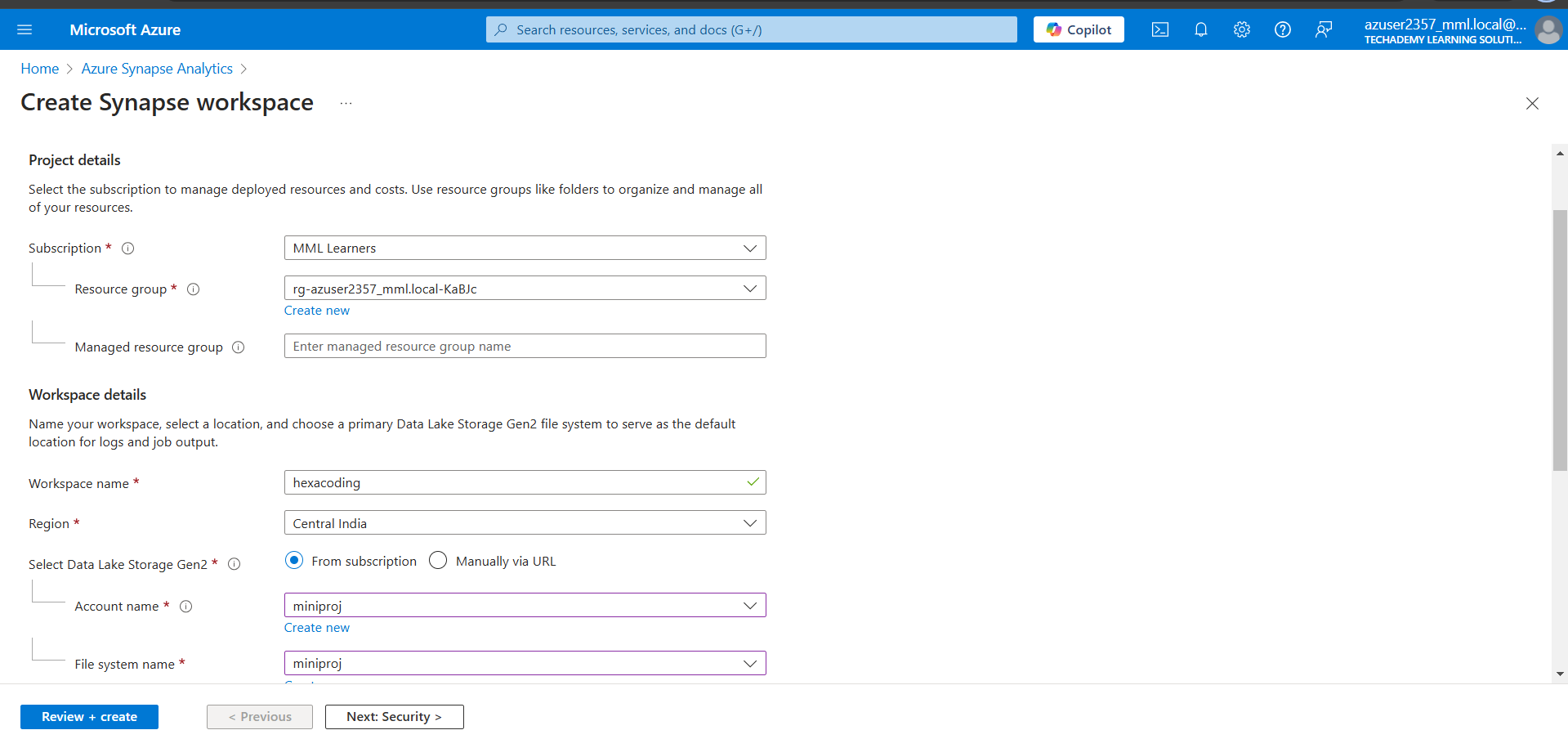
**Step 1: Set Up the Environment**

**1.Create an Azure Synapse Analytics Workspace**:

* Go to the Azure portal.
* Click **Create a resource** > Search for "Azure Synapse Analytics" > Select and create a workspace.



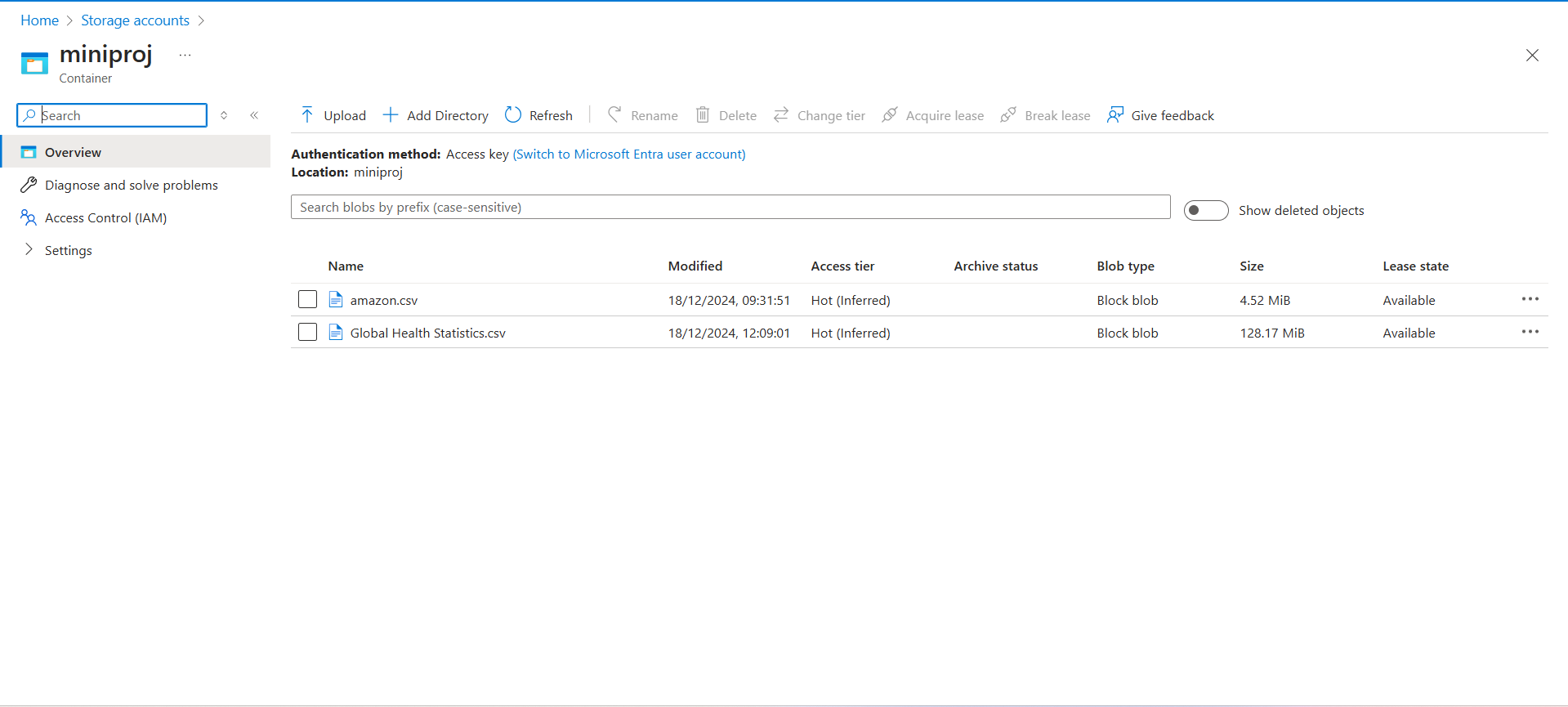
**2.Link the Data Lake Storage Gen2 Account**:



**Step 2: Ingest Data**:

**1.Prepare Your Data Source**:

* Identify the data source you want to use (e.g., a local CSV file, an Azure SQL Database, or an API).
* If using a CSV file, upload it to your linked Data Lake Storage Gen2 account.

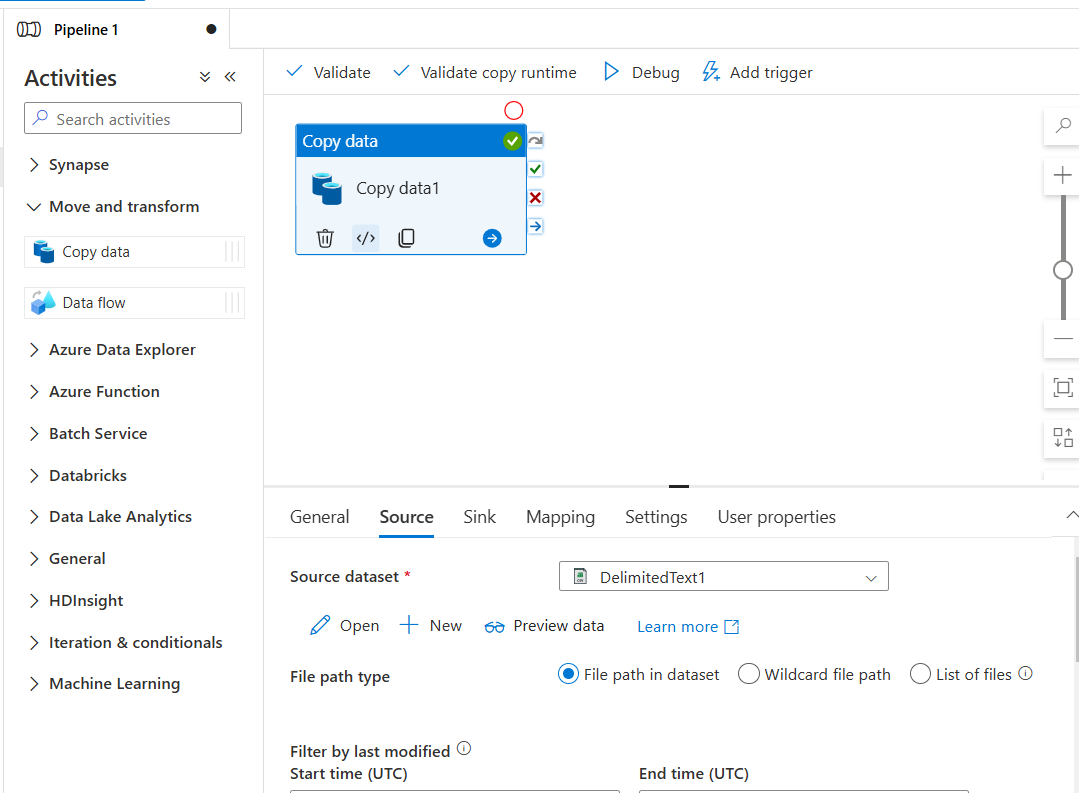


**2.Create a Pipeline in Synapse**:

* In Synapse Studio, go to the **Integrate** tab.
* Click on **+ New** and select **Pipeline**.

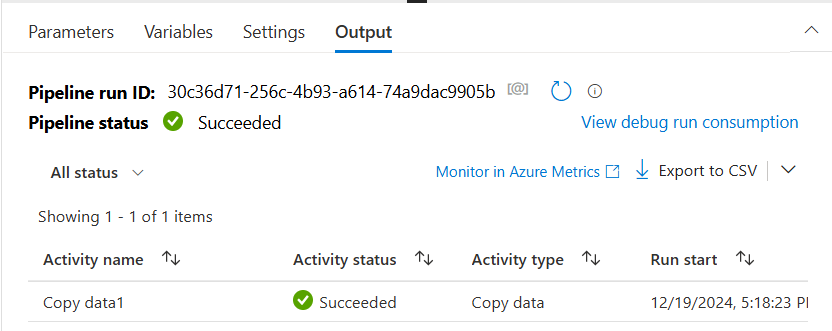
**3.Add a Copy Data Activity**:

* Drag and drop the **Copy Data** activity into the pipeline canvas.



**4.Test the Pipeline**:

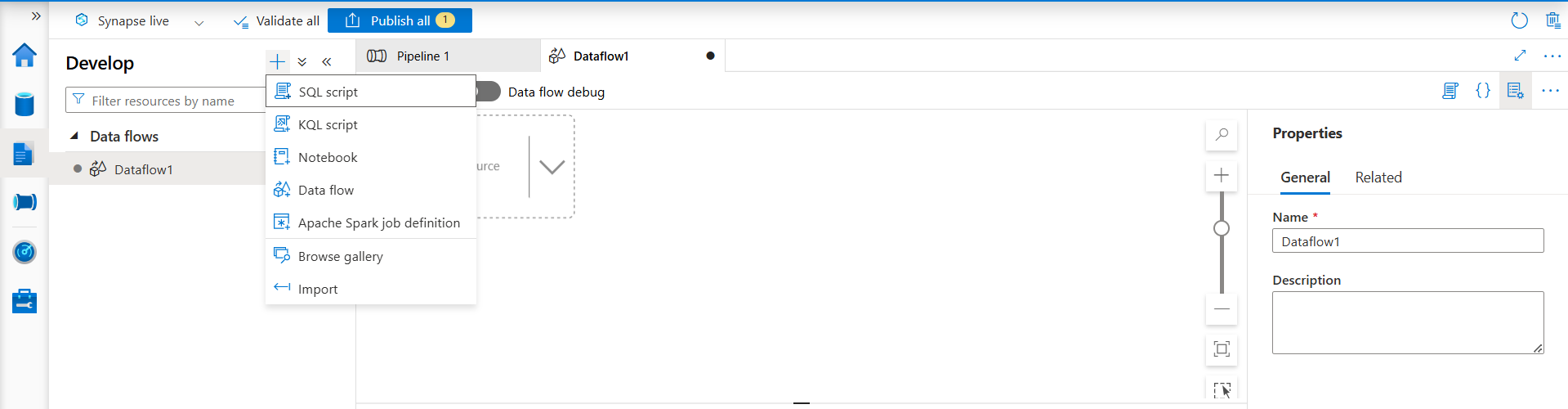
* Click **Debug** to test the pipeline and verify that the data is ingested into your Data Lake.



**Step 3: Create the Dataflow**.

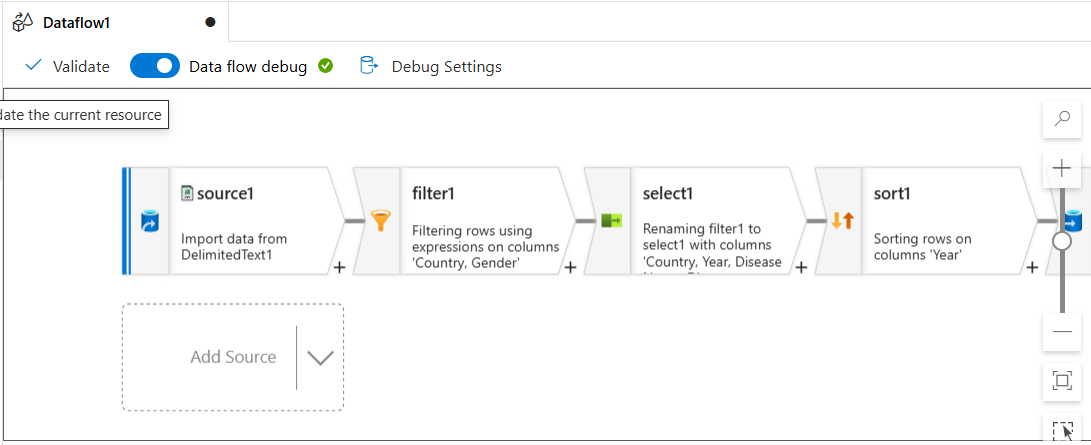
**1. Navigate to the Dataflow Section:**

* In Synapse Studio, go to the **Develop** tab.
* Click **+ New** and select **Dataflow**.



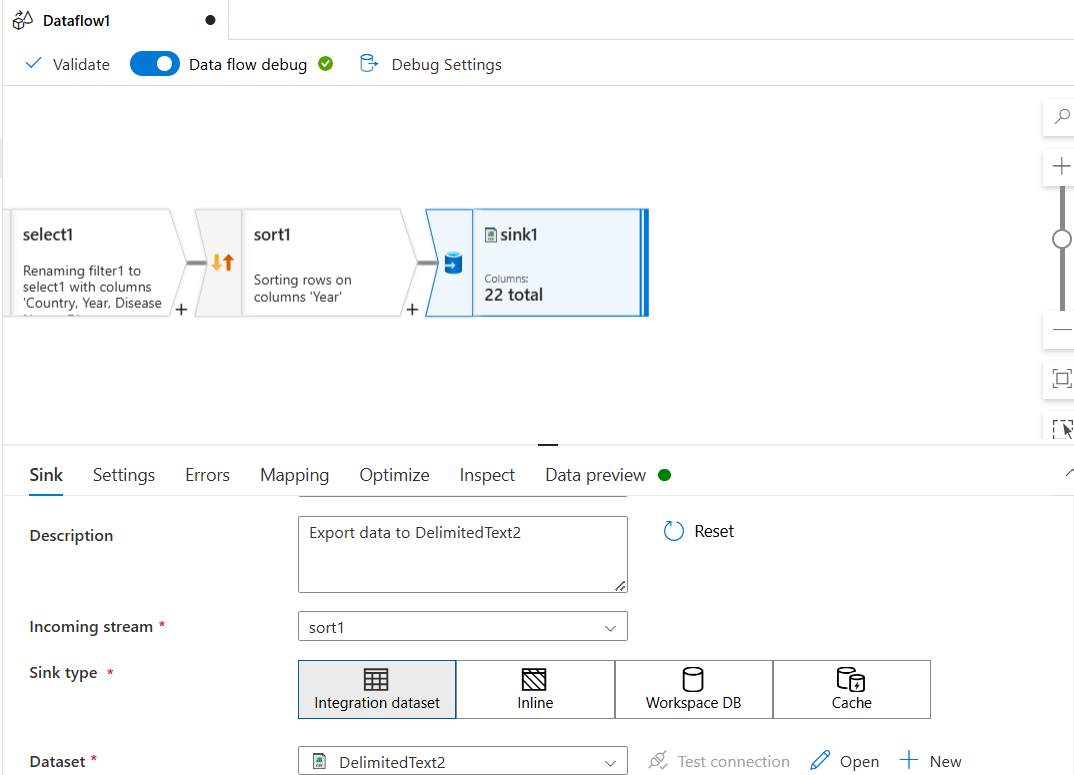
**3. Add Transformations:**

* Perform the required transformations on the data:
  + Use operations like **Filter**, **Aggregate**, **Join**, **Derived Column**, or **Pivot/Unpivot** as per your requirements.
  + For each transformation, configure the logic in the settings pane.



**4. Configure the Sink in the Dataflow:**

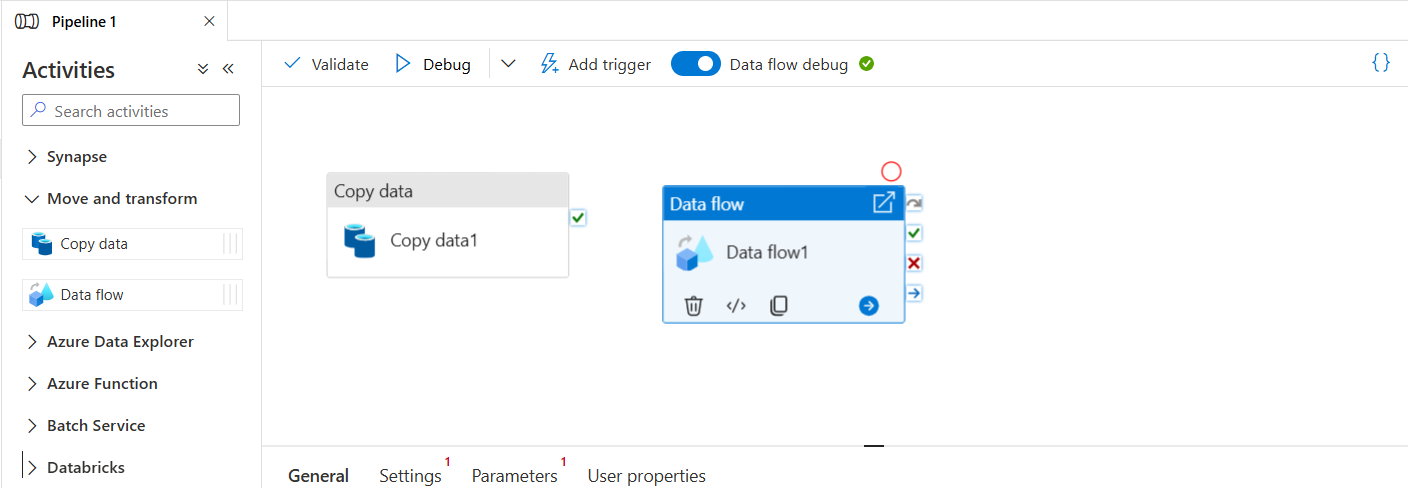
* Add a sink transformation to write the transformed data back to a destination.
* Configure the sink:
  + Choose a dataset for the output destination (e.g., another folder in your Data Lake or a Synapse Analytics table).
  + Define the file format (e.g., CSV, Parquet) or table schema.



**Step 4: Integrate the Dataflow with the Pipeline**

**1. Add a Dataflow Activity in the Pipeline:**

* Go back to the **Integrate** tab in Synapse Studio.
* Open the pipeline you created earlier for data ingestion.
* Drag and drop the **Dataflow** activity from the activities pane to the pipeline canvas.

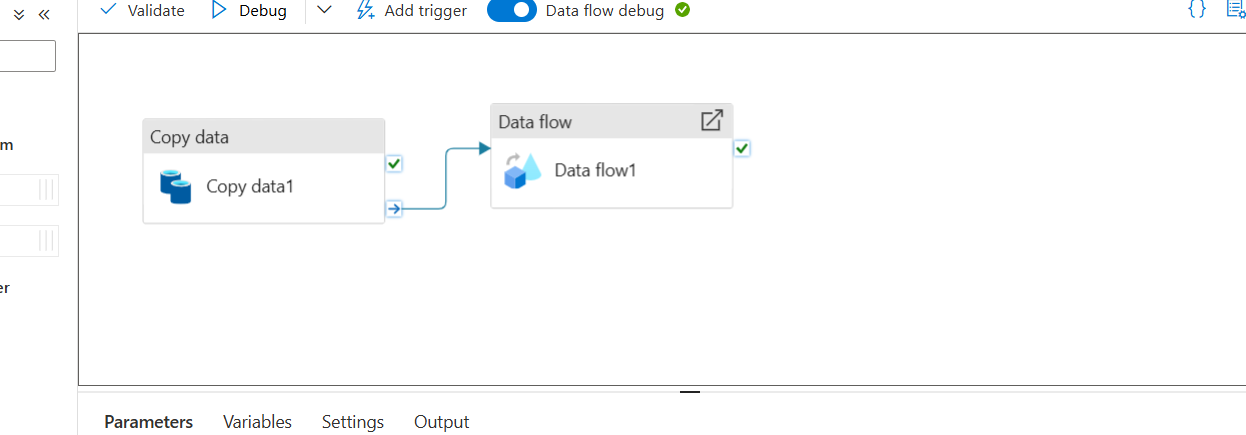


**2. Link the Dataflow to the Dataflow Activity:**

* In the **Dataflow** activity settings, click on **Source** and select the dataflow you created in Step 3.
* Make sure the dataflow is properly linked to the activity.

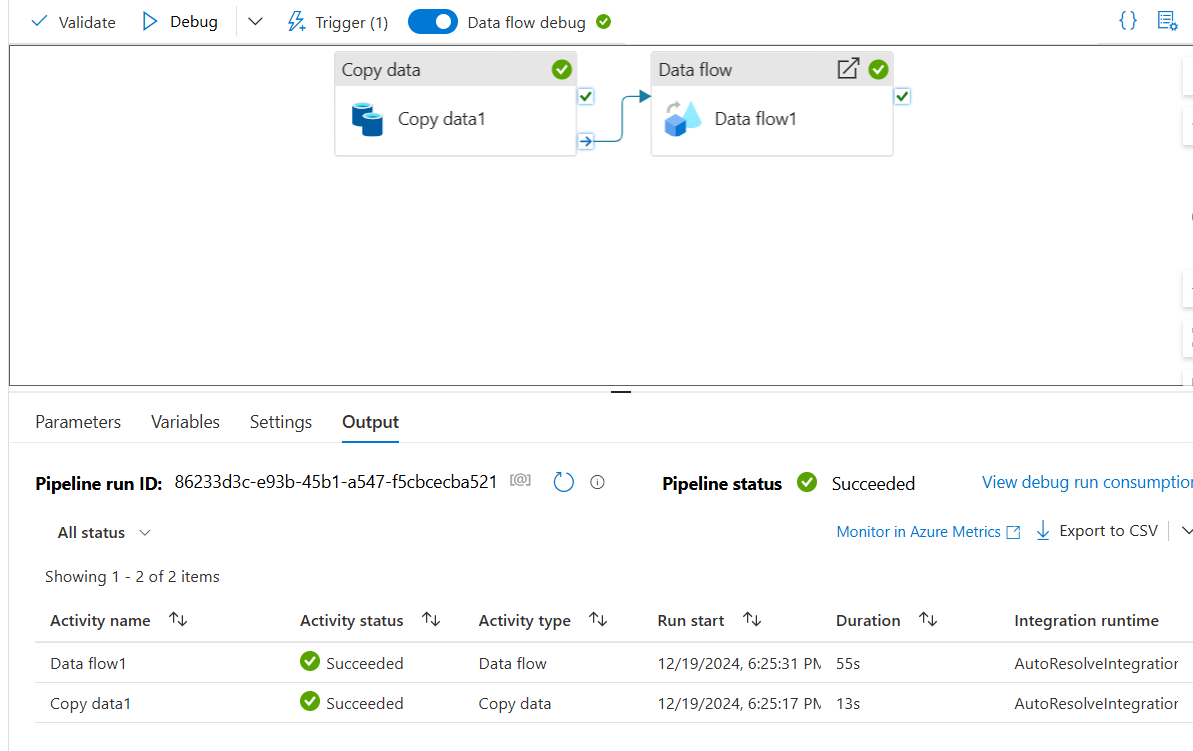
**4. Connect the Dataflow Activity to the Pipeline:**

* Ensure the dataflow activity is connected to the rest of your pipeline



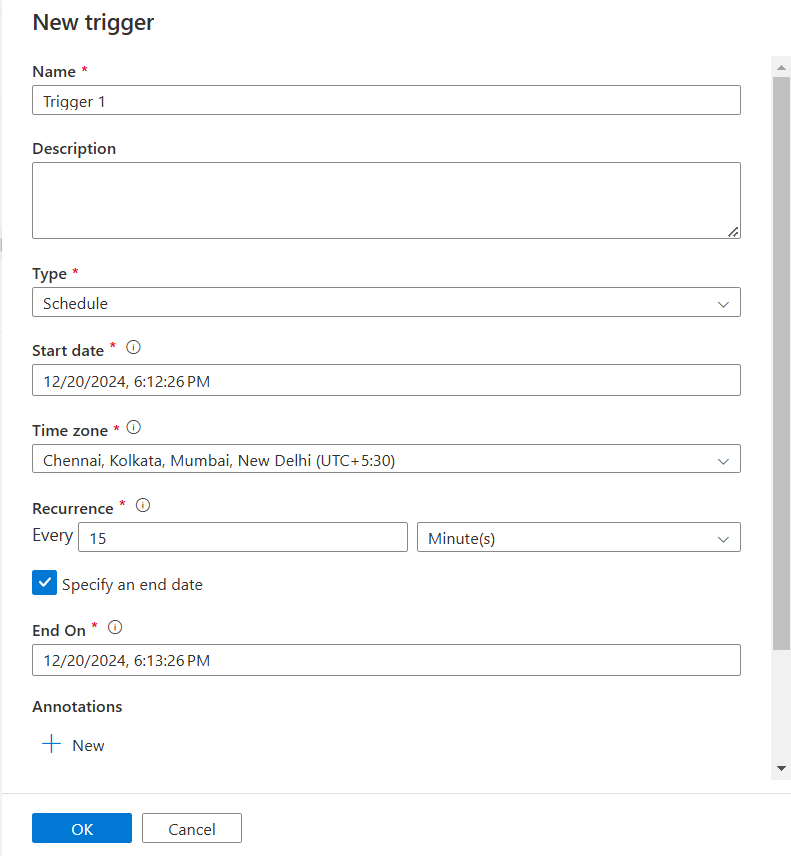
**5. Debug the Pipeline:**

* Click **Debug** to test the pipeline and make sure the dataflow runs smoothly within it.
* Check the output to ensure that the data is correctly transformed and written to the destination.



**Step 5: Schedule and Monitor the Pipeline**

* 1. **Schedule the Pipeline**



**2. Monitor Pipeline Runs:**

* Go to the **Monitor** tab in Synapse Studio.
* Select **Pipelines** from the left panel.

