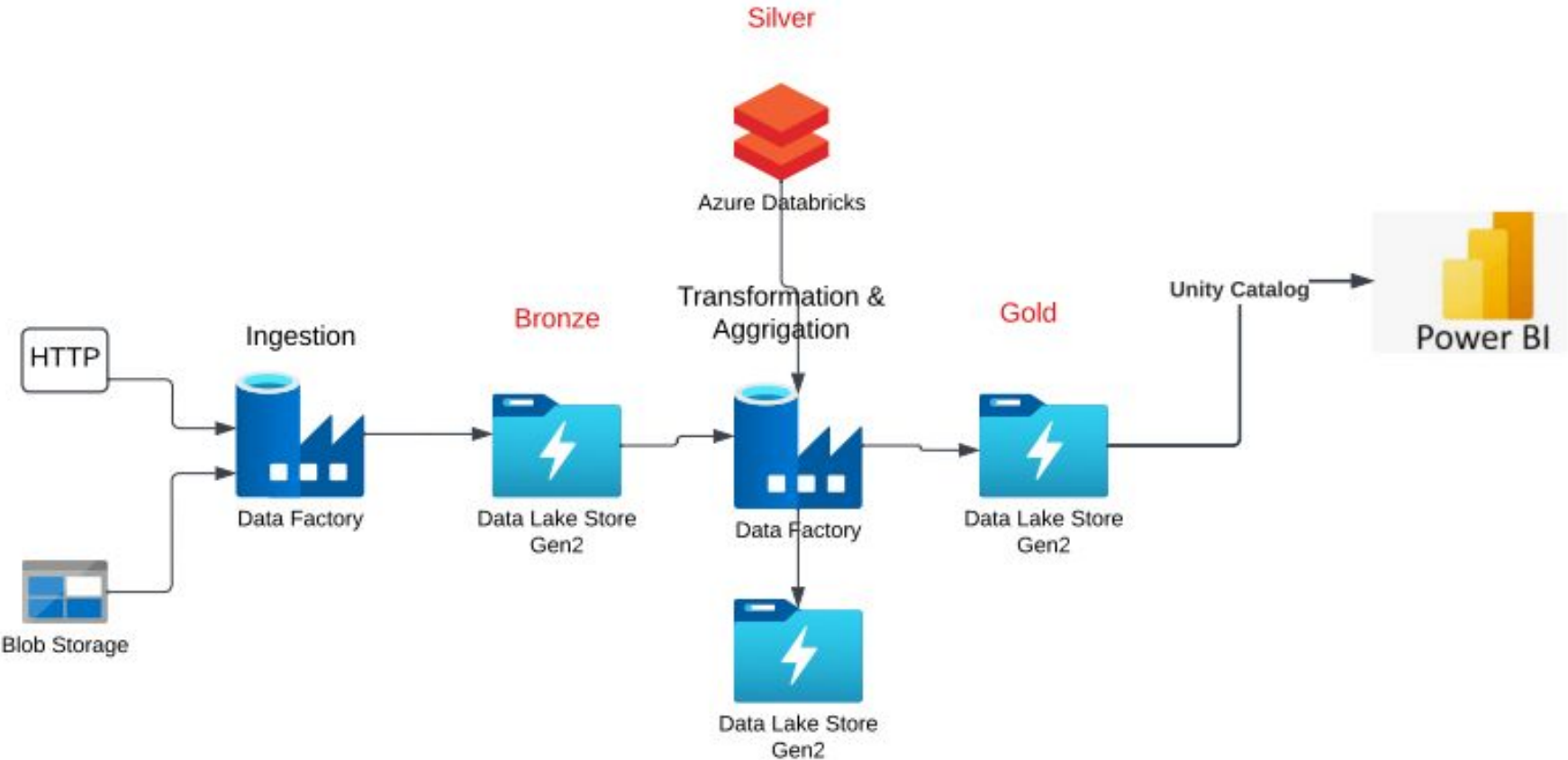


# Solution Architecture



### Ingestion Layer:

- Data is ingested from sources like **HTTP APIs** and **Blob Storage**.
- **Azure Data Factory (ADF)** is used for orchestrating the movement of data into the Data Lake Gen2 (Bronze Layer).

### Bronze Layer:

- Raw data is stored in **Azure Data Lake Storage (ADLS) Gen2**.

### Silver / Transformation Layer:

- Data is processed and transformed using **Azure Databricks**.
- Intermediate or transformed data (Silver Layer) is written back to ADLS.


### Gold Layer:


- The cleaned, aggregated, and analytics-ready data is stored in **ADLS Gen2**
- **Unity Catalog** in Databricks is referenced for governance, access control, and data discovery.


### Consumption Layer:


- Analytics and reporting tools such as **Power BI** consume the Gold Layer for insights and visualizations.


# Link Services:


 1

 Factory settings


 Connections


 **Linked services**

 Integration runtimes


 Microsoft Purview


Source control

 Git configuration

 ARM template

Author

 Triggers

 Global parameters





Linked service defines the connection information to a data store or compute. [Learn more](#)

+ New

Filter by name

Annotations : Any

Showing 1 - 4 of 4 items

Name ↑↓	Type ↑↓	Relationship
 <a href="#">ls_AzureBlobStorage</a>	Azure Blob Storage	1
 <a href="#">ls_AzureDatabricks</a>	Azure Databricks	0
 <a href="#">ls_AzureDataLakeStorage</a>	Azure Data Lake Storage Gen2	0
 <a href="#">ls_HttpServer</a>	HTTP	0



Parameters Variables Settings **Output**

**Pipeline run ID:** d1d8792d-97fa-4740-9c6a-6e9998d172be

**Pipeline status** ✔ Succeeded

All status ▼

[Monitor](#)

Showing 1 - 4 of 4 items

Activity name <span style="font-size: 0.8em;">↑↓</span>	Activity status <span style="font-size: 0.8em;">↑↓</span>	Activity type <span style="font-size: 0.8em;">↑↓</span>	Run start <span style="font-size: 0.8em;">↑↓</span>	Duration <span style="font-size: 0.8em;">↑↓</span>	Integration runtime	User property
Delete population from blob	<span style="color: green;">✔</span> Succeeded	Delete	12/16/2024, 4:07:48 PM	4s	AutoResolveIntegrator	
Copy population tsv staging i...	<span style="color: green;">✔</span> Succeeded	Copy data	12/16/2024, 4:07:34 PM	13s	AutoResolveIntegrator	
Copy population to bronze	<span style="color: green;">✔</span> Succeeded	Copy data	12/16/2024, 4:07:21 PM	13s	AutoResolveIntegrator	
Blob Validation of population...	<span style="color: green;">✔</span> Succeeded	Validation	12/16/2024, 4:07:16 PM	4s		

- Validating the file in blob (population)
- File extracting from tsv.gz to csv file format.
- Moving file from blob (population) to blob (population-staging-data)
- Deleting the file from blob (population)

## Factory Resources

Filter resources by name

### Pipelines

pl\_ingest\_http\_files

pl\_ingestion\_population\_age

### Change Data Capture (preview)

### Datasets

#### http\_data

ds\_http\_file\_list

ds\_http\_file\_load

ds\_http\_file\_read

#### population\_data

ds\_population\_csv\_unzip

ds\_population\_tsv\_gz

ds\_population\_tsv\_staging

### Data flows

### Power Query

pl\_ingestion\_popul... ds\_http\_file\_list ds\_http\_file\_read ds\_http\_file\_load pl\_ingest\_http\_files

Validate Debug Add trigger

Lookup

Lookup http json

ForEach

ForEach http file

Activities

Copy http file

Parameters Variables Settings **Output**

**Pipeline run ID:** 3414acbc-91dd-44c2-a5a8-6719bf88ed9c

**Pipeline status** ✓ Succeeded [View debug run](#)

All status ▼ List ▼ [Monitor in Azure Metrics](#) [Export](#)


Showing 1 - 6 of 6 items

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User prop
Copy http file	✓ Succeeded	Copy data	12/18/2024, 9:26:58 PM	23s	AutoResolveIntegration	
Copy http file	✓ Succeeded	Copy data	12/18/2024, 9:26:58 PM	12s	AutoResolveIntegration	
Copy http file	✓ Succeeded	Copy data	12/18/2024, 9:26:58 PM	13s	AutoResolveIntegration	
Copy http file	✓ Succeeded	Copy data	12/18/2024, 9:26:58 PM	14s	AutoResolveIntegration	
ForEach http file	✓ Succeeded	ForEach	12/18/2024, 9:26:57 PM	26s		
Lookup http json	✓ Succeeded	Lookup	12/18/2024, 9:25:59 PM	58s	AutoResolveIntegration	

# Databricks - unity catalog setup

1. Create : Access Connector for Azure Databricks
2. Gen2 : Storage blob data contributor add role assignment.
3. Databricks : Create metastore and assign workspace
4. In workspace - catalog - Add external storage to access the data from ADLS Gen2

Microsoft Azure

 databricks

CTRL + P

project1

+ New

Workspace

Recents

Catalog

Workflows

Compute

SQL

SQL Editor

Queries

Dashboards

Genie

Catalog Explorer >







External Data

External Locations ⓘ

Credentials ⓘ

Connections ⓘ

3 locations

Name	Credential	URL	Owner
 uc_bronze_ext_loc	0d18efd9-5d29-41...	abfss://bronze@project1adls.dfs.core.windows.net/	 com#ext#@...
 uc_gold_ext_loc	0d18efd9-5d29-41...	abfss://gold@project1adls.dfs.core.windows.net/	 m#ext#@...at...
 uc_silver_ext_loc	0d18efd9-5d29-41...	abfss://silver@project1adls.dfs.core.windows.net/	 #ext#@...t...

Filter resources by name

Pipelines3

pLpopulatin\_transformation

pLingest\_http\_files

pLingestion\_population\_age

Change Data Capture (preview)0

Datasets6

http\_data3

population\_data3

ds\_population\_csv\_unzip

ds\_population\_tsv\_gz

ds\_population\_tsv\_staging

Data flows0

Power Query0

»

✓ Validate

▶ Debug

⚡ Add trigger

Notebook

Population transformation

Notebook

Population aggriqation

Parameters

Variables

Settings

Output

Pipeline run ID: 6b3e51af-477a-4f93-ab01-f348582e2895

Pipeline status 

✓

 Succeeded

All status 

▼

Showing 1 - 2 of 2 items

Activity name	Activity status	Activity type	Run start	Duration	Int
Population aggriqation	<div>✓</div> Succeeded	Notebook	12/19/2024, 12:53:37 A	19s	Au
Population transformation	<div>✓</div> Succeeded	Notebook	12/19/2024, 12:53:17 A	19s	Au

Filter resources by name

Pipelines 4

p\_Lexecute

p\_Lingest\_http\_files

p\_Lingestion\_population\_age

p\_Lpopulatin\_transformation

Change Data Capture (preview) 0

Datasets 6

http\_data 3

population\_data 3

ds\_population\_csv\_unzip

ds\_population\_tsv\_gz

ds\_population\_tsv\_staging

Data flows 0

Power Query 0

» Validate Debug Add trigger



Parameters Variables Settings **Output**

Pipeline run ID: 55a31594-b6e3-4acf-887d-779817fa6ba9

Pipeline status **Succeeded**

All status

[Monitor](#)

Showing 1 - 3 of 3 items

Activity name	Activity status	Activity type	Run start	Duration	Integ
Execute Pipeline3	<b>Succeeded</b>	Execute Pipeline	12/19/2024, 12:59:23 A	1m 12s	
Execute Pipeline2	<b>Succeeded</b>	Execute Pipeline	12/19/2024, 12:58:46 A	38s	
Execute Pipeline1	<b>Succeeded</b>	Execute Pipeline	12/19/2024, 12:58:09 A	36s	







## Factory Resources

Filter resources by name

### Pipelines 3

pl\_population\_transformation

pl\_http\_data\_ingestion

pl\_population\_data\_ingestion

### Change Data Capture (preview) 0

### Datasets 6

http\_data 3

population 3

ds\_population\_csv

ds\_population\_tsv\_zip

ds\_population\_staging\_file

### Data flows 0

### Power Query 0

## Activities

Search activities

Move and transform

Synapse

Azure Data Explorer

Azure Function

Batch Service

Databricks

Notebook

Jar

Python

Data Lake Analytics

General

HDInsight

Iteration & conditionals

Machine Learning

Validate Cancel options Add trigger

Notebook



population transformation

Parameters Variables Settings Output

Pipeline run ID: bc1ede82-6515-4dc3-a7e2-1d42a98d9759

All status

Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start
---------------	-----------------	---------------	-----------