

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.

The screenshot shows the Microsoft Azure portal interface. At the top, there's a search bar and a user profile. Below the navigation bar, the 'Recent' tab is active, displaying a list of resources. The list is filtered by 'Subscription equals all', 'Resource Group equals all', 'Type equals all', and 'Location equals all'. The table below shows the following resources:

Name	Type	Location	Resource Group	Subscription	Last accessed
HexaDataFactory1	Data factory (V2)	South India	rg-azuser4013_mml.loca...	MML Learners	3 minutes ago
hexastorage1	Storage account		rg-azuser4013_mml.loca...	MML Learners	10 minutes ago
HexaWorkspace	Azure Databricks Service	South India	rg-azuser4013_mml.loca...	MML Learners	20 hours ago
rg-azuser4013_mml.local-3YyWq	Resource group	Central India	rg-azuser4013_mml.loca...	MML Learners	4 days ago

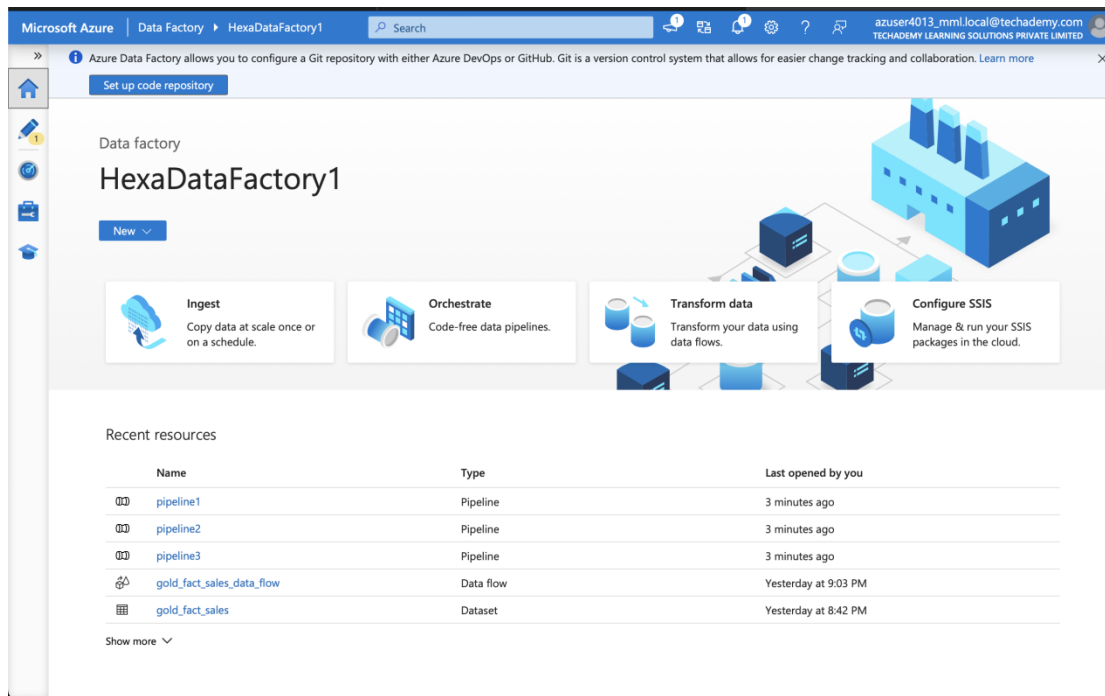
At the bottom, it says 'Showing 1 - 4 of 4. Display count: auto' and there is a 'Give feedback' link.

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

The screenshot shows the 'Containers' view for the 'hexastorage1' storage account. The left sidebar contains a navigation menu with options like Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, Storage browser, Partner solutions, Resource visualizer, Data storage, Containers, File shares, Queues, Tables, Security + networking, Data management, Settings, and Configuration. The main area displays a list of containers with the following columns: Name, Last modified, Anonymous access level, and Lease state. The containers listed are \$logs, bronze, gold, raw, and silver.

Name	Last modified	Anonymous access level	Lease state
\$logs	21/8/2025, 5:16:40 PM	Private	Available
bronze	25/8/2025, 10:31:47 AM	Private	Available
gold	25/8/2025, 10:32:00 AM	Private	Available
raw	23/8/2025, 10:44:37 AM	Private	Available
silver	25/8/2025, 10:31:55 AM	Private	Available

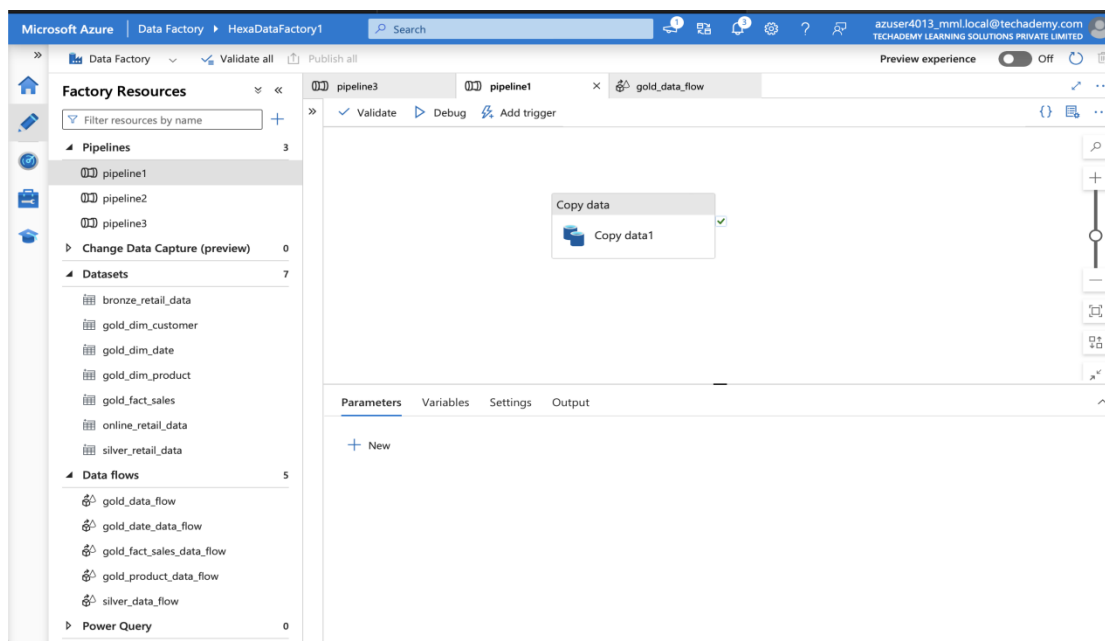
- An ADF workspace configured with Managed Identity for secure access.



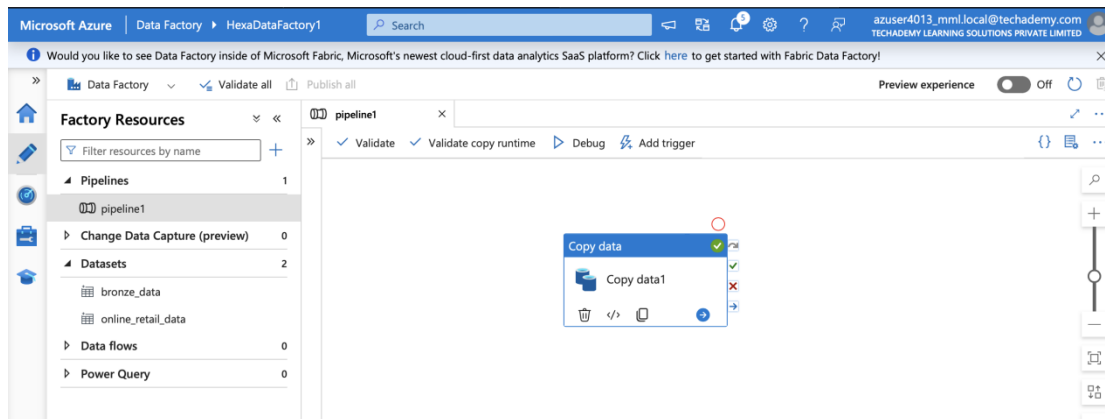
4. DATA INGESTION (BRONZE LAYER)

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

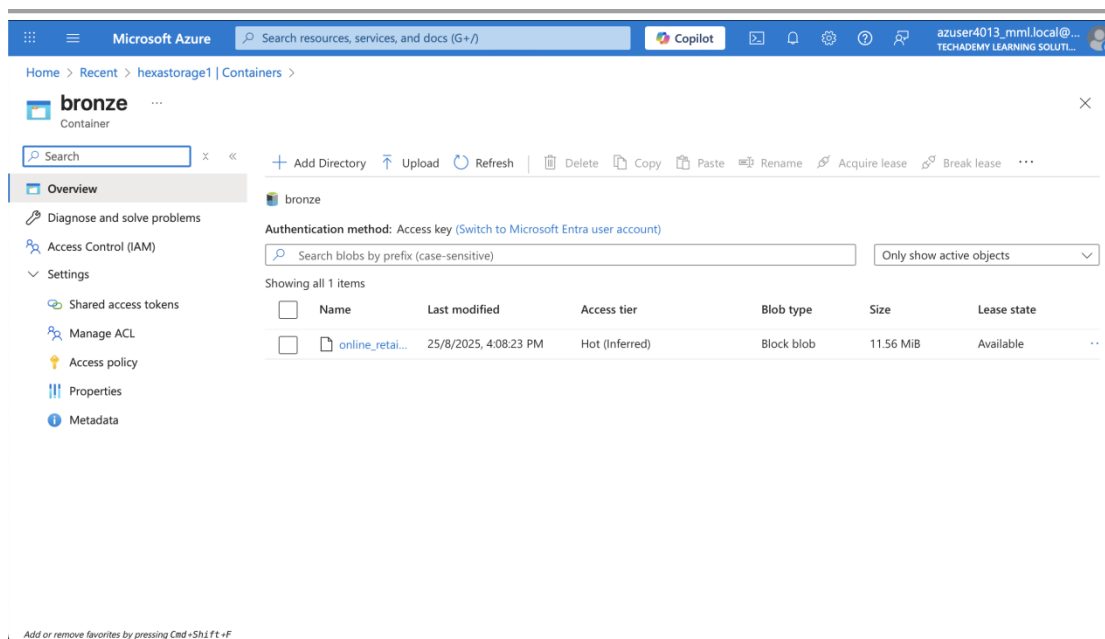
- ADF Copy Activity pipeline (Designer view)



- Monitoring tab showing successful pipeline run



- ADLS Bronze container showing ingested files

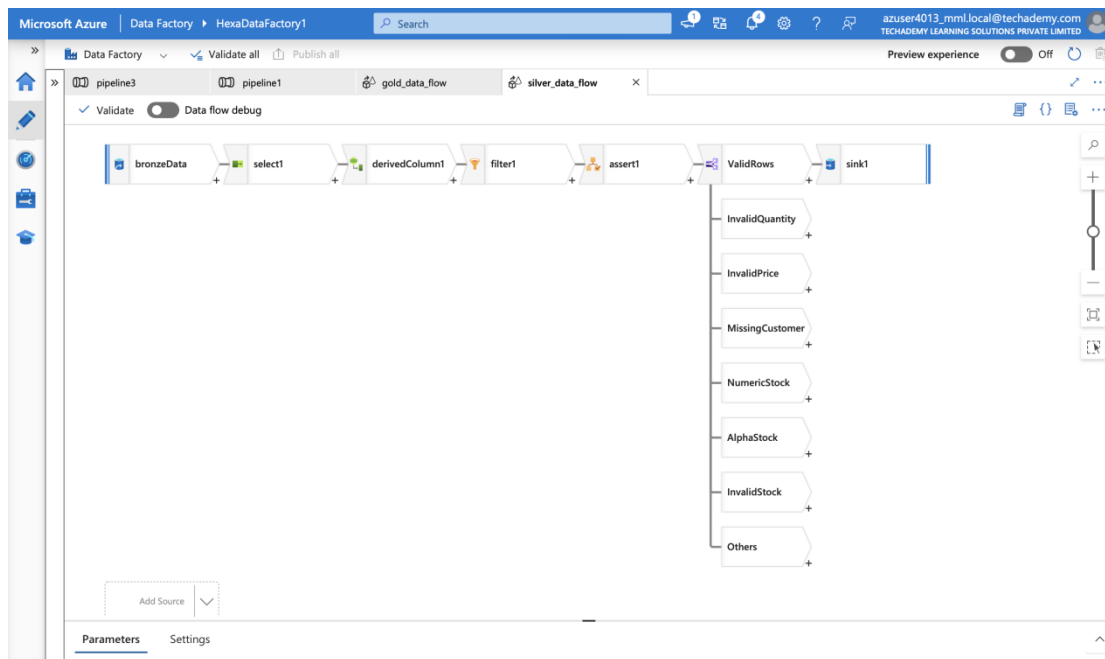


5. DATA TRANSFORMATION (SILVER LAYER)

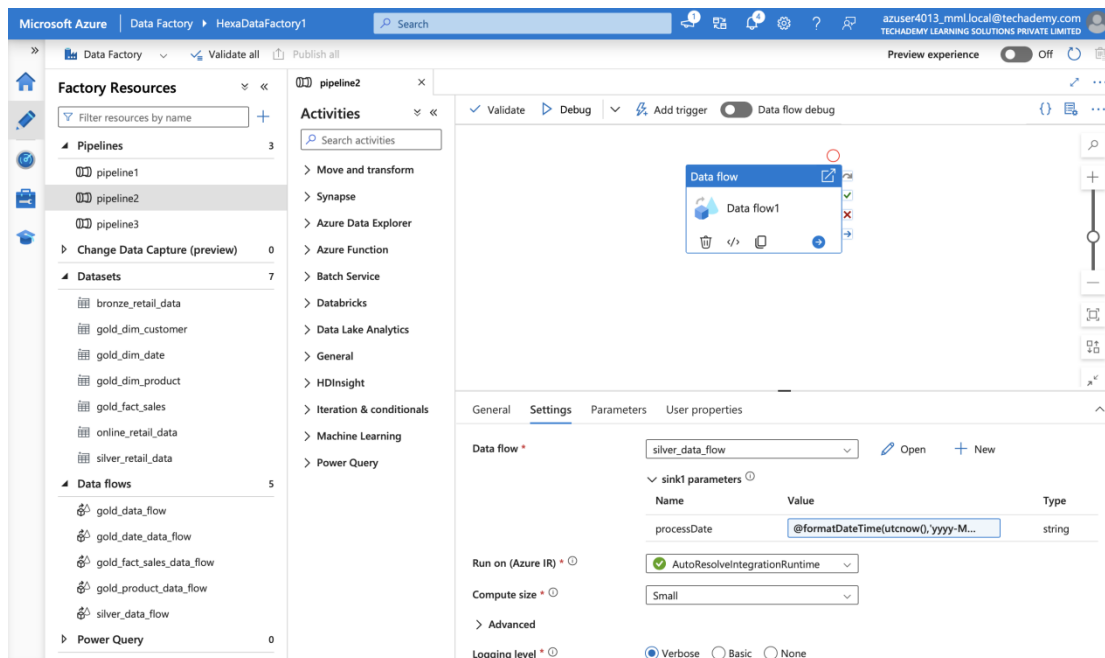
Data was transformed in ADF Mapping Data Flows, including:

- Removing null/duplicate records.
- Standardizing date and numeric formats.
- Creating derived attributes.
- The transformed data was stored in the **Silver container**.

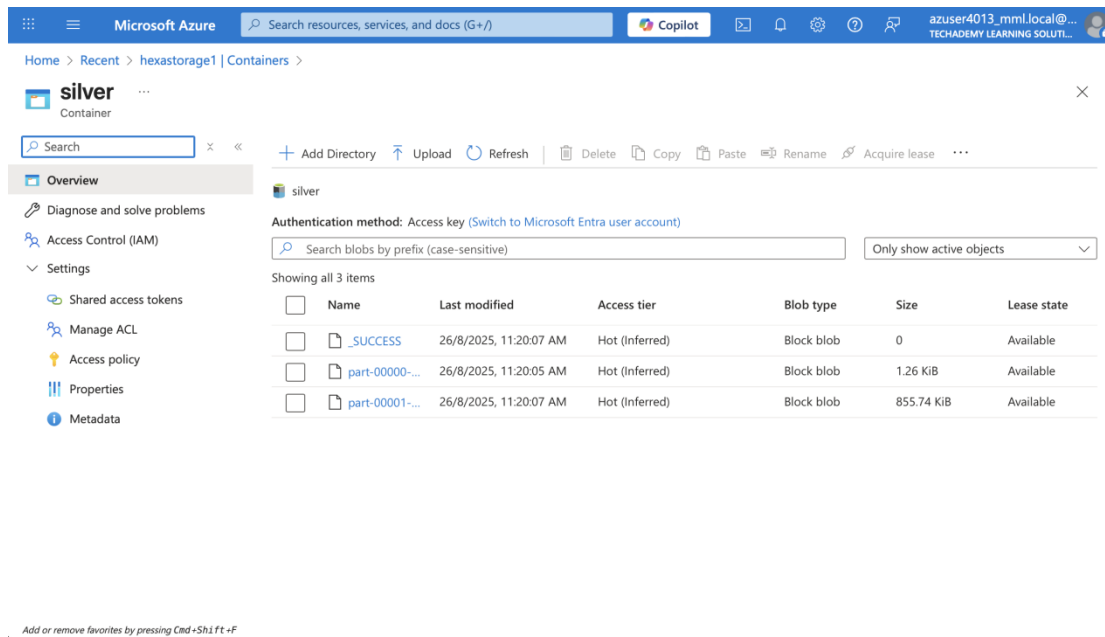
- ADF Mapping Data Flow (showing transformations: removing nulls/duplicates, derived columns)



- Pipeline execution success in Monitoring tab



- ADLS Silver container with transformed files



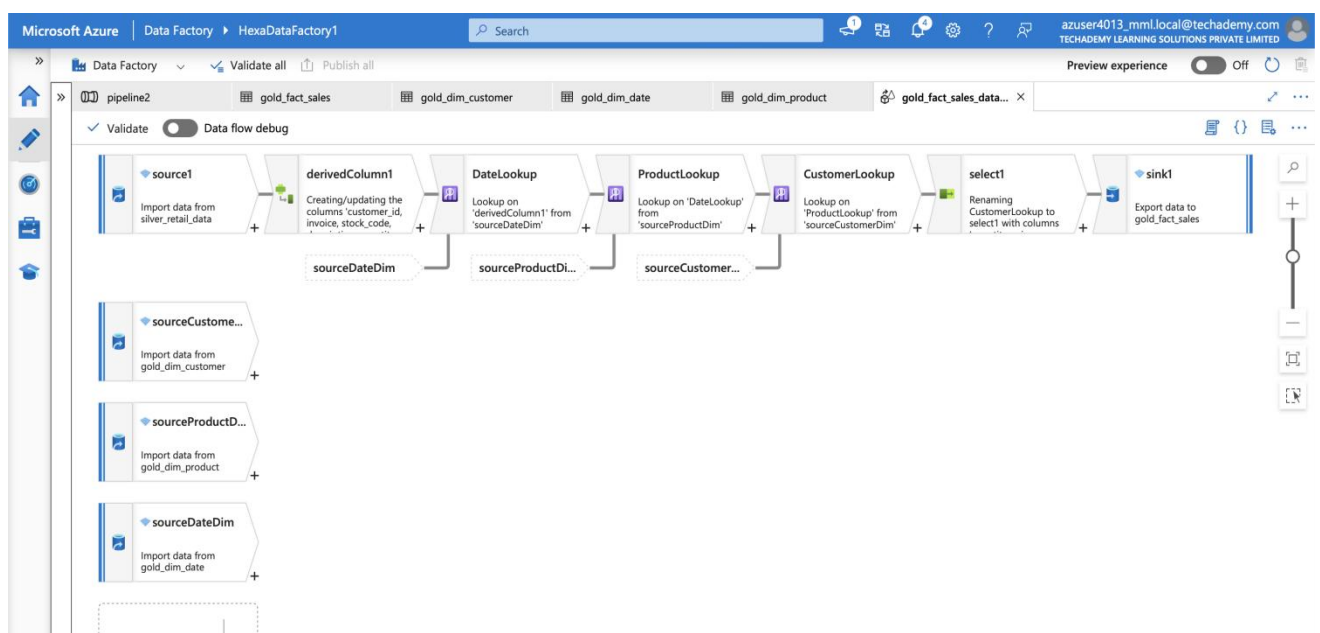
6. CURATED DATA (GOLD LAYER)

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



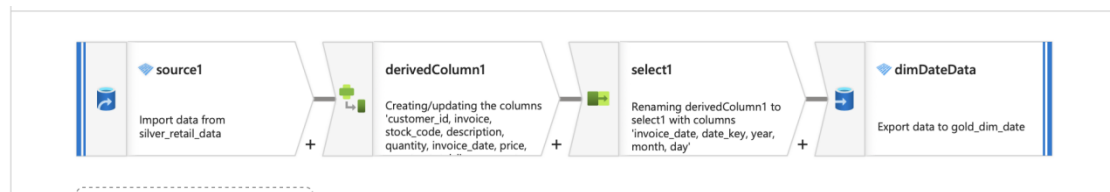
➤ ADF Gold product Data flow:



➤ ADF Gold Customer Data flow:



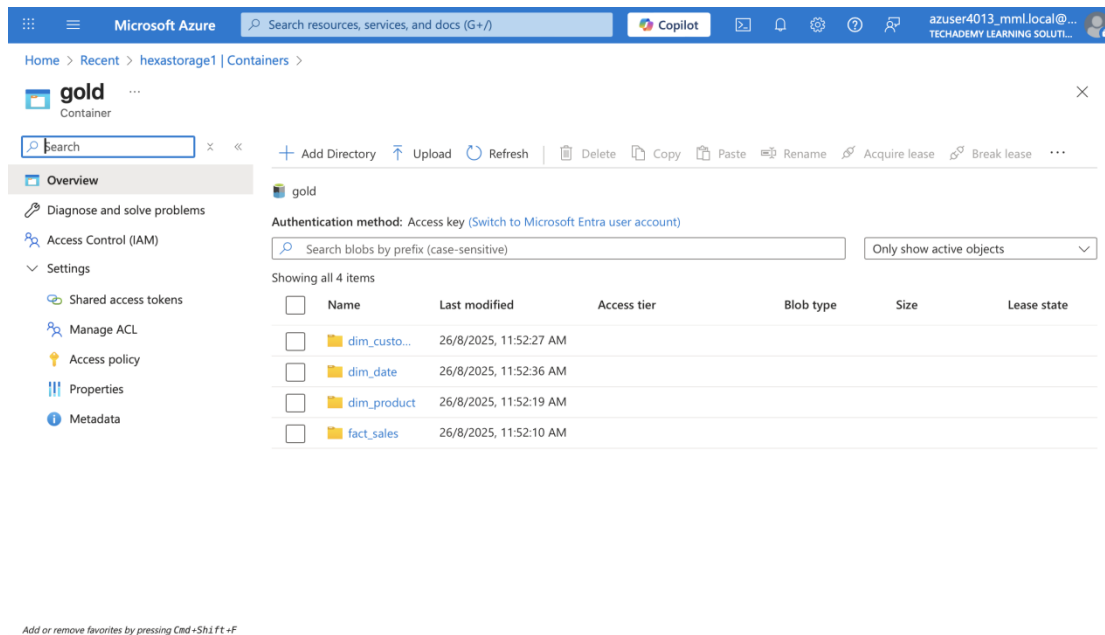
➤ ADF Gold Date data flow:



The screenshot shows the Microsoft Azure Data Factory interface for 'HexaDataFactory1'. The left sidebar displays 'Factory Resources' including Pipelines (pipeline1, pipeline2, pipeline3), Datasets (bronze_retail_data, gold_dim_customer, gold_dim_date, gold_dim_product, gold_fact_sales, online_retail_data, silver_retail_data), and Data flows (gold_fact_sales_data_flow, gold_data_flow, gold_date_data_flow, gold_product_data_flow, silver_data_flow). The main canvas shows 'pipeline3' with four data flow tasks: 'goldDataFlow', 'goldProductDataFlow', 'goldDateDataFlow', and 'goldFactSales'. The 'Settings' tab for 'gold_fact_sales_data_flow' is active, showing parameters for 'source1' and 'sink1'.

Name	Value	Type
processDate	@formatDateTime(utcnow(), 'yyyy-MM-')	string
year	@formatDateTime(utcnow(), 'yyyy')	string
month	@formatDateTime(utcnow(), 'MM')	string

➤ ADLS Gold container with structured files

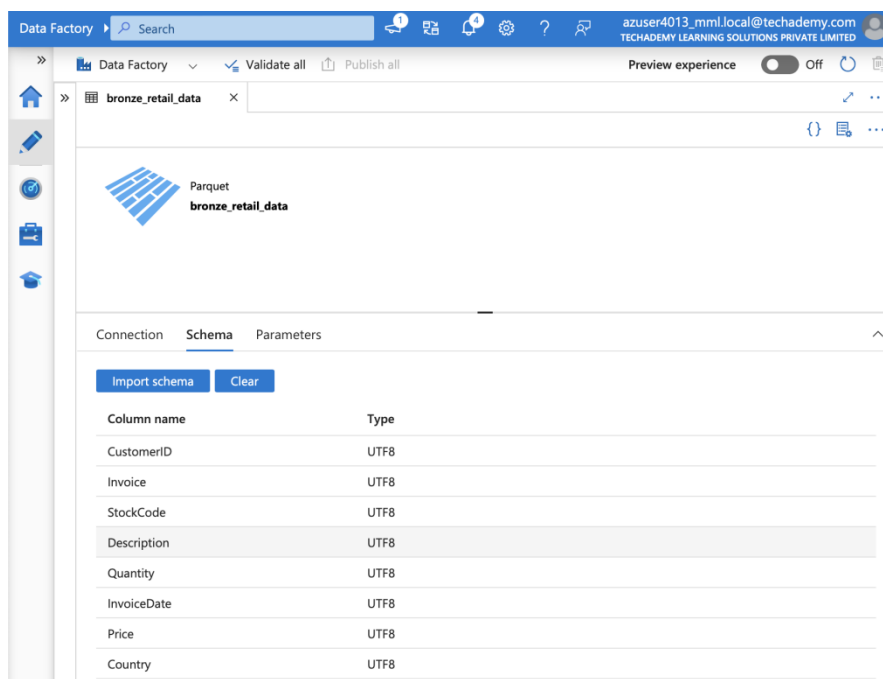


7. VALIDATION & OUTPUTS

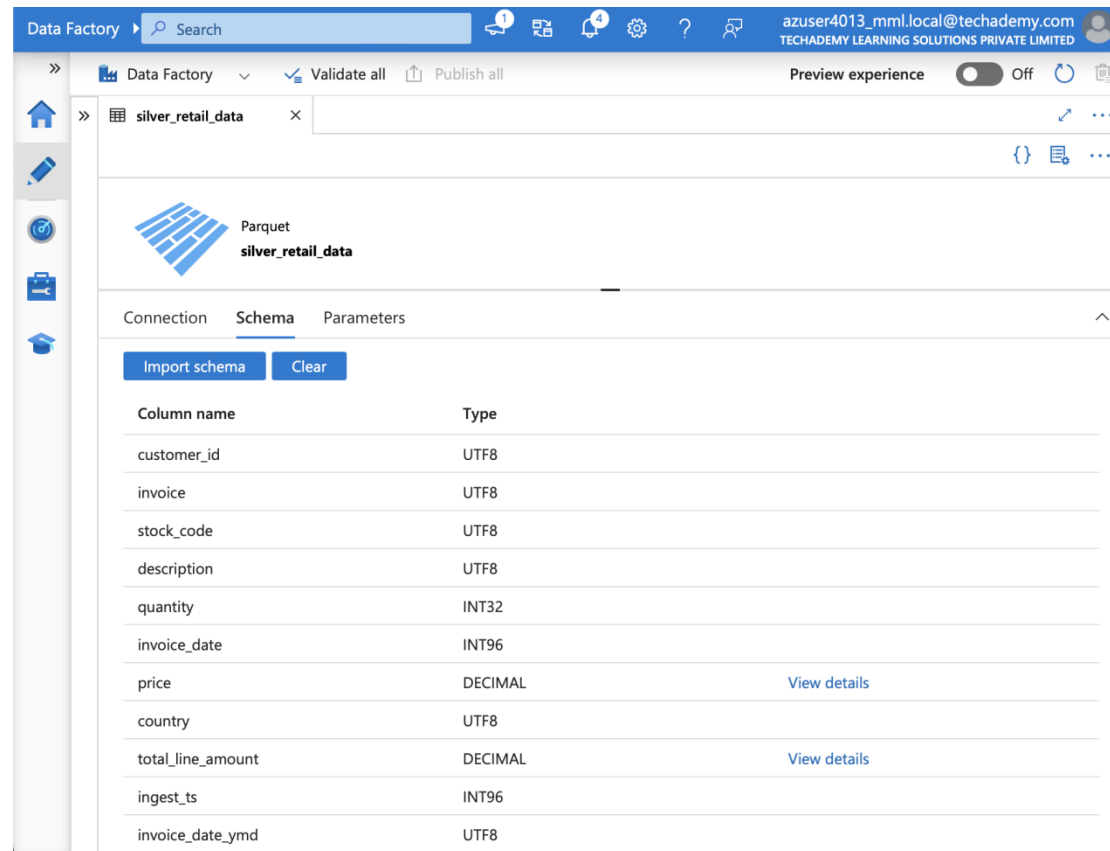
Validation checks confirmed:

- 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:



Data Factory

Search

azuser4013_mml.local@techademy.com
TECHADEMY LEARNING SOLUTIONS PRIVATE LIMITED

Data Factory Validate all Publish all Preview experience Off

silver_retail_data

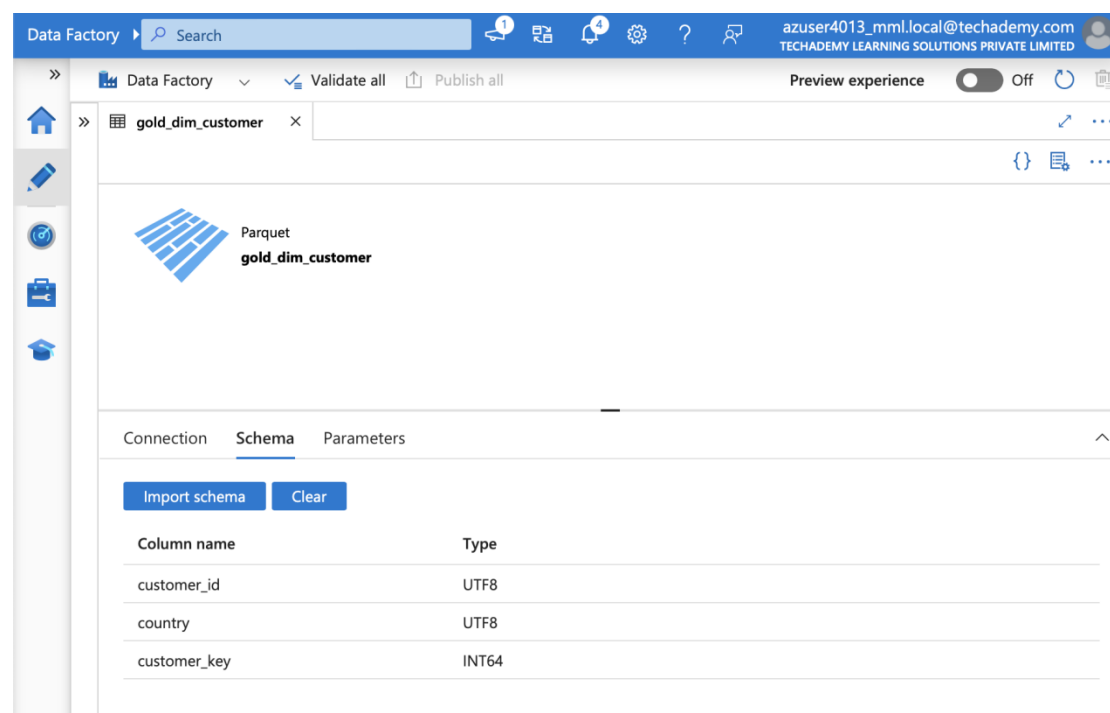
Parquet
silver_retail_data

Connection Schema Parameters

Import schema Clear

Column name	Type	
customer_id	UTF8	
invoice	UTF8	
stock_code	UTF8	
description	UTF8	
quantity	INT32	
invoice_date	INT96	
price	DECIMAL	View details
country	UTF8	
total_line_amount	DECIMAL	View details
ingest_ts	INT96	
invoice_date_ymd	UTF8	

➤ Gold-Customer data schema:



Data Factory

Search

azuser4013_mml.local@techademy.com
TECHADEMY LEARNING SOLUTIONS PRIVATE LIMITED

Data Factory Validate all Publish all Preview experience Off

gold_dim_customer

Parquet
gold_dim_customer

Connection Schema Parameters

Import schema Clear

Column name	Type	
customer_id	UTF8	
country	UTF8	
customer_key	INT64	

➤ Gold-Product data Schema:

The screenshot shows the Data Factory interface with the 'gold_dim_product' dataset selected. The 'Schema' tab is active, displaying a table with the following columns and types:

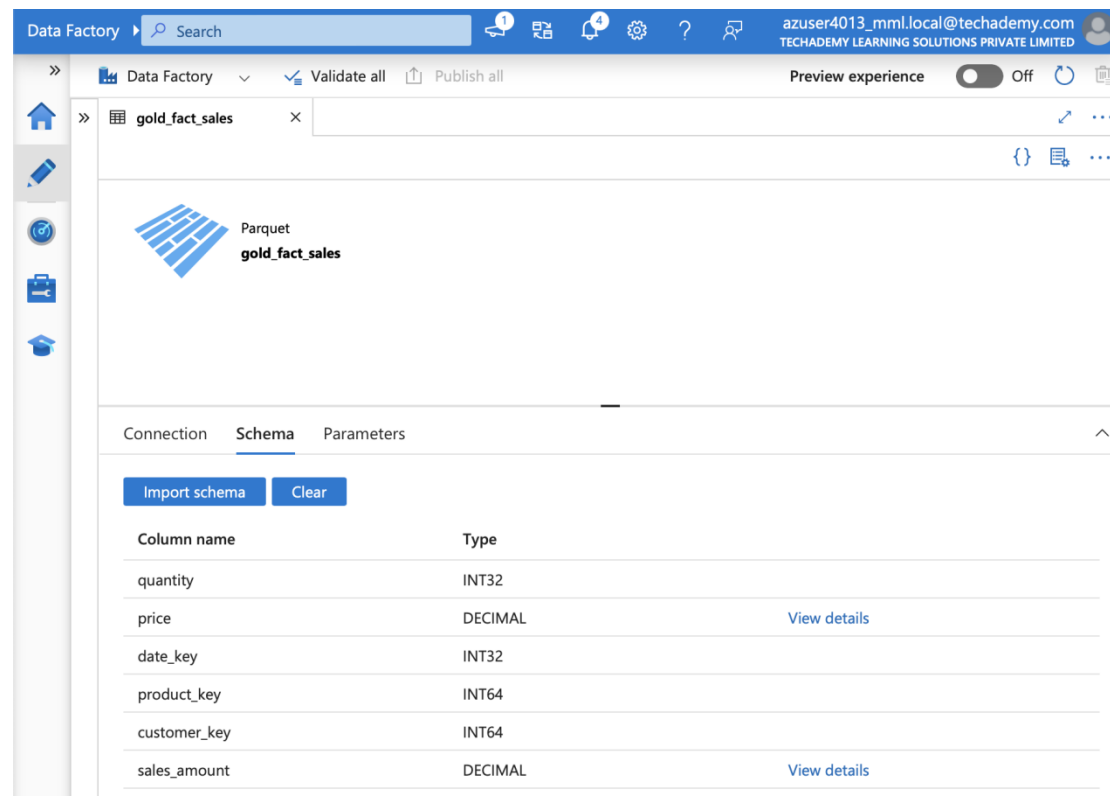
Column name	Type
stock_code	UTF8
description	UTF8
product_key	INT64

➤ Gold-Date data Schema:

The screenshot shows the Data Factory interface with the 'gold_dim_date' dataset selected. The 'Schema' tab is active, displaying a table with the following columns and types:

Column name	Type
invoice_date	INT96
date_key	INT32
year	INT32
month	INT32
day	INT32

➤ Gold Fact table Schema:



Column name	Type	
quantity	INT32	
price	DECIMAL	View details
date_key	INT32	
product_key	INT64	
customer_key	INT64	
sales_amount	DECIMAL	View details

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.