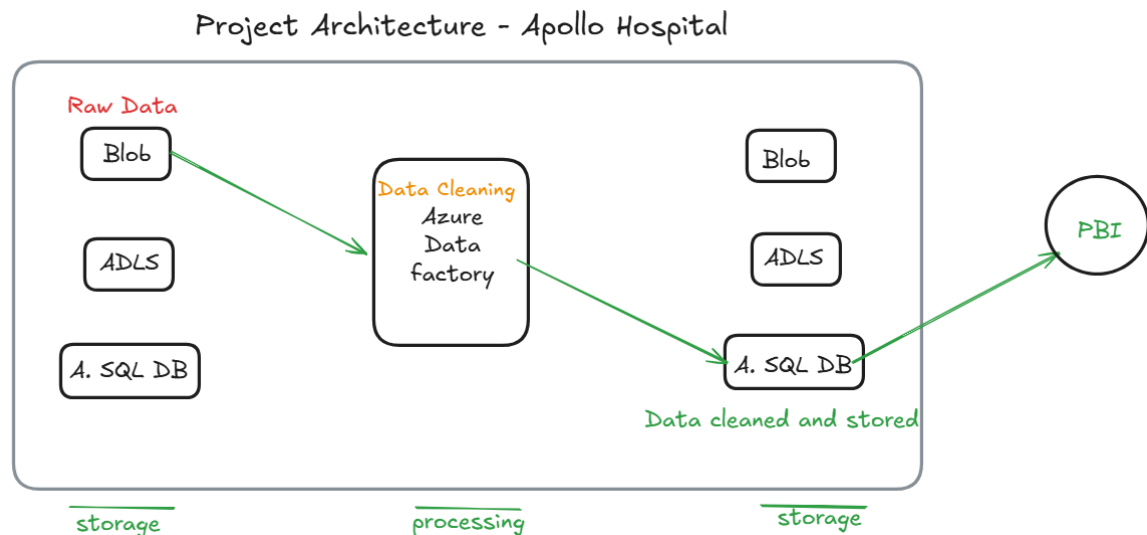


# PROJECT – APOLLO

By Shanti Jogi

## 1. Architecture



## 2. Resources

- Blob storage
- ADLS
- Azure SQL DB
- ADF

### Blob storage Created:

The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes the Microsoft Azure logo, an 'Upgrade' button, a search bar, and various icons for Copilot, notifications, and settings. The main content area displays the 'Overview' page for a storage account named 'blobapollo'. The account is located in the 'eastus' region and is part of the 'ApolloHG' resource group. The 'Properties' section shows the account kind as 'StorageV2 (general purpose v2)' and the provisioning state as 'Succeeded'. The 'Blob service' section lists various features and their status, such as 'Hierarchical namespace' (Disabled), 'Default access tier' (Hot), and 'Blob soft delete' (Enabled (7 days)). The 'Security' section shows that 'Require secure transfer for REST API operations' is 'Enabled' and 'Infrastructure encryption' is 'Disabled'. The 'Networking' section shows that 'Allow access from' is 'All networks' and 'Access for trusted Microsoft services' is 'Yes'.

## ADLS Created:

The screenshot shows the Microsoft Azure portal interface for the 'adlsapollo' storage account. The left sidebar contains navigation options like Overview, Activity log, Tags, and Storage browser. The main content area displays 'Storage account metrics' with four cards: Blob containers, File shares, Tables, and Queues, each showing counts and total data stored. Below these is a 'Recently viewed' section and a link to 'Other ways to manage data'.

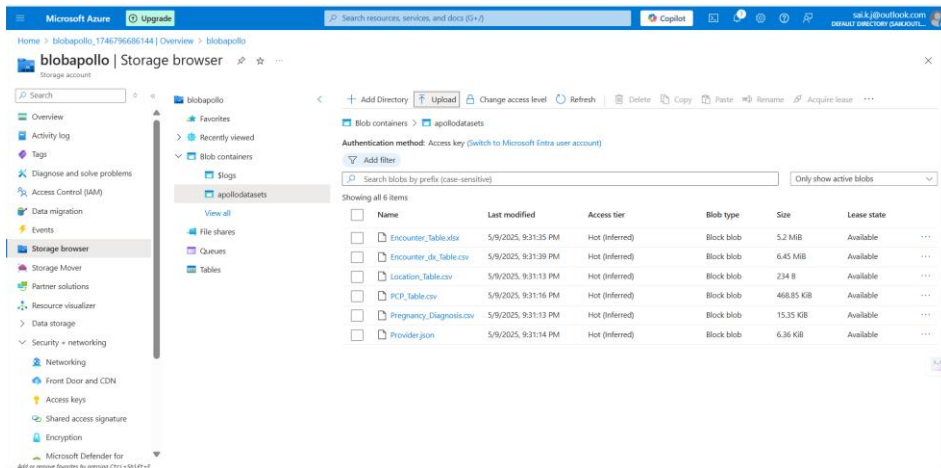
## Azure SQL DB created:

The screenshot shows the Microsoft Azure portal interface for the 'apolldb' database. The left sidebar contains navigation options like Overview, Activity log, Tags, and Query editor (preview). The main content area displays a 'Welcome to SQL Database Query Editor' message with authentication options: SQL server authentication (Login, Password) and Microsoft Entra authentication (Continue as sai.k.j@outlook.com).

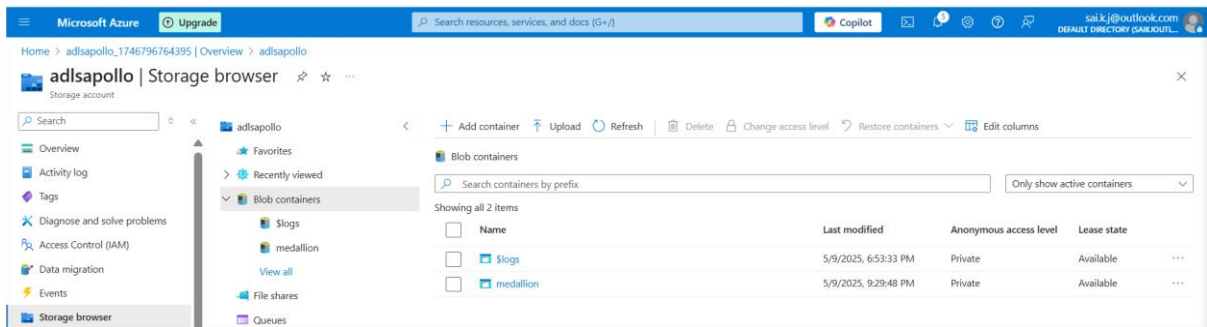
## ADF Created:

The screenshot shows the Microsoft Azure portal interface for the 'apolloadf09' data factory. The left sidebar contains navigation options like Overview, Activity log, Tags, and Data Factory Studio. The main content area displays the 'Azure Data Factory Studio' logo and a 'Launch studio' button. Below this are four cards: Quick Starts, Tutorials, Template Gallery, and Training Modules.

## Raw Data in blob

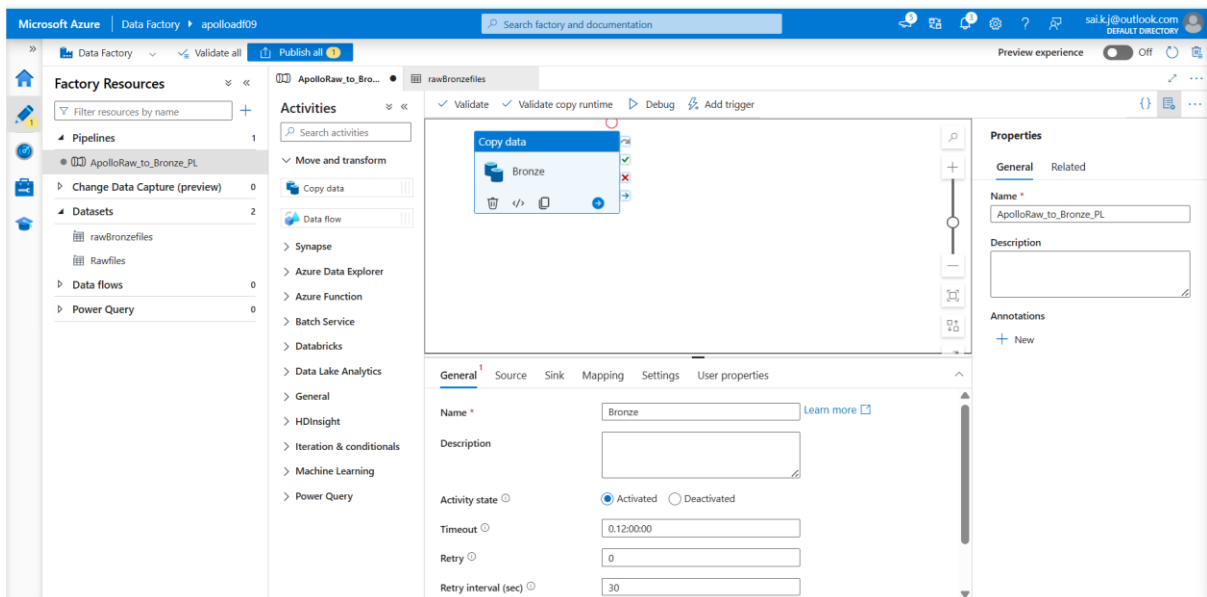


## For Medallion architecture, created a directory in ADLS



## Lets create Pipelines following medallion architecture

### Bronze Pipeline created:



## Trigger the pipeline

Microsoft Azure | Data Factory | apolloadf09

Search factory and documentation

saik.j@outlook.com

Run Succeeded  
Successfully ran ApolloRaw\_to\_Bronze\_PL (Pipeline).  
View pipeline run

All pipeline runs > ApolloRaw\_to\_Bronze\_PL - Activity runs

Copy data

Bronze

Activity runs

Pipeline run ID 23a62f6f-5c3d-4a7d-990c-478acbd55fd5

All status

Monitor in Azure Metrics | Export to CSV

Showing 1 - 1 items

Activity name	Activity status	Activity name	Run start	Duration	Integration runtime	User properties	Activity run ID
Bronze	Succeeded	Copy data	5/9/2025, 9:54:15 PM	12s	AutoResolveIntegrationRuntime (East US)		7e4e0f65-d65c-4d8d-9717-560

## Silver Pipeline Created:

Microsoft Azure | Data Factory | apolloadf09

Search factory and documentation

saik.j@outlook.com

Factory Resources

Pipelines

- ApolloRaw\_to\_Bronze\_PL
- Bronze\_to\_Silver\_PL

Change Data Capture (preview)

Datasets

- Provider\_DF
- Encounter\_DF
- EncounterDX\_DF
- Location\_DF
- Patient\_DF
- PCP\_DF
- PregDiag\_DF

Data flows

- Encounter
- EncounterDX
- Location
- Patient
- PCP
- PregDiag
- Provider

Activities

Search activities

Move and transform

Copy data

Data flow

Synapse

Azure Data Explorer

Azure Function

Batch Service

Databricks

Data Lake Analytics

General

HDInsight

Iteration & conditionals

Machine Learning

Power Query

Data flow debug

Parameters

Variables

Settings

Output

+ New

Lets create Gold pipeline:  
for that we need to setup a final dataflow

Microsoft Azure | Data Factory | apolloadf09

Search factory and documentation

saik.j@outlook.com

Factory Resources

Pipelines

- ApolloRaw\_to\_Bronze\_PL
- Bronze\_to\_Silver\_PL
- Apollo\_PL
- Silver\_to\_Gold\_PL

Change Data Capture (preview)

Datasets

- Encounter\_DF
- EncounterDX\_DF
- Location\_DF
- Patient\_DF
- PCP\_DF
- PregDiag\_DF
- Provider\_DF

Data flows

- Apollo\_DF
- Encounter\_DF
- EncounterDX\_DF
- Location\_DF
- Patient\_DF
- PCP\_DF
- PregDiag\_DF
- Provider\_DF

Power Query

18 Columns

Encounter

EncounterDX

PregDiag

Patient

PCP

Provider

location

join1

select1

join2

select2

join3

select3

join4

join5

select4

join6

select2

## Gold Pipeline created

The screenshot shows the Microsoft Azure Data Factory portal for the 'apoloadf09' workspace. The 'Factory Resources' pane on the left lists various pipelines, including 'Silver\_to\_Gold\_PL'. The 'Activities' pane shows a search bar and a list of activity types. The main canvas displays a 'Data flow' activity named 'Apollo'. The 'General' tab is selected, showing the 'Name' as 'Apollo' and the 'Activity state' as 'Activated'. A notification at the top right indicates 'Publishing completed Successfully published'.

## Now setup Master pipeline to execute all pipelines

The screenshot shows the Microsoft Azure Data Factory portal for the 'apoloadf09' workspace. The 'Factory Resources' pane on the left lists various pipelines, including 'Apollo\_PL'. The 'Activities' pane shows a search bar and a list of activity types. The main canvas displays a sequence of 'Execute Pipeline' activities. The first activity is 'Bronze ApolloRaw\_to\_Bronz...', followed by 'Silver Bronze\_to\_Silver\_PL', and finally 'Gold Silver\_to\_Gold\_PL'. The 'General' tab is selected, showing the 'Invoked pipeline' as 'Silver\_to\_Gold\_PL' and the 'Wait on completion' checkbox checked.

## Trigger the pipeline and check in gold →ADLS

## As per the BRD followed medallion architecture, Files saved as is in bronze layer

Microsoft Azure | Upgrade | Search resources, services, and docs (G+)

Home > adlsapollo

adlsapollo | Storage browser

Search

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

adlsapollo

Favorites  
Recently viewed  
Blob containers  
\$logs  
medallion  
View all  
File shares  
Queues  
Tables

+ Add Directory | Upload | Refresh | Delete | Copy | Paste | Rename | Acquire lease | Break lease | Edit columns

Blob containers > medallion > Bronze

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

Search blobs by prefix (case-sensitive)

Only show active objects

Showing all 7 items

<input type="checkbox"/>	Name	Last modified	Access tier	Blob type	Size	Lease state
<input type="checkbox"/>	[-]					...
<input type="checkbox"/>	Encounter_Table.xlsx	5/10/2025, 12:17:22 PM	Hot (Inferred)	Block blob	5.2 MiB	Available
<input type="checkbox"/>	Encounter_dx_Table.csv	5/10/2025, 12:17:22 PM	Hot (Inferred)	Block blob	6.45 MiB	Available
<input type="checkbox"/>	Location_Table.csv	5/10/2025, 12:17:22 PM	Hot (Inferred)	Block blob	234 B	Available
<input type="checkbox"/>	PCP_Table.csv	5/10/2025, 12:17:22 PM	Hot (Inferred)	Block blob	468.85 KiB	Available
<input type="checkbox"/>	Patient_Table.csv	5/10/2025, 12:17:22 PM	Hot (Inferred)	Block blob	1.18 MiB	Available
<input type="checkbox"/>	Pregnancy_Diagnosis.csv	5/10/2025, 12:17:22 PM	Hot (Inferred)	Block blob	15.35 KiB	Available
<input type="checkbox"/>	Provider.json	5/10/2025, 12:17:22 PM	Hot (Inferred)	Block blob	6.36 KiB	Available

## after transformations files moved to silver layer in one format

Microsoft Azure | Upgrade | Search resources, services, and docs (G+)

Home > adlsapollo

adlsapollo | Storage browser

Search

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

adlsapollo

Favorites  
Recently viewed  
Blob containers  
\$logs  
medallion  
View all  
File shares  
Queues  
Tables

+ Add Directory | Upload | Refresh | Delete | Copy | Paste | Rename | Acquire lease | Break lease | Edit columns

Blob containers > medallion > Silver

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

Search blobs by prefix (case-sensitive)

Only show active objects

Showing all 7 items

<input type="checkbox"/>	Name	Last modified	Access tier	Blob type	Size	Lease state
<input type="checkbox"/>	[-]					...
<input type="checkbox"/>	Encounter.csv	5/10/2025, 12:20:53 PM	Hot (Inferred)	Block blob	5.3 MiB	Available
<input type="checkbox"/>	EncounterDX.csv	5/10/2025, 12:20:59 PM	Hot (Inferred)	Block blob	6.07 MiB	Available
<input type="checkbox"/>	Location.csv	5/10/2025, 12:20:38 PM	Hot (Inferred)	Block blob	175 B	Available
<input type="checkbox"/>	PCP.CSV	5/10/2025, 12:20:40 PM	Hot (Inferred)	Block blob	420.02 KiB	Available
<input type="checkbox"/>	Patient.csv	5/10/2025, 12:21:04 PM	Hot (Inferred)	Block blob	1.13 MiB	Available
<input type="checkbox"/>	PregDiag.csv	5/10/2025, 12:21:32 PM	Hot (Inferred)	Block blob	13.7 KiB	Available
<input type="checkbox"/>	Provider.csv	5/10/2025, 12:21:41 PM	Hot (Inferred)	Block blob	1.69 KiB	Available

## And final analytical file for biz transformations moved in Gold Layer

Microsoft Azure | Upgrade | Search resources, services, and docs (G+)

Home > adlsapollo

adlsapollo | Storage browser

Search

Overview  
Activity log  
Tags  
Diagnose and solve problems  
Access Control (IAM)  
Data migration  
Events  
Storage browser  
Partner solutions  
Resource visualizer  
Data storage

adlsapollo

Favorites  
Recently viewed  
Blob containers  
\$logs  
medallion  
View all  
File shares  
Queues  
Tables

+ Add Directory | Upload | Refresh | Delete | Copy | Paste | Rename | Acquire lease | Break lease | Edit columns

Blob containers > medallion > Gold

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

Search blobs by prefix (case-sensitive)

Only show active objects

Showing all 1 items

<input type="checkbox"/>	Name	Last modified	Access tier	Blob type	Size	Lease state
<input type="checkbox"/>	[-]					...
<input type="checkbox"/>	Apollo.csv	5/10/2025, 12:22:07 PM	Hot (Inferred)	Block blob	188 B	Available

## Successfully moved data to ADLS → Gold Layer for further analysis....