# **Execution Proof Document**

## **Data Lake Analytics**

## 1. Introduction

This document provides execution evidence and project results for the **Data Lake Analytics Solution** built using **Azure Data Factory (ADF)** and **Azure Data Lake Storage (ADLS)**.

The objective is to demonstrate:

The successful end-to-end execution of ETL pipelines in ADF.

The organization of data across **Bronze, Silver, and Gold layers** in ADLS.

The validation of transformations and schema consistency.

The readiness of curated data for downstream analytics and reporting.

## 2. Project Architecture

The architecture follows a **multi-layered data lake design**:

**Bronze Layer:** Stores raw ingested data.

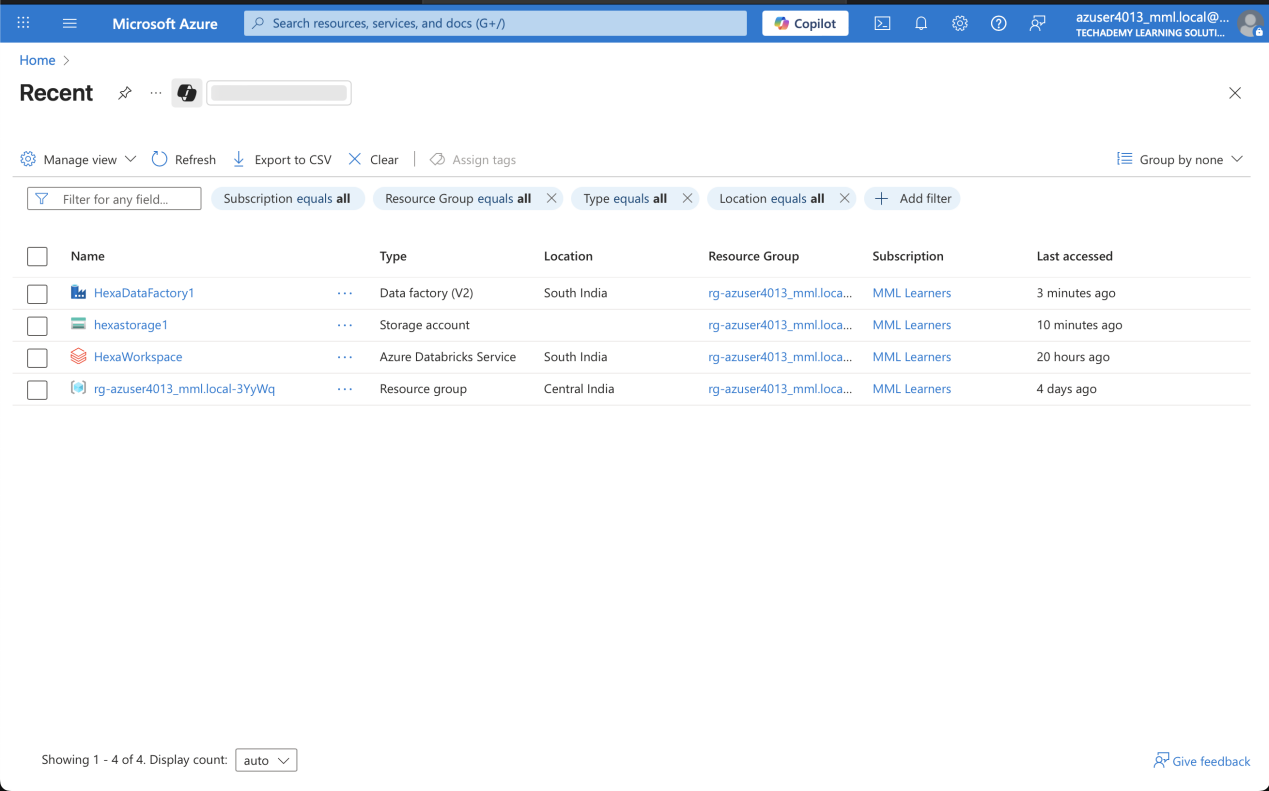
**Silver Layer:** Contains cleaned and transformed data.

**Gold Layer:** Holds curated, analytics-ready tables (fact and dimension data).

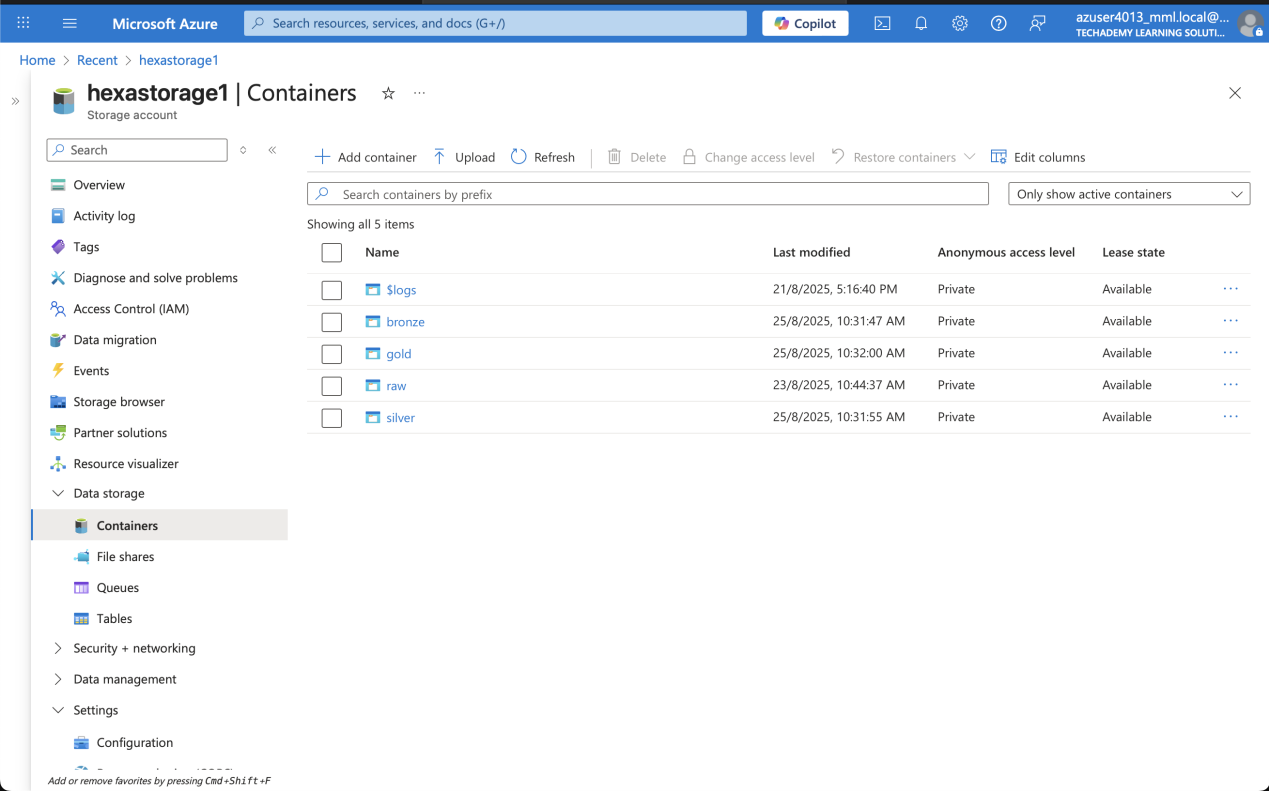
## 3. Environment Setup Evidence

The environment was successfully set up in Azure, including:

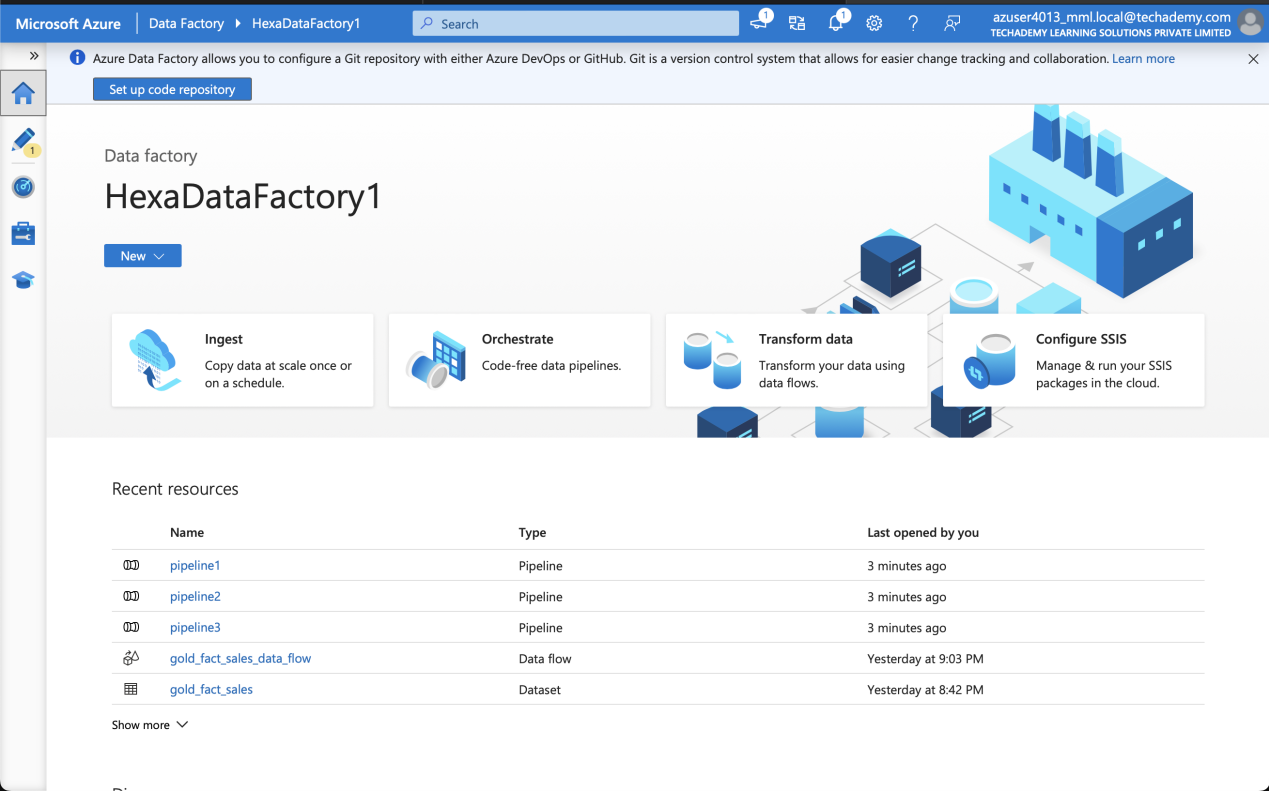
* A Resource Group for project resources.



* An ADLS Gen2 account with **bronze, silver, and gold** containers.



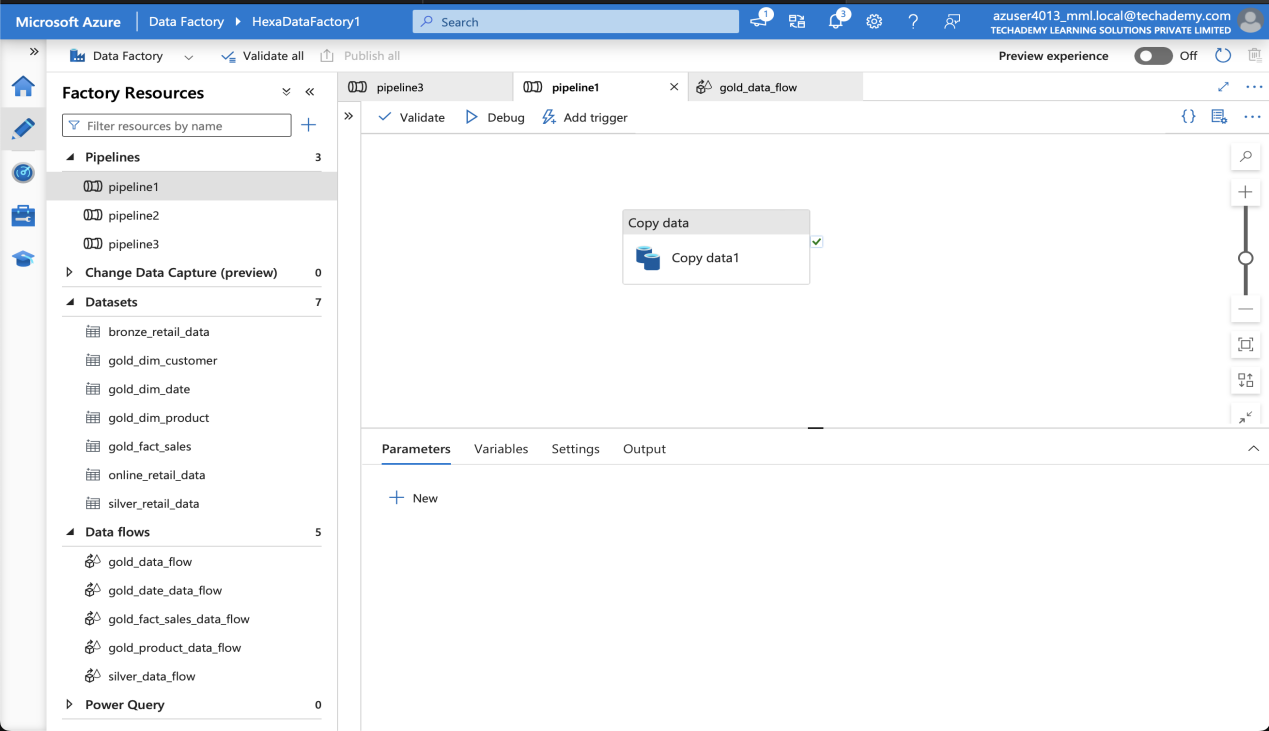
* An ADF workspace configured with Managed Identity for secure access.



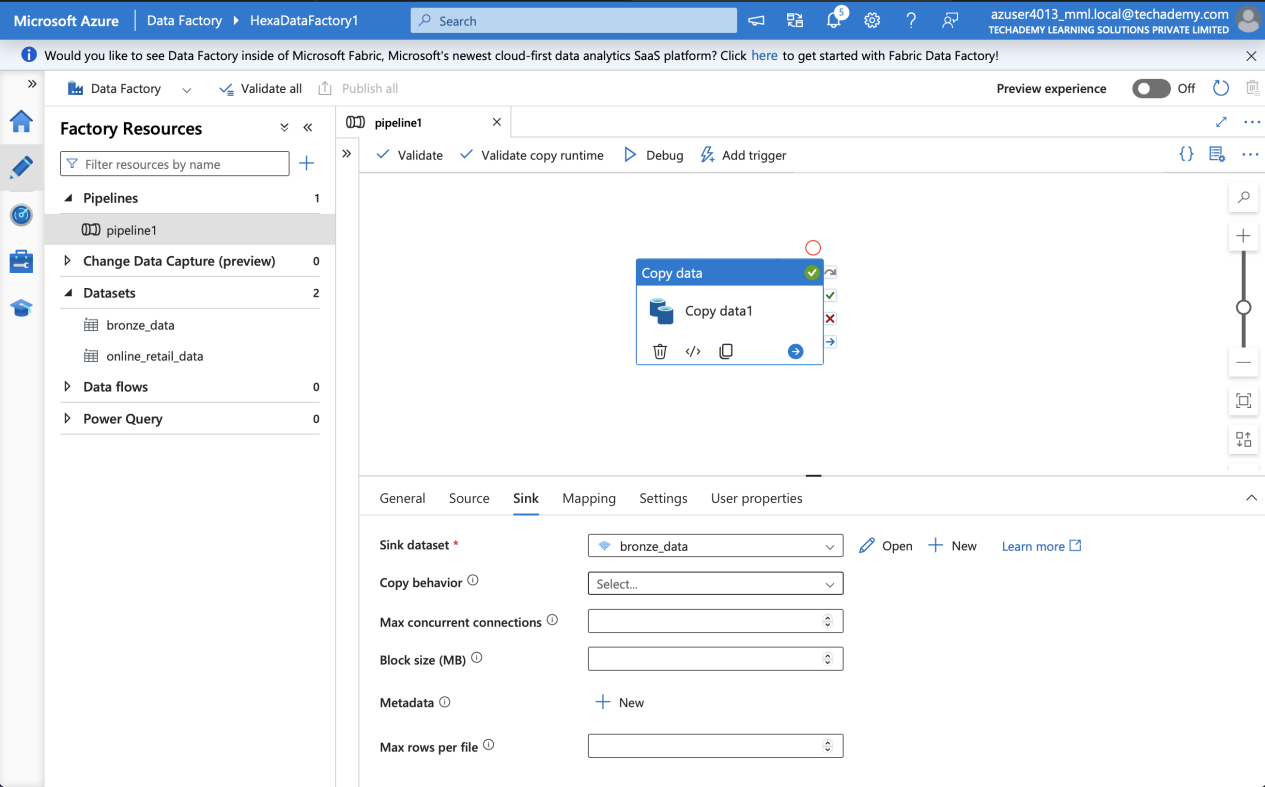
## 4. Data Ingestion (Bronze Layer)

Raw data was ingested from source files into the **Bronze container** using ADF Copy Activity pipelines. Monitoring results confirm successful execution.

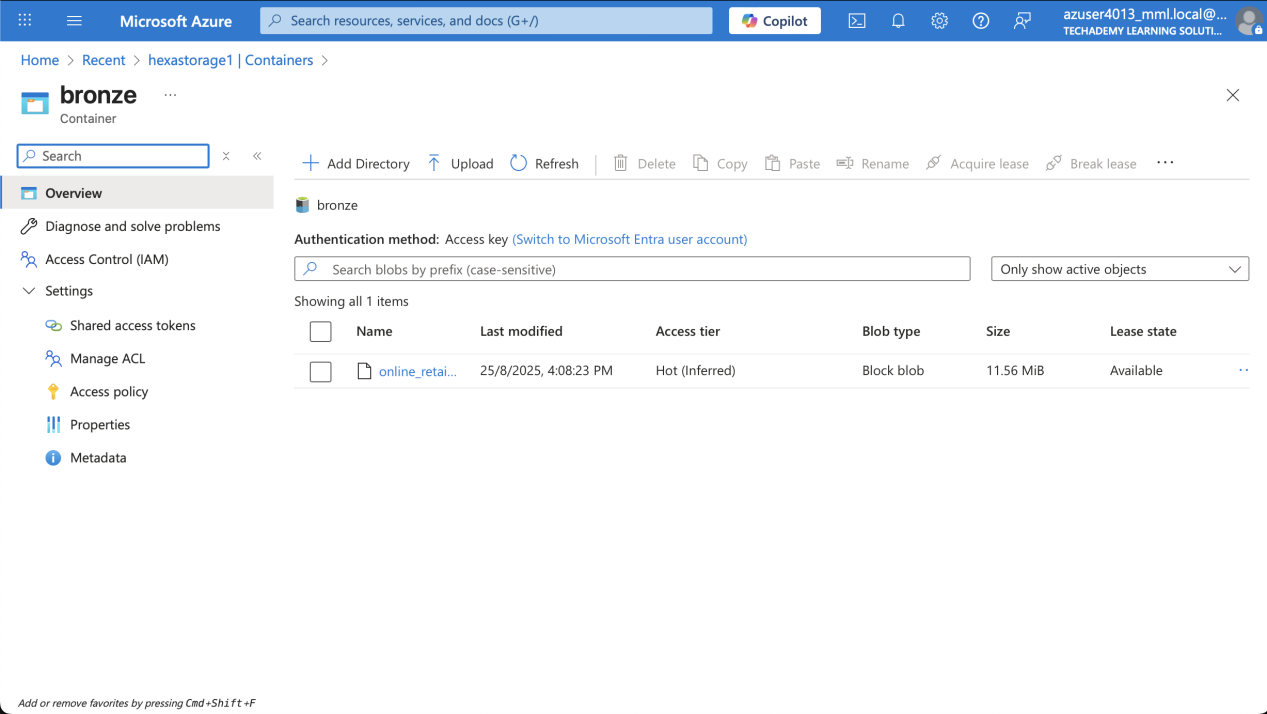
* ADF Copy Activity pipeline (Designer view)



* Monitoring tab showing successful pipeline run



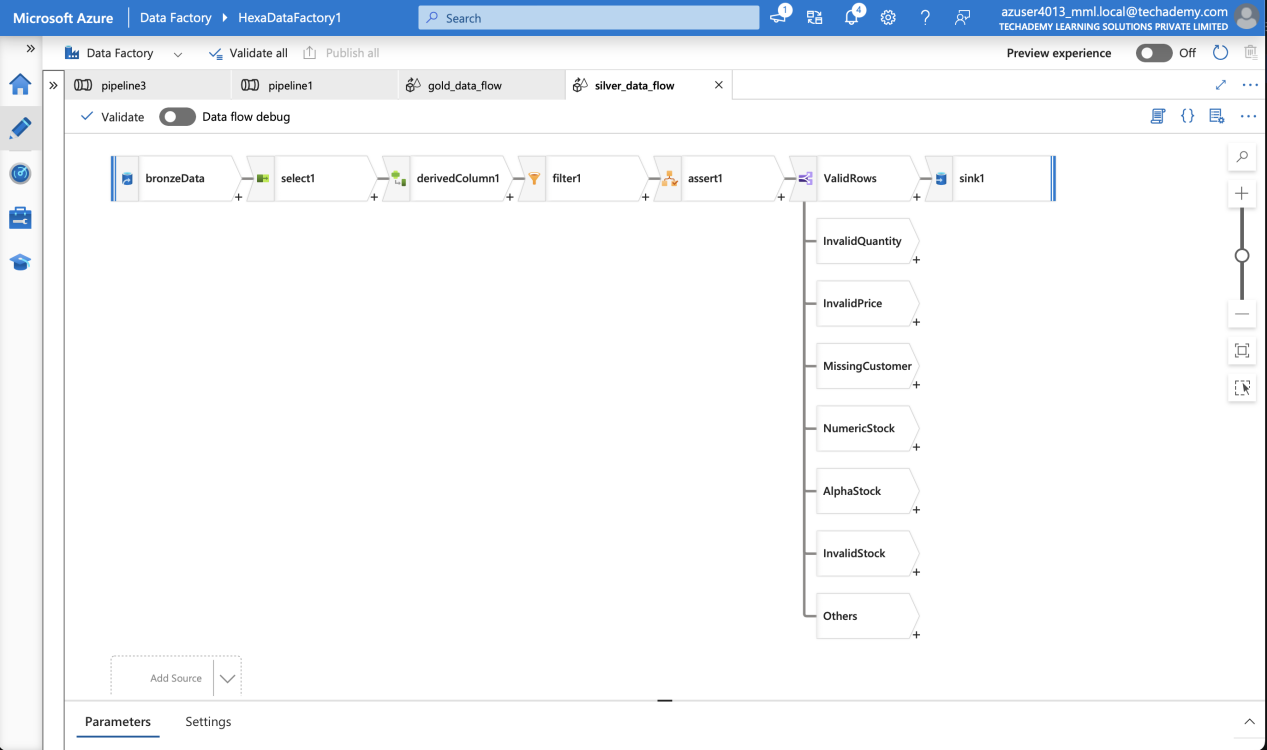
* ADLS Bronze container showing ingested files



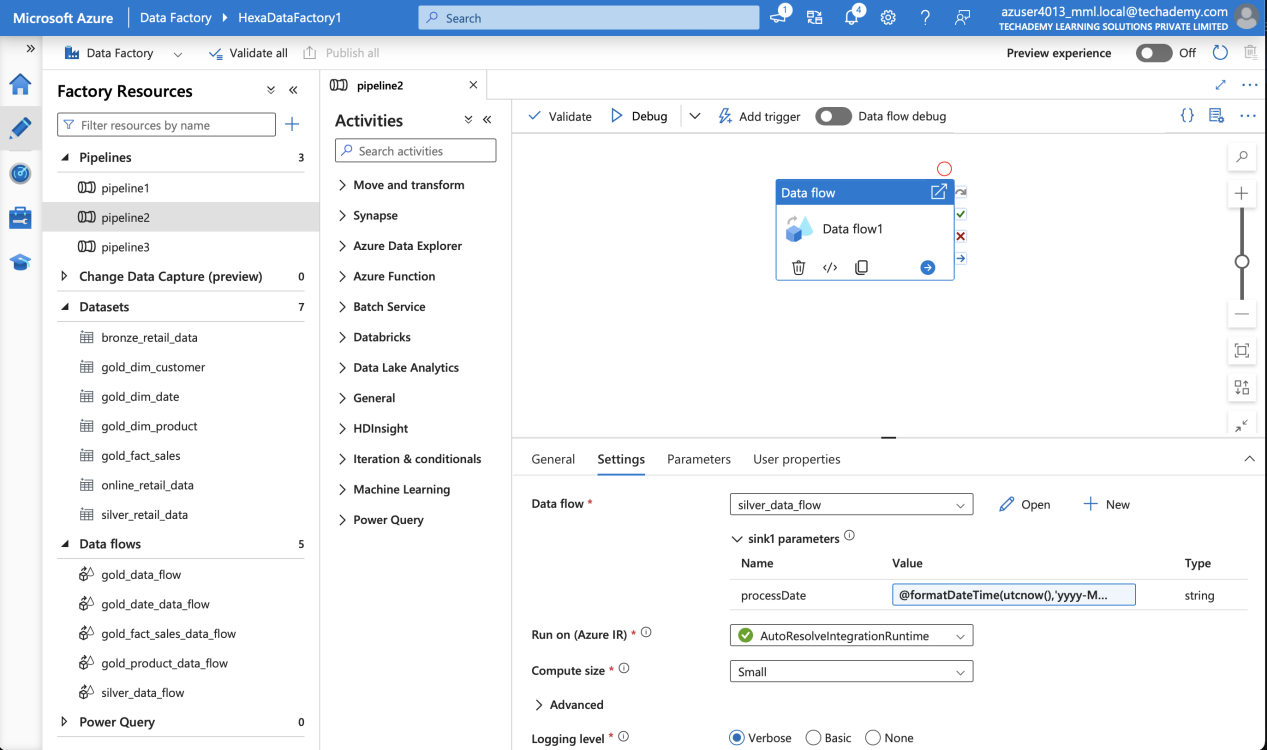
## 5. Data Transformation (Silver Layer)

Data was transformed in ADF Mapping Data Flows, including:

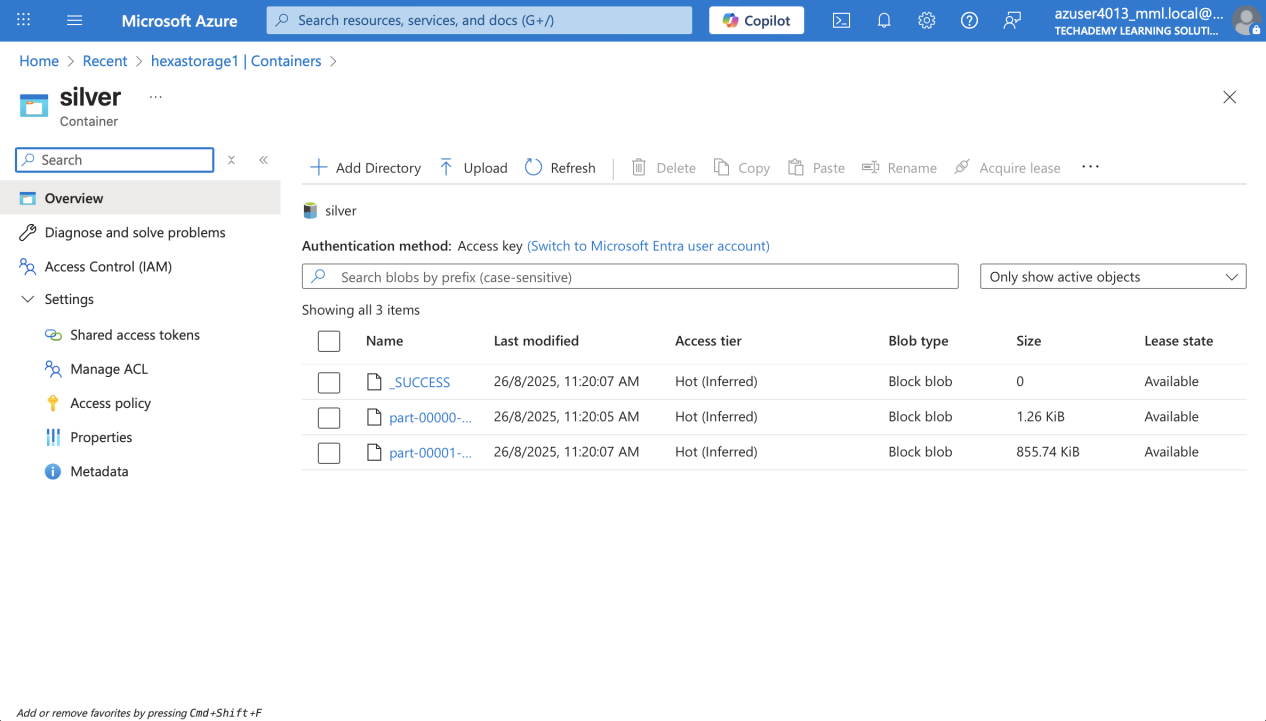
* Removing null/duplicate records.
* Standardizing date and numeric formats.
* Creating derived attributes.
* The transformed data was stored in the **Silver container**.
* ADF Mapping Data Flow (showing transformations: removing nulls/duplicates, derived columns)



* Pipeline execution success in Monitoring tab



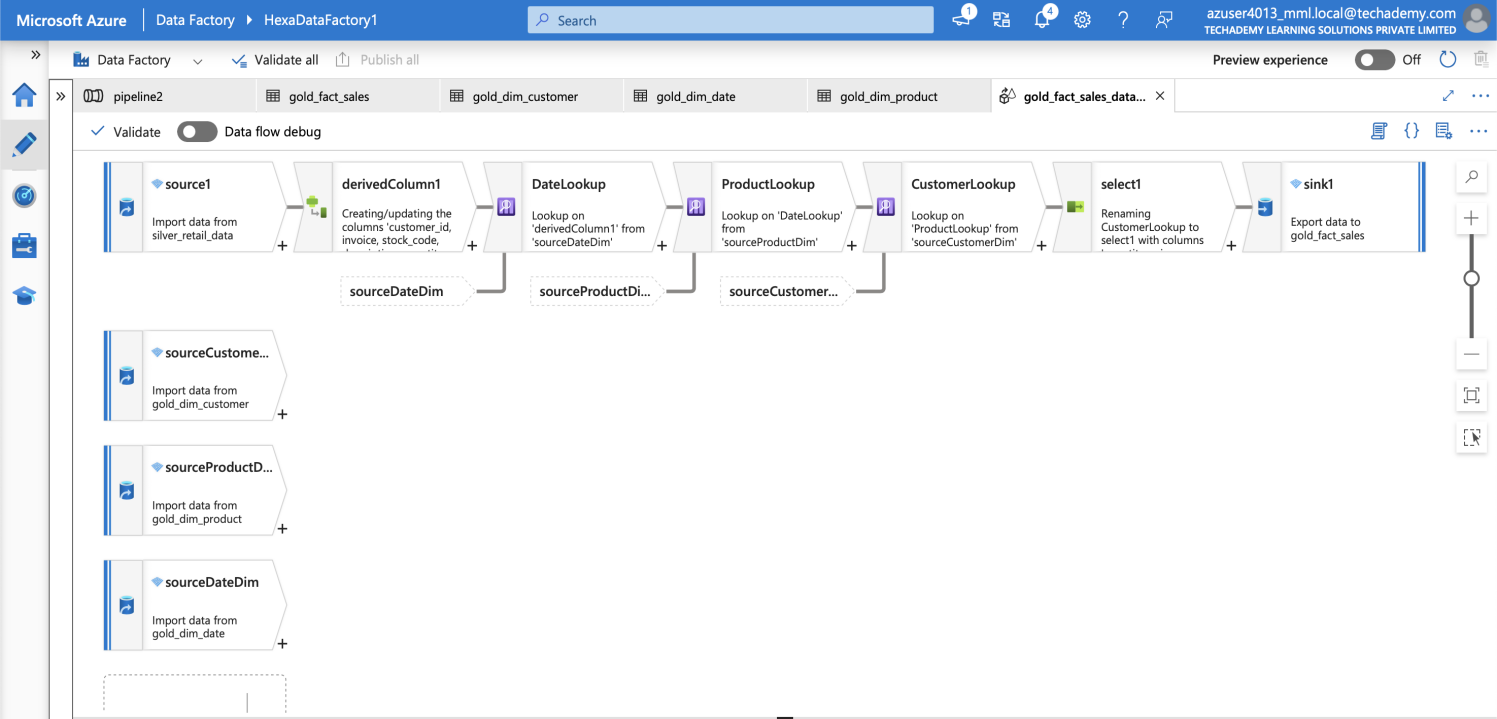
* ADLS Silver container with transformed files



## 6. Curated Data (Gold Layer)

Curated tables were created and stored in the **Gold layer**. These include:

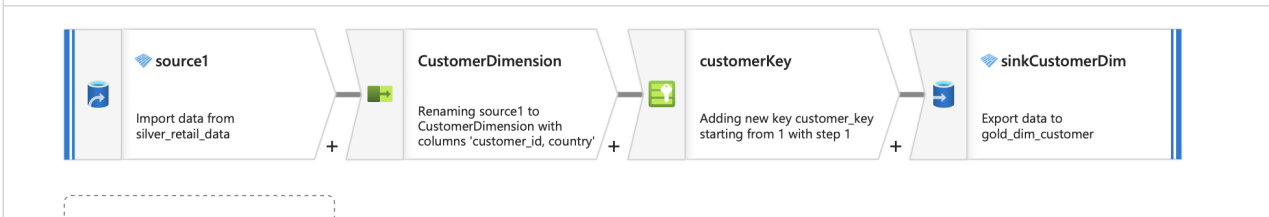
* **Dimension tables:** Customer, Product, Date.
* **Fact table:** Sales (linked with surrogate keys).  
  This structured data is now **analytics-ready**.
* ADF pipeline creating dimension and fact tables



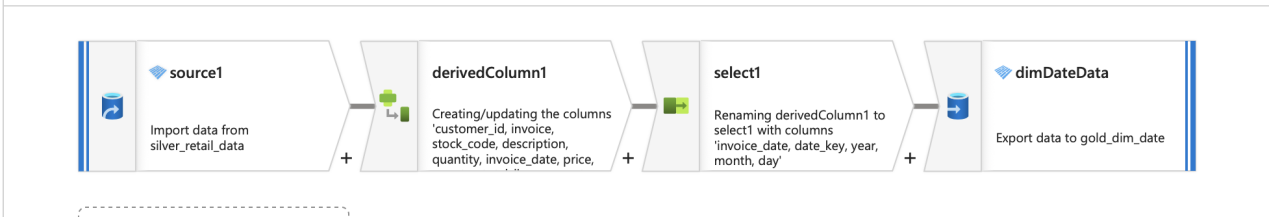
* ADF Gold product Data flow:

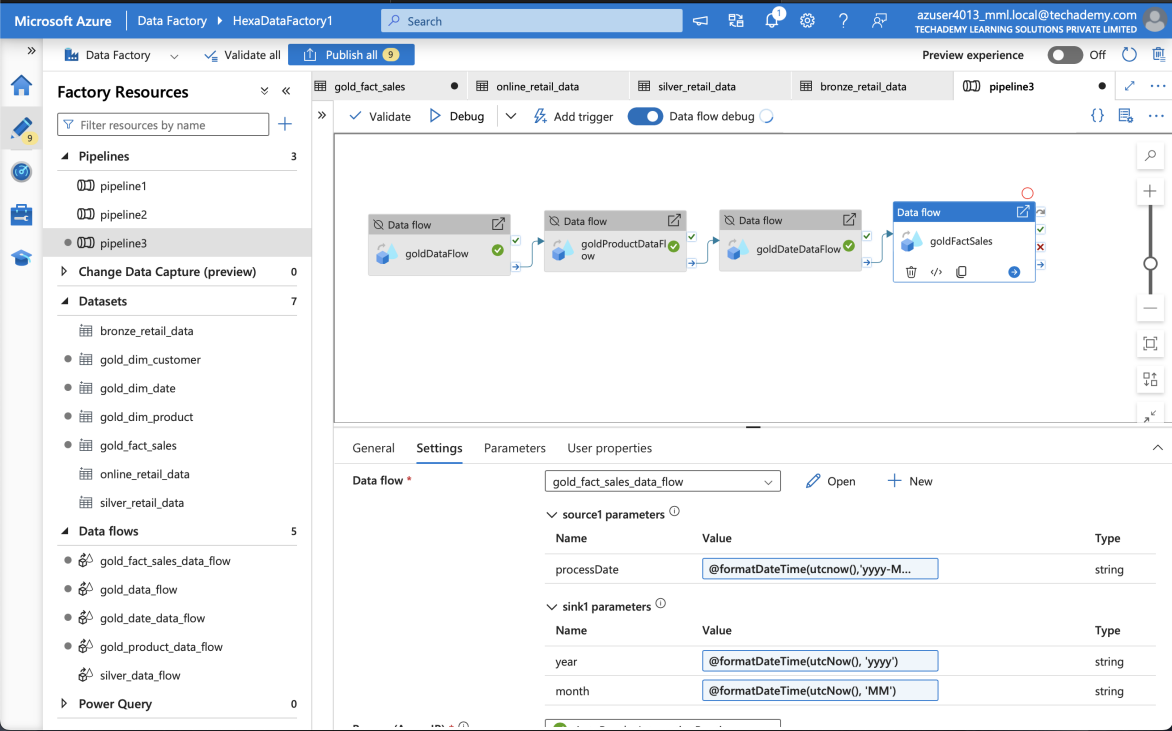


* ADF Gold Customer Data flow:

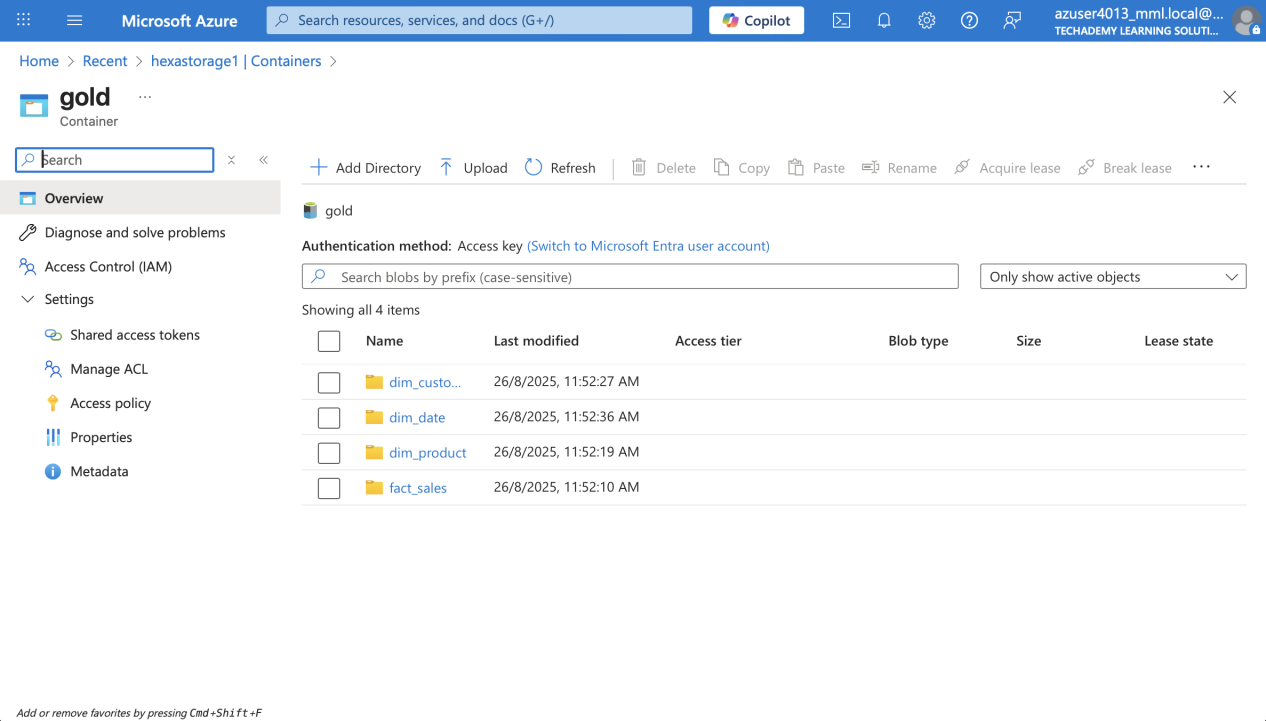


* ADF Gold Date data flow:





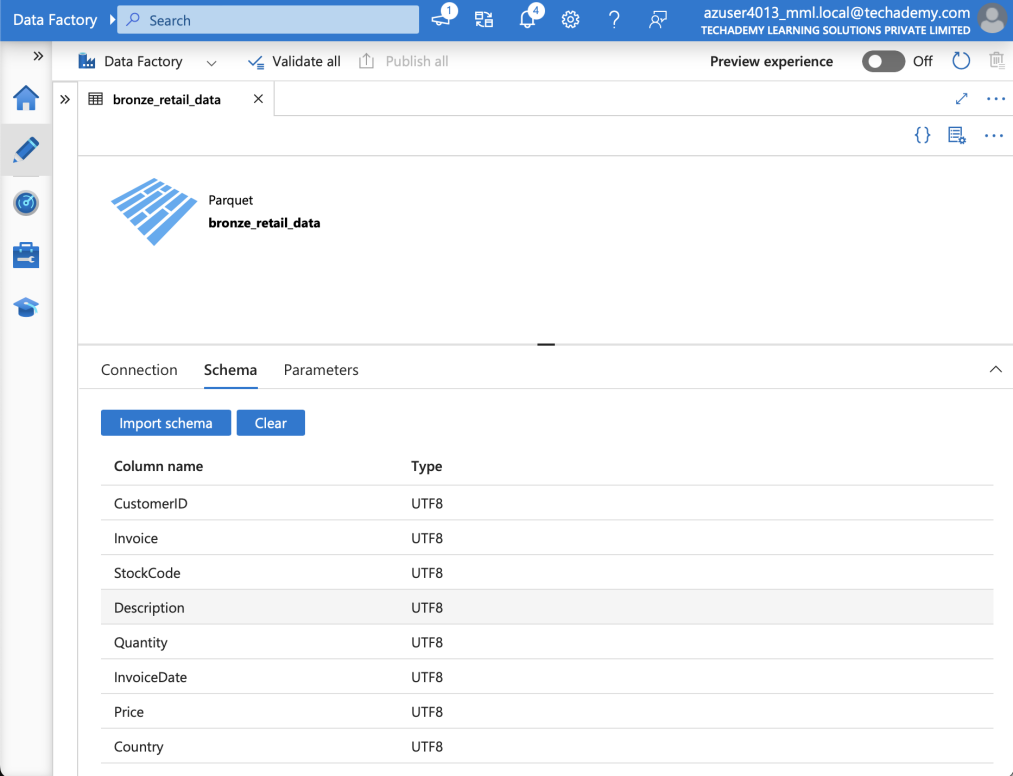
* ADLS Gold container with structured files



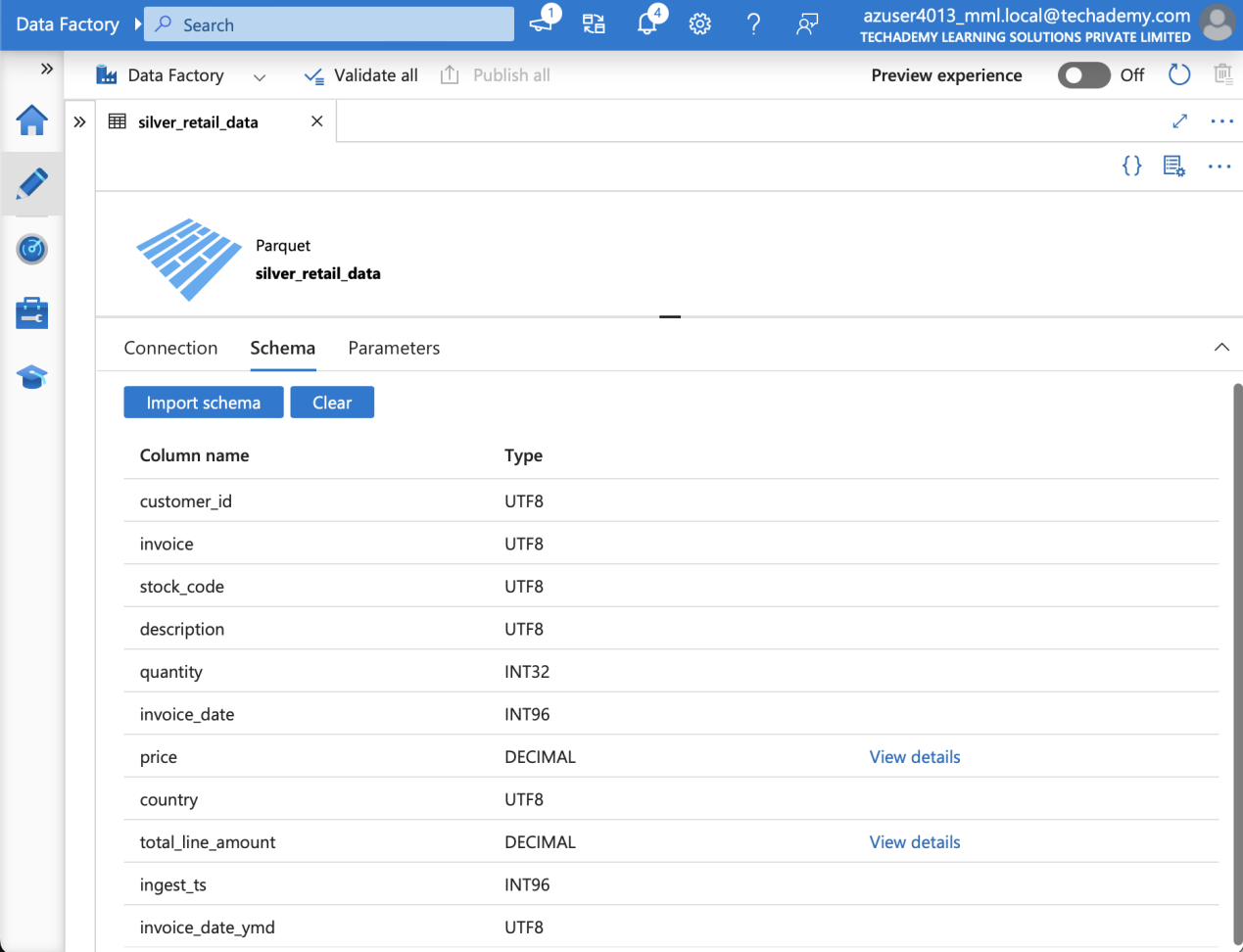
## 7. Validation & Outputs

Validation checks confirmed:

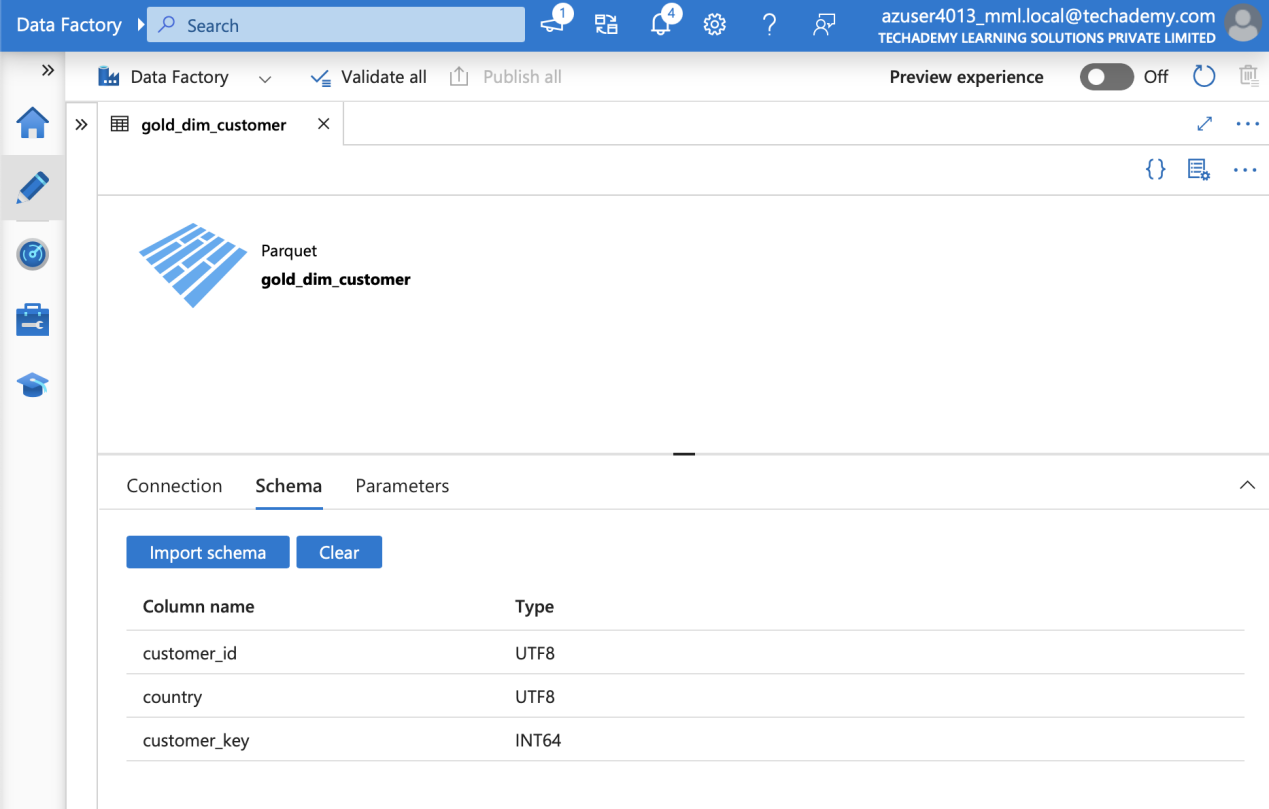
* Correct record counts between source and target.
* Schema alignment after transformations.
* Clean and structured data in the Gold layer, ready for reporting.
* Bronze/ Original data schema:



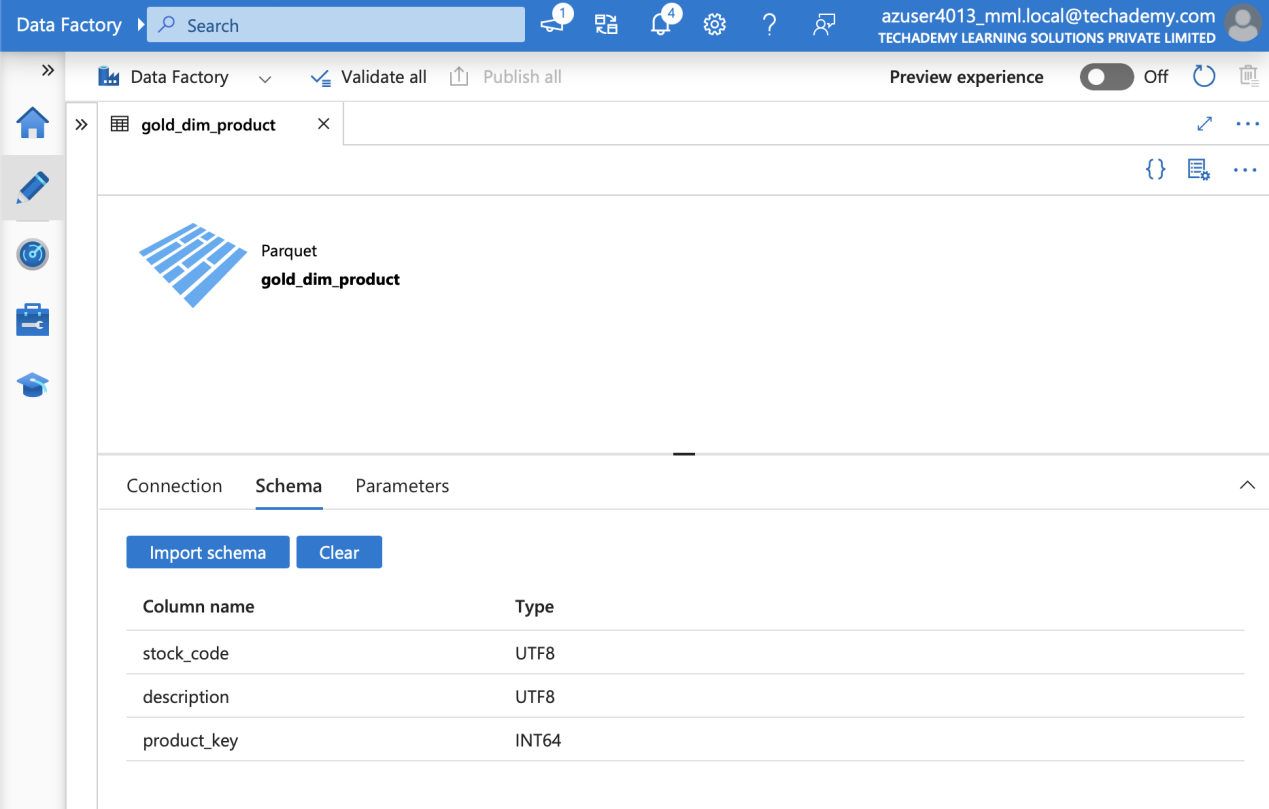
* Silver data Schema:



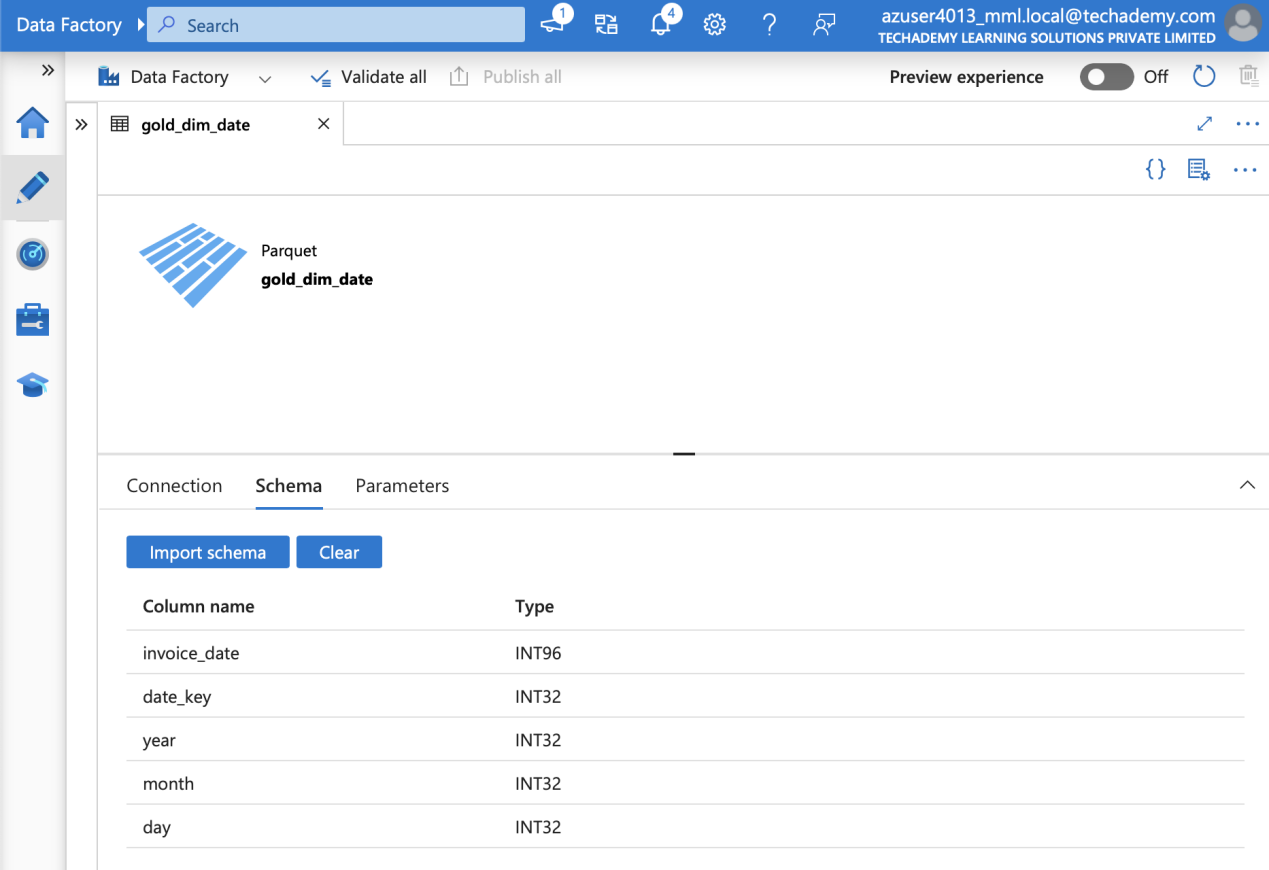
* Gold-Customer data schema:



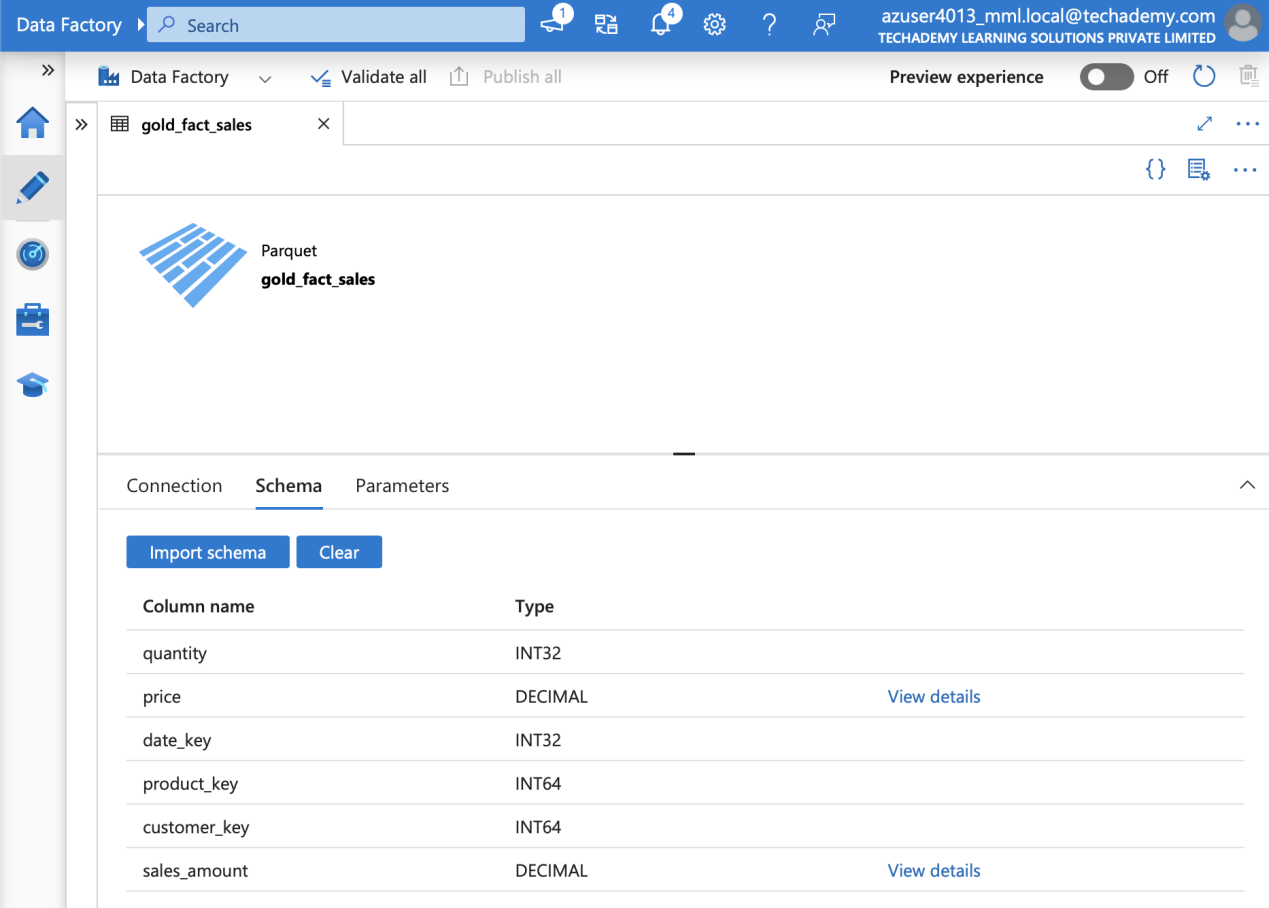
* Gold-Product data Schema:



* Gold-Date data Schema:



* Gold Fact table Schema:



## 8. Summary & Conclusion

The project successfully demonstrated the implementation of a **multi-layered data lake architecture** using ADF and ADLS.

All pipelines executed successfully, moving data from Bronze → Silver → Gold.

Validation outputs confirm correctness and consistency.

The Gold layer now provides curated data for business analytics.