

Case Study 4 – Dynamic Ingestion of Tables using Metadata Control in Azure Data Factory

by Sudarshan Zunja

Setting Configurations and Environments

Source Dataset – SQL Server Database Tables

Target Sink – Azure Blob Storage Container

Transformation Intermediaries – Azure Databricks

Microsoft Azure | Data Factory | notafactory

Factory Resources

- Pipelines: 1
 - pipeline1
- Change Data Capture (preview): 0
- Datasets: 3
 - DS_SinkStorage
 - DS_SQLServer
 - DS_SQLServerParam
- Data flows: 1
 - dataflow1
- Power Query: 0

SQL Server
DS_SQLServerParam

Connection Schema Parameters

Linked service * LS_SQLServerParam Test connection Edit + New Learn more

Table dbo/metadatacontrol2 Refresh Preview data

☐ Enter manually

Microsoft Azure | Data Factory | notafactory

Factory Resources

- Pipelines: 1
 - pipeline1
- Change Data Capture (preview): 0
- Datasets: 3
 - DS_SinkStorage
 - DS_SQLServer
 - DS_SQLServerParam
- Data flows: 1
 - dataflow1
- Power Query: 0

SQL Server
DS_SQLServer

Connection Schema Parameters

Linked service * LS_SQLServerParam Test connection Edit + New Learn more

Table dbo/metadatacontrol2 Refresh Preview data

☐ Enter manually

Microsoft Azure | Data Factory | notafactory

Search factory and documentation

TrainingUser7@dgmtech.cloud
DGM TECHNOLOGIES PRIVATE LIMITED

Preview experience: Off

Factory Resources

- Pipelines: 1
 - pipeline1
- Change Data Capture (preview): 0
- Datasets: 3
 - DS_SinkStorage
 - DS_SQLServer
 - DS_SQLServerParam
- Data flows: 1
 - dataflow1
- Power Query: 0

DelimitedText DS_SinkStorage

Connection Schema Parameters

Linked service: LS_BlobStorage [Test connection](#) [Edit](#) [New](#) [Learn more](#)

File path: rawdata / @dataset().pSinkPath / File name [Browse](#) [Preview data](#)

Compression type: No compression

Column delimiter: Comma (,)

Row delimiter: Default (\r\n, or \n)

Encoding: Default(UTF-8)

Quote character: Double quote (")

Escape character: Backslash (\)

First row as header: ☐

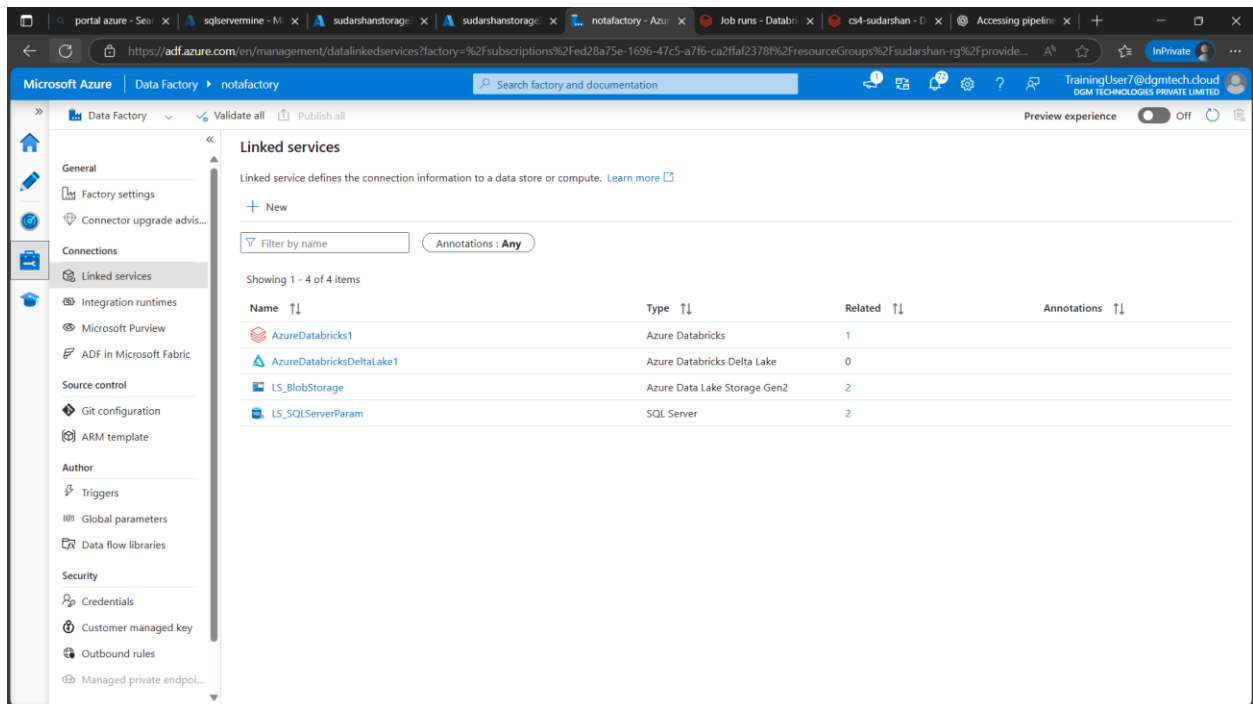
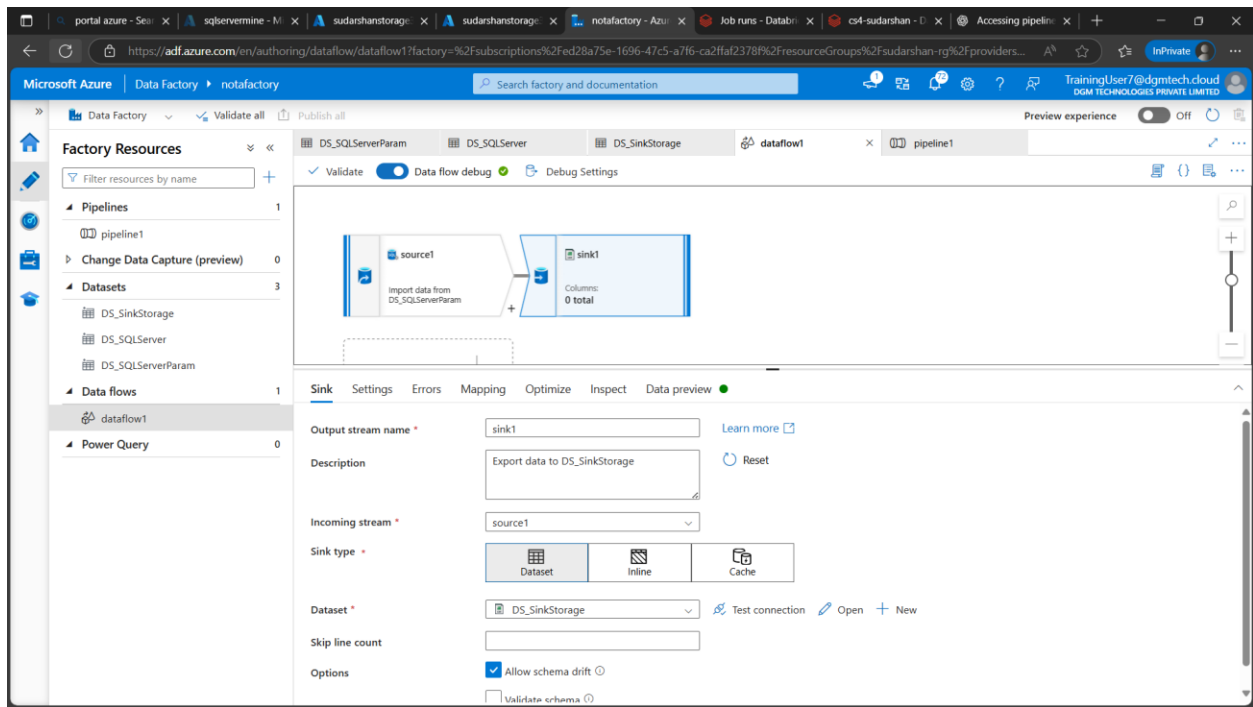
Data Flow Structure

This screenshot shows the initial setup of a data flow in the Microsoft Azure Data Factory portal. The left-hand 'Factory Resources' pane lists the available components: Pipelines (1), Change Data Capture (0), Datasets (3), Data flows (1), and Power Query (0). The main canvas displays a data flow named 'dataflow1' with two components: 'source1' (Import data from DS_SQLServerParam) and 'sink1' (Export data to DS_SinkStorage). Below the canvas, the 'Parameters' tab is active, showing a table of parameters.

Name	Type	Default value
pSourceQuery	string	Enter expression...
pSinkPath	string	Enter expression...

This screenshot shows the 'Source settings' configuration for the 'source1' component of the data flow. The 'Output stream name' is set to 'source1'. The 'Description' is 'Import data from DS_SQLServerParam'. The 'Source type' is set to 'Dataset'. The 'Dataset' is selected as 'DS_SQLServerParam'. The 'Options' section includes 'Allow schema drift' (checked), 'Infer drifted column types' (unchecked), 'Validate schema' (unchecked), and 'Sampling' (set to 'Disable').

Property	Value
Output stream name	source1
Description	Import data from DS_SQLServerParam
Source type	Dataset
Dataset	DS_SQLServerParam
Options	<ul style="list-style-type: none">Allow schema drift: <input checked="" type="checkbox"/>Infer drifted column types: <input type="checkbox"/>Validate schema: <input type="checkbox"/>Sampling: <input type="radio"/> Enable <input checked="" type="radio"/> Disable



Pipeline Structure and Run

Microsoft Azure | Data Factory | notafactory

All pipeline runs > pipeline1 - Activity runs

Lookup1

ForEach1

Notebook1

Activity runs

Pipeline run ID 331d8b30-5839-4494-90de-fd0ec75d99b6

All status List

Showing 1 - 5 items

Activity name	Activity st...	Activit...	Run start	Duration	Integration runtime	User prop...	Activity run ID
Lookup1	Succeeded	Lookup	8/8/2025, 5:19:58 PM	11s	AutoResolveIntegrationRuntime (West US 2)		cbe6179f-5215-4b6c-8d94-52943c13-ccce-41c
ForEach1	Succeeded	ForEach	8/8/2025, 5:20:10 PM	4m 2s			e7388716-b1a6-44da6c7b33-6433-4fc
Data flow1	Succeeded	Data flow	8/8/2025, 5:20:11 PM	3m 37s	AutoResolveIntegrationRuntime (West US 2)		da6c7b33-6433-4fc
Data flow1	Succeeded	Data flow	8/8/2025, 5:23:49 PM	21s	AutoResolveIntegrationRuntime (West US 2)		b2696a6d-db44-46
Notebook1	Succeeded	Notebook	8/8/2025, 5:24:12 PM	37s	AutoResolveIntegrationRuntime (West US 2)		

Databricks Transformations and Jobs Runs

Microsoft Azure databricks

Jobs & Pipelines

Create new

- Ingestion pipeline
- ETL pipeline
- Job

Jobs & pipelines Job runs

Job Run Start: 08/06/2025 05:30 PM End: 08/08/2025 05:30 PM Run status Error code

Top 5 error codes (2 errors)

RunExecutionError 2

Start time	Job	Run as	Launched	Duration	Status	Error code	Run parameters
Aug 08, 2025, 05:24 PM	ADF_notafactory_pipeline...	TrainingUser7	By runs submit API	16s	Succeeded		
Aug 08, 2025, 05:17 PM	ADF_notafactory_pipeline...	TrainingUser7	By runs submit API	12s	Succeeded		
Aug 08, 2025, 05:12 PM	ADF_notafactory_pipeline...	TrainingUser7	By runs submit API	15s	Succeeded		
Aug 08, 2025, 05:10 PM	ADF_notafactory_pipeline...	TrainingUser7	By runs submit API	19s	Failed	RunExecutionError	

Microsoft Azure databricks

Runs

ADF_notafactory_pipeline1_Notebook1_b2696a6d-db44-4636-8083-cc22c3eb92a2 run

Output

```
return df

pii_columns = [
    1: "hash",
    2: "hash",
    3: "redact",
    4: "redact"
]

masked_customers_df = mask_pii(cust_silver, pii_columns)
masked_customers_df.show(5)

gold_path = f"{delta_path}Gold"

masked_customers_df.write.mode("overwrite").format("delta").save(gold_path)
```

(2) Spark Jobs

- cust_silver: pyspark.sql.dataframe.DataFrame = [c0: string, c1: string ... 3 more fields]
- masked_customers_df: pyspark.sql.dataframe.DataFrame = [c0: string, c1: string ... 3 more fields]
- pol_silver: pyspark.sql.dataframe.DataFrame = [c0: string, c1: string ... 3 more fields]

Task run

Details

Job ID 1049332639346510

Task run ID 872422149560001

Run as TrainingUser7

Started Aug 08, 2025, 05:24 PM

Ended Aug 08, 2025, 05:24 PM

Duration 16s

Queue duration

Status Succeeded

Notebook

/Users/traininguser7@dgmtech.cloud/cs4-sudarshan

Final Medallion Layers (Bronze-Silver-Gold)

portal azure - Se...sqlservmine - M...sudarshanstorage...sudarshanstorage...notafactory - Azu...ADF_notafactory...cs4-sudarshan - ...Accessing pipeline...

https://portal.azure.com/#@dgmttech.cloud/resource/subscriptions/ed28a75e-1696-47c5-a716-ca2faf2378f/resourceGroups/sudarshan-rg/providers/Microsoft.Storage/storageA...InPrivate

Microsoft AzureSearch resources, services, and docs (G+?)CopilotTrainingUser7@dgmttec...DGM TECHNOLOGIES PRIVATE U...

Home > sudarshanstorage3

Storage account

Overview

Activity log

Tags

Diagnose and solve problems

Access Control (IAM)

Data migration

Events

Storage browser

Partner solutions

Resource visualizer

Data storage

Security + networking

Networking

Access keys

Shared access signature

Encryption

Microsoft Defender for Cloud

Data management

sudarshanstorage3

Favorites

Recently viewed

Blob containers

Logs

rawdata

View all

File shares

Queues

Tables

Enhance the security of this storage account

Does this storage account follow security best practices +1

Storage browser

+

 Add Directory

↑

 Upload

↻

 Refresh

🗑

 Delete

📄

 Copy

📄

🔄

🔒

🔓

...

Blob containers > rawdata > rawdata

Authentication method: Access key [\(Switch to Microsoft Entra user account\)](#)

🔍

 Search blobs by prefix (case-sensitive)

Only show active objects

Showing all 3 items

Add or remove favorites by pressing Ctrl+Shift+F