

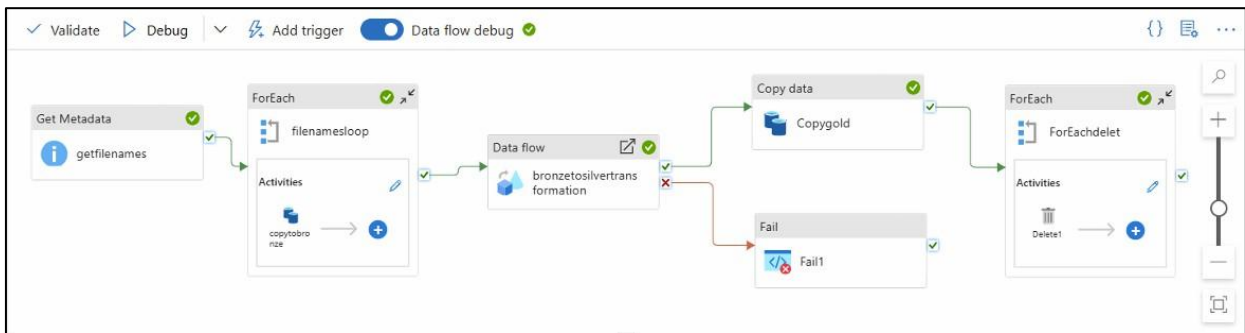
# **End-to-End Azure Data Engineering Project: Validating and Matching Trade Transactions Using ADF and ADLS**

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

Parameters

Variables

Settings

Output

Pipeline run ID: a7653d21-7e2c-40fa-b3a0-02bcd3452f43

Pipeline status ✓ Succeeded

View debug run consumption

All status ▼

List ▼

Monitor in Azure Metrics 📊

Export to CSV 📄

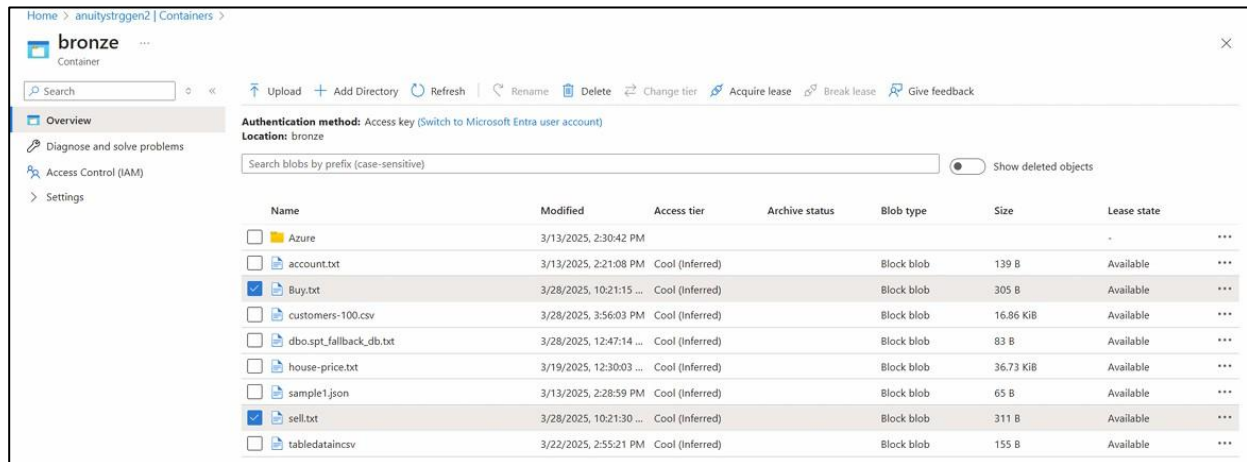
Showing 1 - 9 of 9 items

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



Home > anu1ystrggen2 | Containers >

bronze Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Authentication method: Access key (Switch to Microsoft Entra user account)

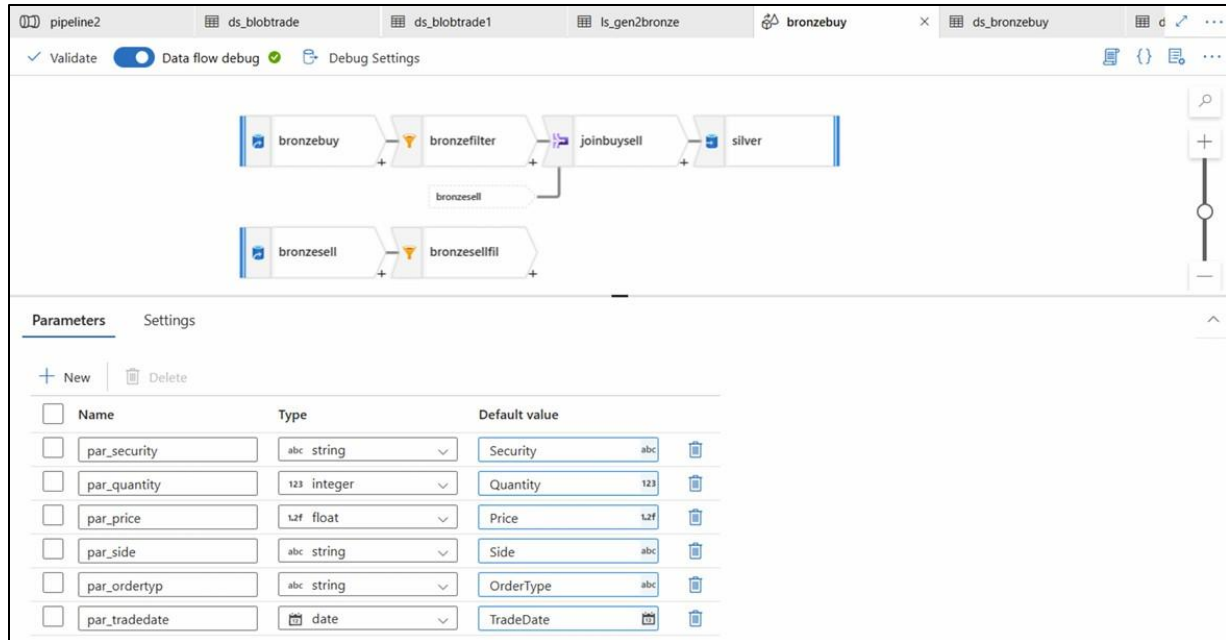
Location: bronze

Search blobs by prefix (case-sensitive) Show deleted objects

| Name   | Modified                | Access tier     | Archive status | Blob type  | Size      | Lease state |
|--|-------------------------|-----------------|----------------|------------|-----------|-------------|
| <input type="checkbox"/> Azure                   | 3/13/2025, 2:30:42 PM   |                 |                |            |           | -           |
| <input type="checkbox"/> account.txt             | 3/13/2025, 2:21:08 PM   | Cool (Inferred) |                | Block blob | 139 B     | Available   |
| <input checked="" type="checkbox"/> Buy.txt      | 3/28/2025, 10:21:15 ... | Cool (Inferred) |                | Block blob | 305 B     | Available   |
| <input type="checkbox"/> customers-100.csv       | 3/28/2025, 3:56:03 PM   | Cool (Inferred) |                | Block blob | 16.86 KiB | Available   |
| <input type="checkbox"/> dbo.spt_fallback_db.txt | 3/28/2025, 12:47:14 ... | Cool (Inferred) |                | Block blob | 83 B      | Available   |
| <input type="checkbox"/> house-price.txt         | 3/19/2025, 12:30:03 ... | Cool (Inferred) |                | Block blob | 36.73 KiB | Available   |
| <input type="checkbox"/> sample1.json            | 3/13/2025, 2:28:59 PM   | Cool (Inferred) |                | Block blob | 65 B      | Available   |
| <input checked="" type="checkbox"/> sell.txt     | 3/28/2025, 10:21:30 ... | Cool (Inferred) |                | Block blob | 311 B     | Available   |
| <input type="checkbox"/> tabledata1.csv          | 3/22/2025, 2:55:21 PM   | Cool (Inferred) |                | Block blob | 155 B     | Available   |

### 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.



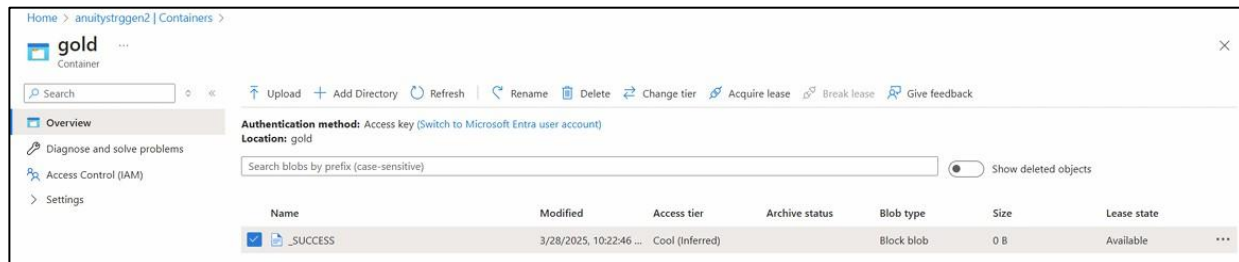
## Files in silver layer:

The screenshot shows the Microsoft Azure portal interface for a container named 'silver'. The table below lists the files and their properties.

| Name  | Modified                | Access tier     | Archive status | Blob type  | Size  | Lease state |
|---|-------------------------|-----------------|----------------|------------|-------|-------------|
| <input checked="" type="checkbox"/> _SUCCESS                              | 3/28/2025, 10:21:48 ... | Cool (Inferred) |                | Block blob | 0 B   | Available   |
| <input type="checkbox"/> dbo.MSreplication_options.txt                    | 3/25/2025, 12:30:59 ... | Cool (Inferred) |                | Block blob | 155 B | Available   |
| <input type="checkbox"/> dbo.spt_fallback_db.txt                          | 3/25/2025, 12:31:02 ... | Cool (Inferred) |                | Block blob | 83 B  | Available   |
| <input type="checkbox"/> dbo.spt_fallback_dev.txt                         | 3/25/2025, 12:31:00 ... | Cool (Inferred) |                | Block blob | 102 B | Available   |
| <input type="checkbox"/> dbo.spt_fallback_usg.txt                         | 3/25/2025, 12:31:11 ... | Cool (Inferred) |                | Block blob | 93 B  | Available   |
| <input type="checkbox"/> dbo.spt_monitor.txt                              | 3/25/2025, 12:31:01 ... | Cool (Inferred) |                | Block blob | 173 B | Available   |
| <input type="checkbox"/> house-price.txt                                  | 3/23/2025, 11:23:39 ... | Cool (Inferred) |                | Block blob | 0 B   | Available   |
| <input type="checkbox"/> part-00000-56a0495d-6ce6-4f29-a726-f1ead88fce... | 3/28/2025, 9:58:54 PM   | Cool (Inferred) |                | Block blob | 192 B | Available   |
| <input type="checkbox"/> part-00000-688bf13c-ee24-497e-88c1-f48cf72efb... | 3/28/2025, 10:21:47 ... | Cool (Inferred) |                | Block blob | 192 B | Available   |

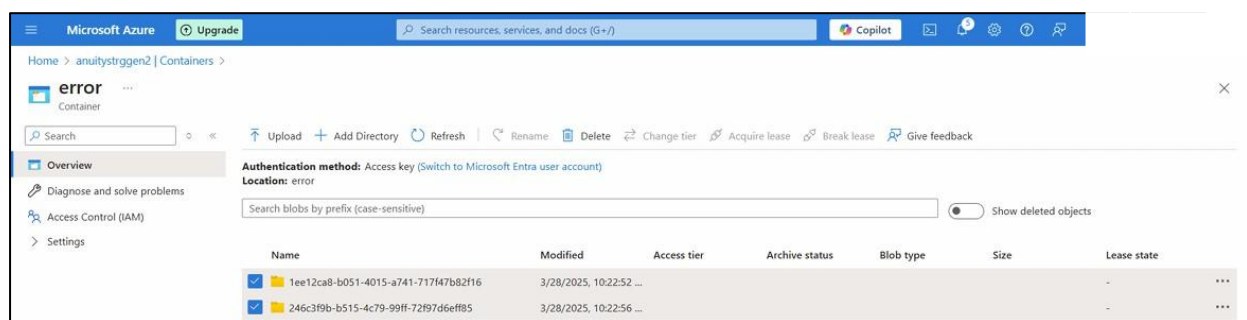
## 3. Column Selection & Gold Layer Writing

- 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.