

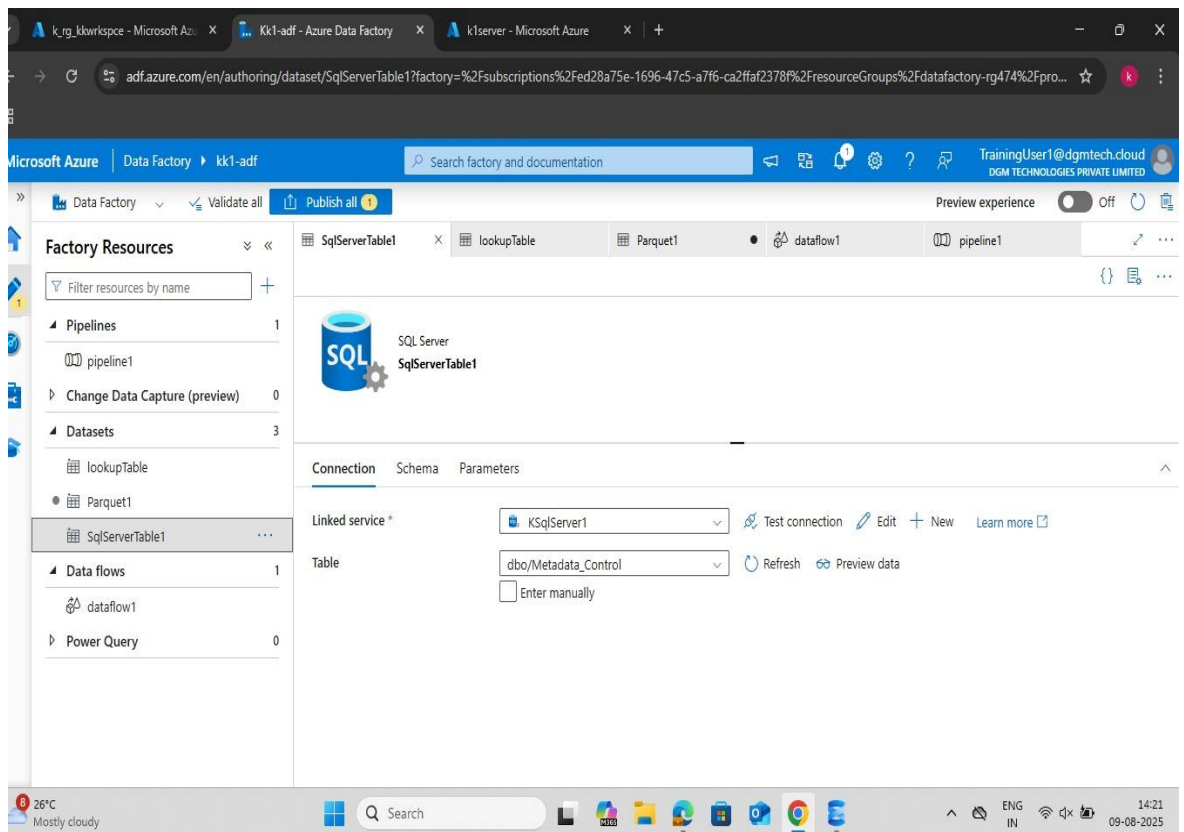
# Case Study 4 – Enterprise Data Governance & Ingestion Framework (Ingestion of Tables using Metadata Control in ADF)

- By *Kafeel Kamran*

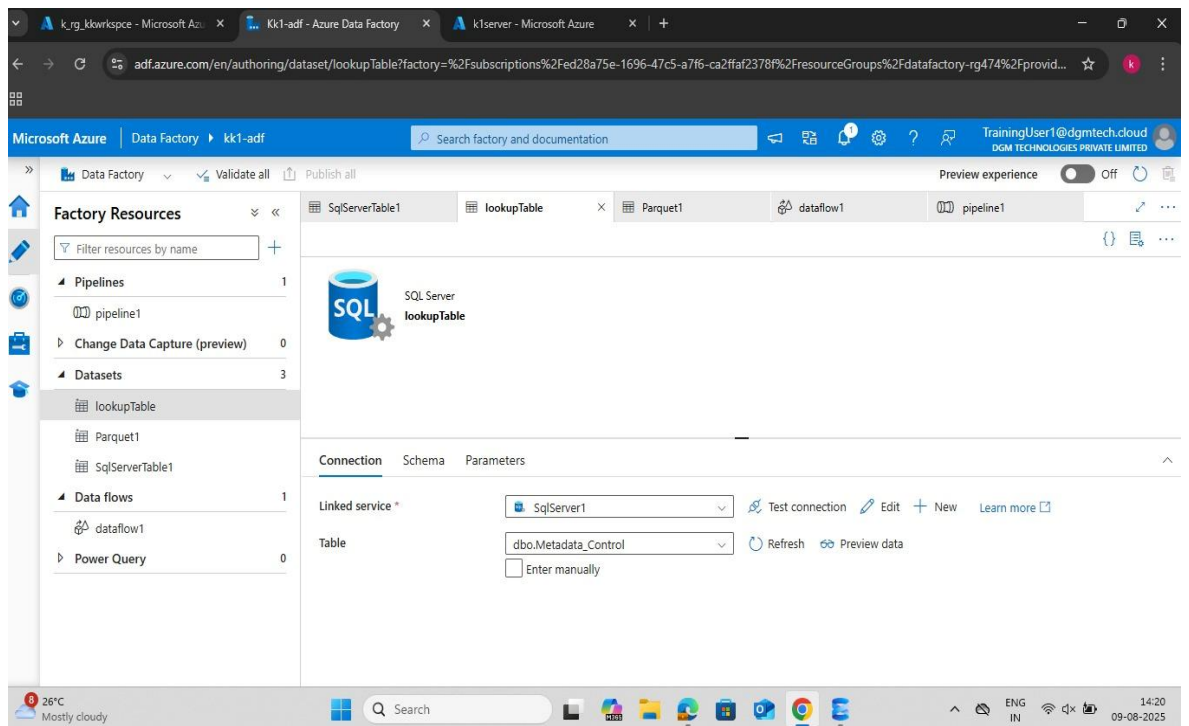
## Setting Configs and Environments

- Source Dataset – SQL Server Database Tables
- Target Sink - Azure Blob Storage Container (Parquets)
- Transformation Intermediaries – Azure Databricks

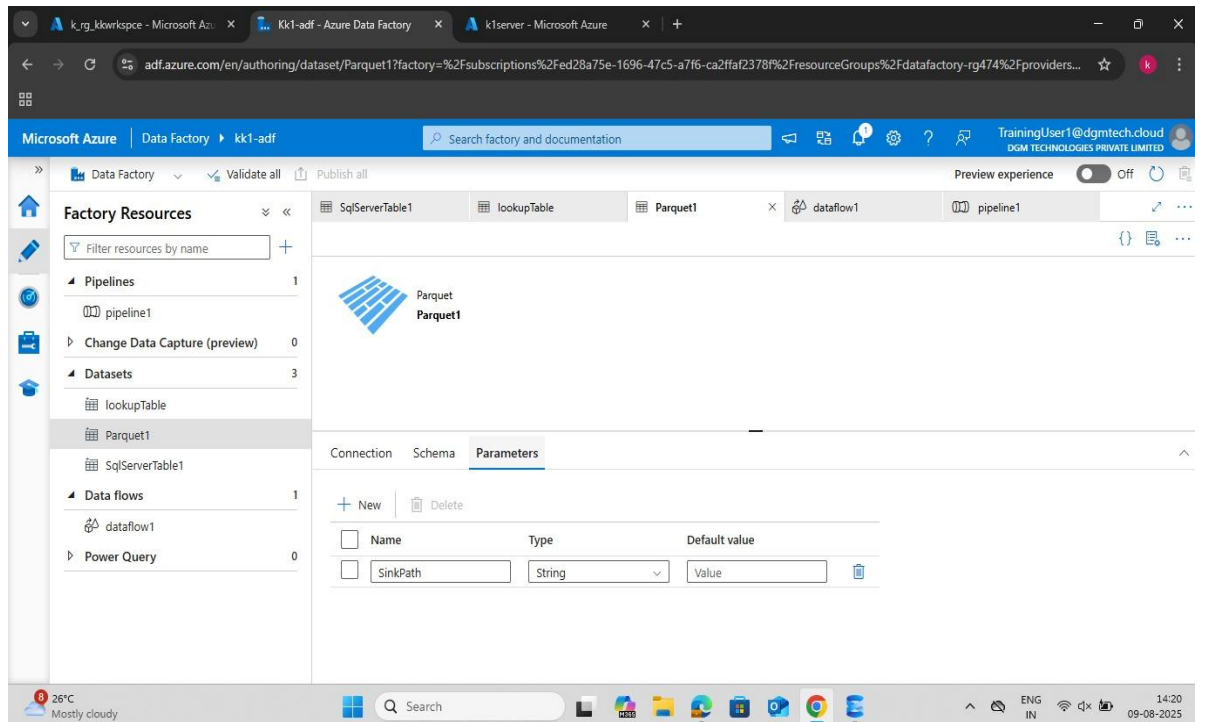
### 1. Source SQL Server with Parameters



## 2. SQL Server (Lookup Table) w/o Params



## 3. Data Sink Storage



## 4. Data Flow Structure

The screenshot shows the Microsoft Azure Data Factory interface. The main canvas displays a data flow diagram with a source node 'source1' (Import data from SqlServerTable1) and a sink node 'sink1' (Export data to Parquet1). The left sidebar shows 'Factory Resources' including Pipelines, Datasets, and Data flows. The bottom pane shows 'Parameters' with fields for Name, Type, and Default value.

Name	Type	Default value
SourceQuery	abc string	Enter expression...
SinkPath	abc string	Enter expression...

## 5. Data Flow Structure (Source)

The screenshot shows the Microsoft Azure Data Factory interface. The main canvas displays a data flow diagram with a source node 'source1' (Import data from SqlServerTable1) and a sink node 'sink1' (Export data to Parquet1). The left sidebar shows 'Factory Resources' including Pipelines, Datasets, and Data flows. The bottom pane shows 'Source settings' with fields for Output stream name, Description, Source type, and Dataset.

Output stream name *	Description	Source type *	Dataset *
source1	Import data from SqlServerTable1	Dataset	SqlServerTable1

## 6. Data Flow Structure (Sink)

The screenshot displays the Microsoft Azure Data Factory portal interface. The left-hand navigation pane shows the 'Factory Resources' section, with 'Data flows' expanded and 'dataflow1' selected. The main workspace shows a visual representation of a data flow with a source node ('source1') and a sink node ('sink1'). The sink node is highlighted, and the 'Sink' tab is active in the configuration pane on the right. The configuration pane includes fields for 'Output stream name' (set to 'sink1'), 'Description' (set to 'Export data to Parquet1'), 'Incoming stream' (set to 'source1'), and 'Sink type' (set to 'Dataset'). The 'Dataset' type is selected, and the 'Dataset' dropdown is set to 'Parquet1'. The 'Options' section shows 'Allow schema drift' checked and 'Validate schema' unchecked. The top of the interface shows the 'Data Factory' breadcrumb and the 'Validate all' button. The bottom of the interface shows the Windows taskbar with the system clock at 16:55 on 09-08-2025.

## 7. Linked Services

The screenshot displays the Microsoft Azure Data Factory portal interface, specifically the 'Linked services' section. The left-hand navigation pane shows the 'Connections' section, with 'Linked services' selected. The main workspace shows a list of linked services. The list has columns for 'Name', 'Type', 'Related', and 'Annotations'. The list contains four items: 'AzureBlobStorage1' (Azure Blob Storage), 'AzureDataBricks1' (Azure Databricks), 'KSqlServer1' (SQL Server), and 'SqlServer1' (SQL Server). The top of the interface shows the 'Data Factory' breadcrumb and the 'Validate all' button. The bottom of the interface shows the Windows taskbar with the system clock at 16:49 on 09-08-2025.

Name	Type	Related	Annotations
AzureBlobStorage1	Azure Blob Storage	1	
AzureDataBricks1	Azure Databricks	1	
KSqlServer1	SQL Server	1	
SqlServer1	SQL Server	1	

## 8. Pipeline Structure and Success

Microsoft Azure | Data Factory | Kk1-adf

All pipeline runs > pipeline1 - Activity runs

Run pipeline was modified after this run. The current pipeline configuration is shown.

Visual representation of the pipeline:

```
graph LR; Lookup1[Lookup1] --> ForEach1[ForEach1]; ForEach1 --> DataFlow1[Data flow1]; DataFlow1 --> Notebook1[Notebook1];
```

Table of pipeline runs:

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties	Activity run ID
Lookup1	Succeeded	Lookup	8/9/2025, 4:31:19 PM	11s	AutoResolveIntegrationRuntime (North Central US)		1b22f1a-c0fc-4d15-9029-2813b5bdc72
ForEach1	Succeeded	ForEach	8/9/2025, 4:31:31 PM	3m 26s	AutoResolveIntegrationRuntime (North Central US)		71f5d802-09c9-4b35-820b-c27ab061e31e
Data flow1	Succeeded	Data flow	8/9/2025, 4:31:32 PM	3m 11s	AutoResolveIntegrationRuntime (North Central US)		d430550d-b441-4078-8cdd-7e026a8c057e
Data flow1	Succeeded	Data flow	8/9/2025, 4:31:32 PM	3m 24s	AutoResolveIntegrationRuntime (North Central US)		84f15ab1-e84e-494b-8da0-26cd45b5f94e
Data flow1	Succeeded	Data flow	8/9/2025, 4:31:32 PM	3m 6s	AutoResolveIntegrationRuntime (North Central US)		d7ae8e68-516f-4542-88f8-f5c570121aa8
Notebook1	Succeeded	Notebook	8/9/2025, 4:34:57 PM	44s	AutoResolveIntegrationRuntime (North Central US)		add96568-24c5-4611-a440-0096d79969f

## 9. Databricks Transformations and Job Run

Microsoft Azure | databricks

Jobs & Pipelines

Create new

- Ingestion pipeline: Ingest data from popular apps, databases and file sources
- ETL pipeline: Build ETL pipelines using SQL and Python
- Job: Orchestrate notebooks, pipelines, queries and more

Jobs & pipelines | Job runs

Job: Run as: Start: 07-08-2025 17:30 | End: 09-08-2025 17:30 | Run status: Error code: Create job

Timeline view:

Start time | Job | Run as | Launched | Duration | Status | Error code | Run parameters

Start time	Job	Run as	Launched	Duration	Status	Error code	Run parameters
Aug 09, 2025, 04:35 PM	ADF_kk1-adf-pipeline1_Note...	TrainingUser1	By runs submit API	26s	Succeeded		



The screenshot shows the Databricks web interface. The top navigation bar includes 'Microsoft Azure' and 'databricks' logos. The left sidebar contains a 'New' button and various navigation links. The main content area displays a notebook run for 'ADF\_kk1-adf\_pipeline1'. The notebook code is visible, showing a Spark job that processes data from a Delta table. The 'Task run' details panel on the right shows the job ID, task run ID, and status (Succeeded).

**Task run Details:**

- Job ID: 27227337146427
- Task run ID: 544895376603606
- Run as: TrainingUser1
- Started: Aug 09, 2025, 04:35 PM
- Ended: Aug 09, 2025, 04:35 PM
- Duration: 26s
- Queue duration: -
- Status: Succeeded
- Lineage: No lineage information for this job.

## 10. Medallion Layers (Bronze-Silver-Gold)

The screenshot shows the Microsoft Azure portal. The top navigation bar includes 'Microsoft Azure' and 'Copilot' logos. The left sidebar contains a 'Home' button and various navigation links. The main content area displays the 'k1datastorage' Storage browser. The 'raw' container is selected, showing a list of files with columns for Name, Last modified, Access tier, Blob type, Size, and Lease state.

**Storage browser Details:**

- Authentication method: Access key (Switch to Microsoft Entra user account)
- Showing all 3 items:

Name	Last modified	Access tier	Blob type	Size	Lease state
[...]	09/08/2025, 16:22:26				
bronze	09/08/2025, 16:35:28				
bronzeGold	09/08/2025, 16:35:25				

## 11. Medallion Silver & Gold (Delta logs)

The screenshot shows the Microsoft Azure portal interface for the 'k1datastorage' storage account. The left sidebar contains navigation options like Overview, Activity log, Tags, and Storage browser. The main pane displays the 'Storage browser' view for the 'k1datastorage' account. The breadcrumb path is 'Blob containers > casestudy > raw > bronzeSilver'. The authentication method is 'Access key'. A search bar is present. The table below lists the contents of the 'bronzeSilver' container, showing 7 items. The items include a directory '[.]', a directory '\_delta\_log', and several 'part-00000...' files, all of which are 'Block blob' type and 'Available' in lease state.

Name	Last modified	Access tier	Blob type	Size	Lease state
[.]					
_delta_log	09/08/2025, 16:35:27				
part-00000...	09/08/2025, 16:35:23	Hot (Inferred)	Block blob	1.63 KiB	Available
part-00000...	09/08/2025, 16:35:26	Hot (Inferred)	Block blob	1.19 KiB	Available
part-00000...	09/08/2025, 16:35:24	Hot (Inferred)	Block blob	1.88 KiB	Available
part-00000...	09/08/2025, 16:28:31	Hot (Inferred)	Block blob	1.11 KiB	Available
part-00000...	09/08/2025, 16:28:29	Hot (Inferred)	Block blob	1.8 KiB	Available
part-00000...	09/08/2025, 16:28:24	Hot (Inferred)	Block blob	1.6 KiB	Available

The screenshot shows the Microsoft Azure portal interface for the 'k1datastorage' storage account. The left sidebar contains navigation options like Overview, Activity log, Tags, and Storage browser. The main pane displays the 'Storage browser' view for the 'k1datastorage' account. The breadcrumb path is 'Blob containers > casestudy > raw > bronzeGold > \_delta\_log'. The authentication method is 'Access key'. A search bar is present. The table below lists the contents of the '\_delta\_log' container, showing 6 items. The items include a directory '[.]', a directory '\_\_tmp\_path...', a directory '\_staged\_co...', and several '0000000000...' files, all of which are 'Block blob' type and 'Available' in lease state.

Name	Last modified	Access tier	Blob type	Size	Lease state
[.]					
__tmp_path...	09/08/2025, 16:35:29				
_staged_co...	09/08/2025, 16:29:41				
0000000000...	09/08/2025, 16:29:43	Hot (Inferred)	Block blob	2.65 KiB	Available
0000000000...	09/08/2025, 16:29:42	Hot (Inferred)	Block blob	1.65 KiB	Available
0000000000...	09/08/2025, 16:35:29	Hot (Inferred)	Block blob	2.65 KiB	Available
0000000000...	09/08/2025, 16:35:28	Hot (Inferred)	Block blob	1.73 KiB	Available