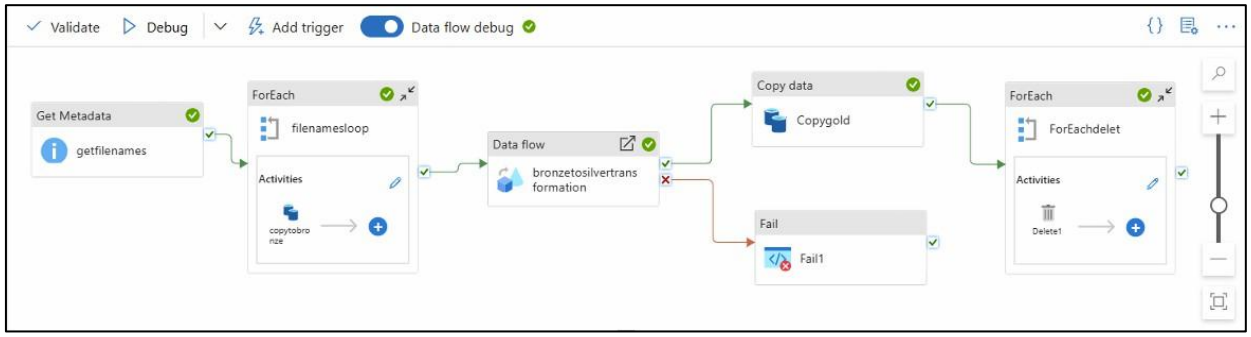
End-to-End Azure Data Engineering Project: Validating and Matching Trade Transactions Using ADF and ADLS

Project Overview:

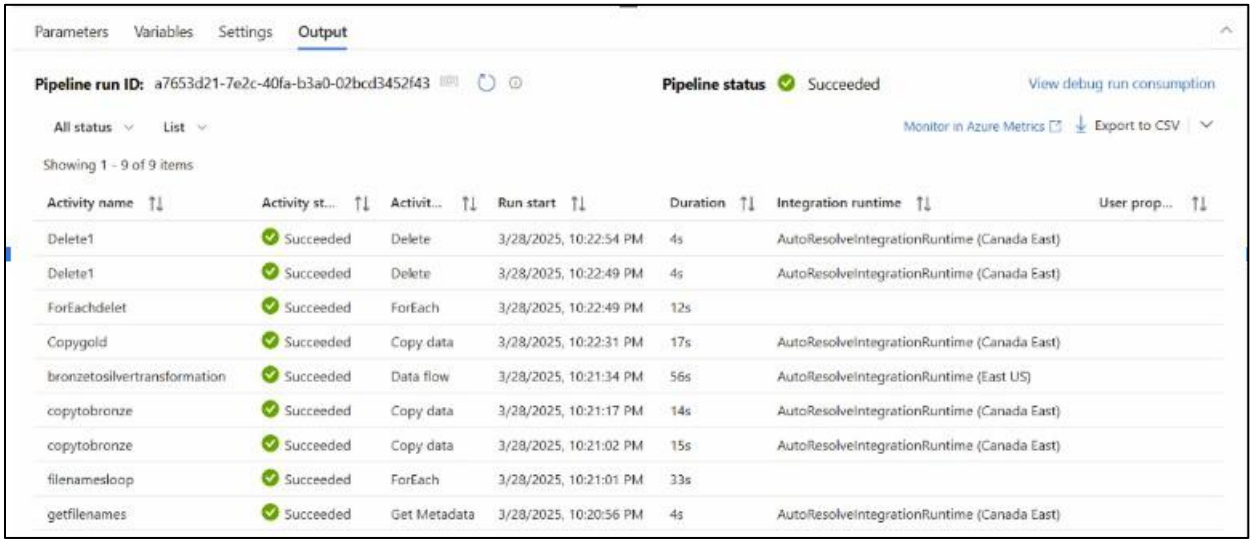
This project focuses on building a data pipeline to validate and match the buy and sell legs of trade transactions, ensuring only correct and complete trades are passed through to downstream systems.

- **Source**: Trade data (buy and sell legs) ingested daily from **Blob Storage (Raw folder)**.
- **Target**: Cleaned and validated trades stored in **Azure Data Lake Gen2 (Gold folder)**, with error records routed to a separate error path.
- **Technology Stack**: Azure Data Factory (ADF) with **Data Flows**, **ADLS Gen2**, and **parameterized pipelines**.

END_TO_END pipeline:



Debug resul

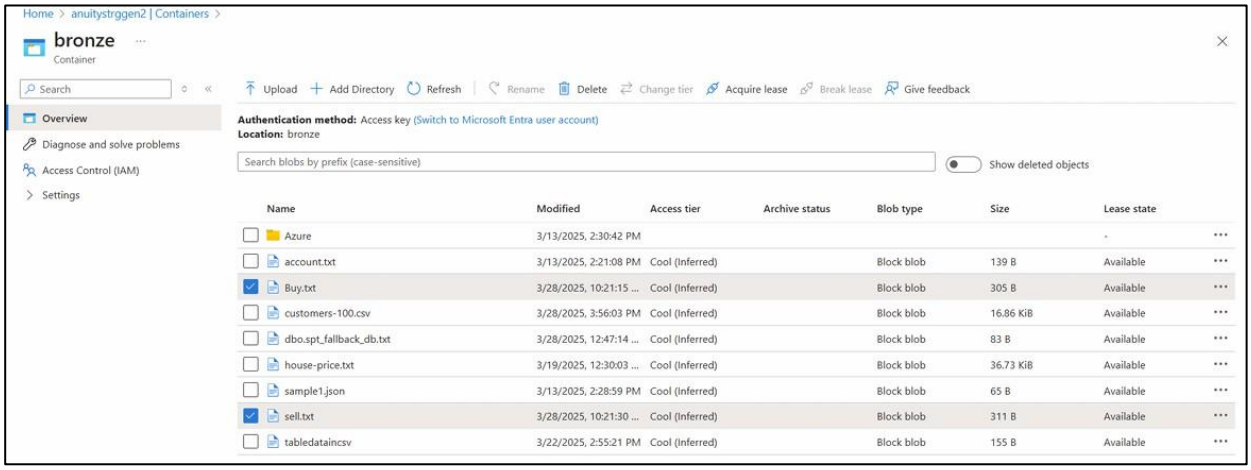


Activity name	Activity st...	Activit...	Run start	Duration	Integration runtime	User prop...
Delete1	Succeeded	Delete	3/28/2025, 10:22:54 PM	4s	AutoResolveIntegrationRuntime (Canada East)	
Delete1	Succeeded	Delete	3/28/2025, 10:22:49 PM	4s	AutoResolveIntegrationRuntime (Canada East)	
ForEachdelet	Succeeded	ForEach	3/28/2025, 10:22:49 PM	12s		
Copygold	Succeeded	Copy data	3/28/2025, 10:22:31 PM	17s	AutoResolveIntegrationRuntime (Canada East)	
bronzetosilvertransformation	Succeeded	Data flow	3/28/2025, 10:21:34 PM	56s	AutoResolveIntegrationRuntime (East US)	
copytobronze	Succeeded	Copy data	3/28/2025, 10:21:17 PM	14s	AutoResolveIntegrationRuntime (Canada East)	
copytobronze	Succeeded	Copy data	3/28/2025, 10:21:02 PM	15s	AutoResolveIntegrationRuntime (Canada East)	
filenamesloop	Succeeded	ForEach	3/28/2025, 10:21:01 PM	33s		
getfilenames	Succeeded	Get Metadata	3/28/2025, 10:20:56 PM	4s	AutoResolveIntegrationRuntime (Canada East)	

Transformation Steps (ETL Process)

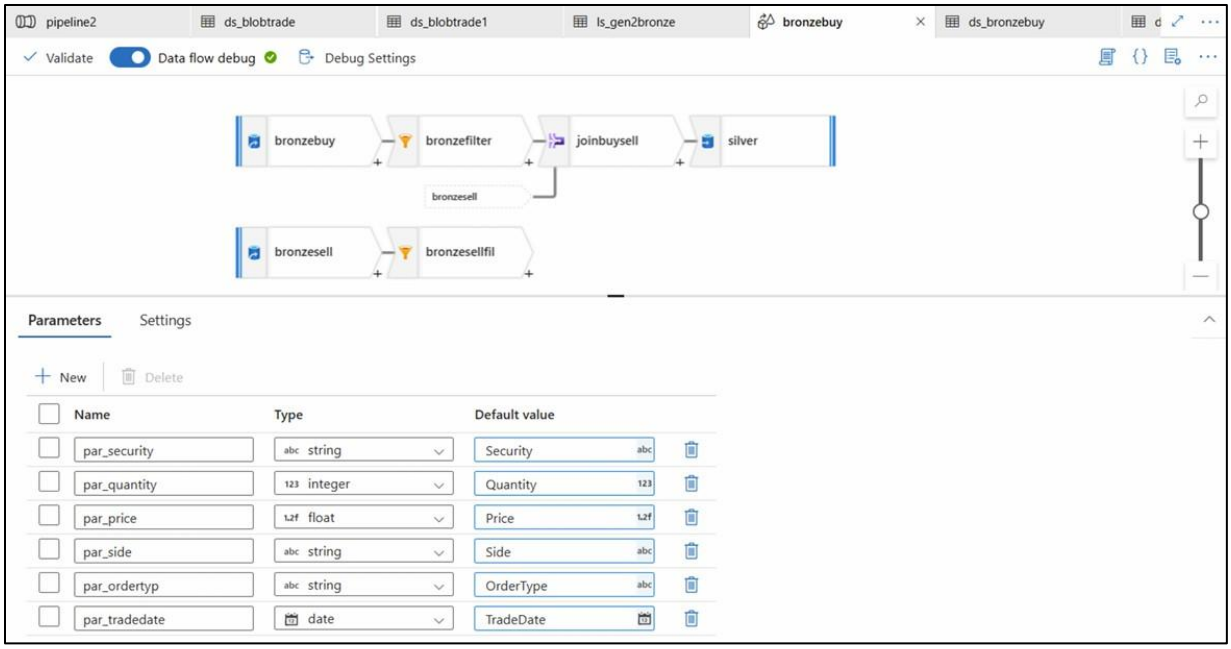
1. Raw Ingestion

- Buy and sell trade files are placed in the **Blob Storage raw folder**.
- Files are then copied into the **Bronze layer** in **ADLS Gen2**, preserving the original data.

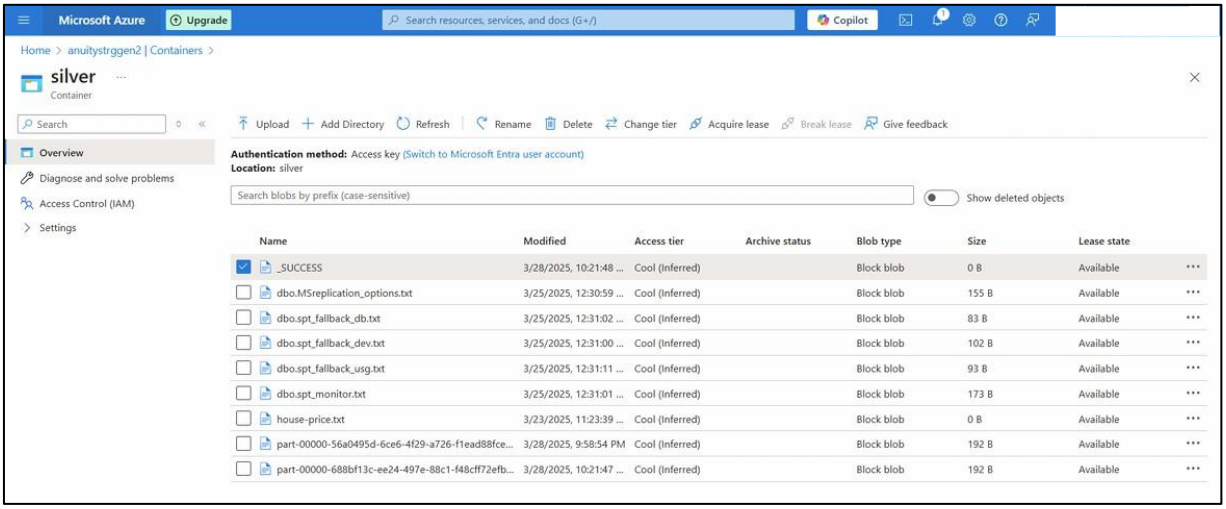


2. Trade Validation and Matching (Silver Layer)

- **Using ADF Data Flows:**
    - Read buy and sell trade files from the Bronze folder.
    - Filter each to remove invalid records (e.g., nulls, format errors).
    - Join buy and sell legs on matching criteria (e.g., trade ID, quantity) to form a
    - complete trade.
    - Store successfully matched trades into the Silver folder in ADLS Gen2.

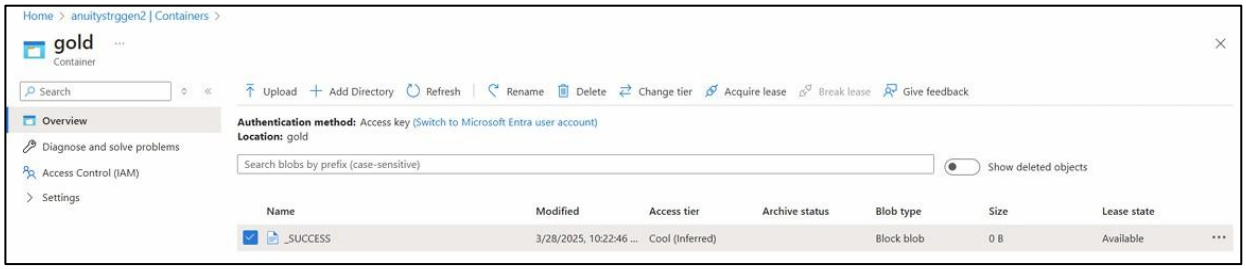


Files in silver layer:



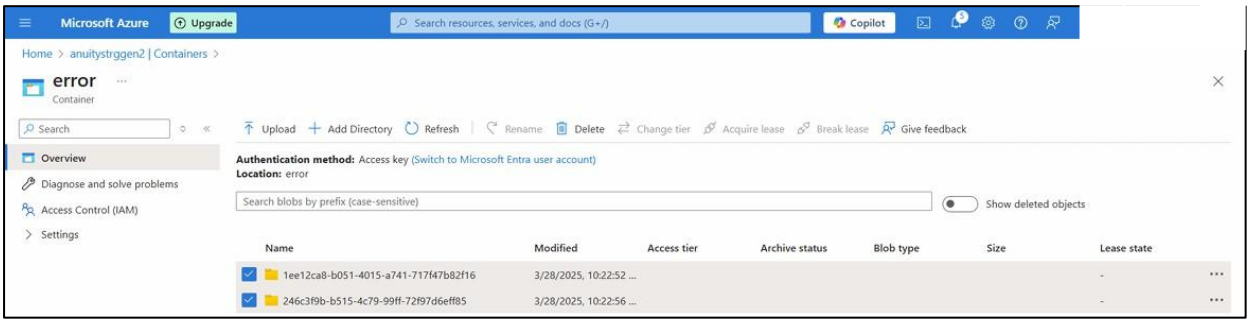
3. Column Selection & Gold Layer Writing

- Drop unnecessary or repeated columns from the Silver data.
- Pass only the required columns using parameters back to the ADF pipeline.
- Store the final, clean trade data in the Gold folder in ADLS Gen2.



4. Clean-Up

- After processing, the original raw files are deleted from Blob Storage to make room for the next batch.



Conclusion

- Delivered an automated, robust pipeline that ensures only valid and matched trades move forward for reporting or settlement.
- Maintained a clear data lineage by separating Bronze, Silver, and Gold layers in ADLS Gen2.
- Empowered business users with access to error files for timely resolution and data quality improvement.
- Enabled scalability and efficiency by cleaning up processed files and preparing the system for the next daily run.