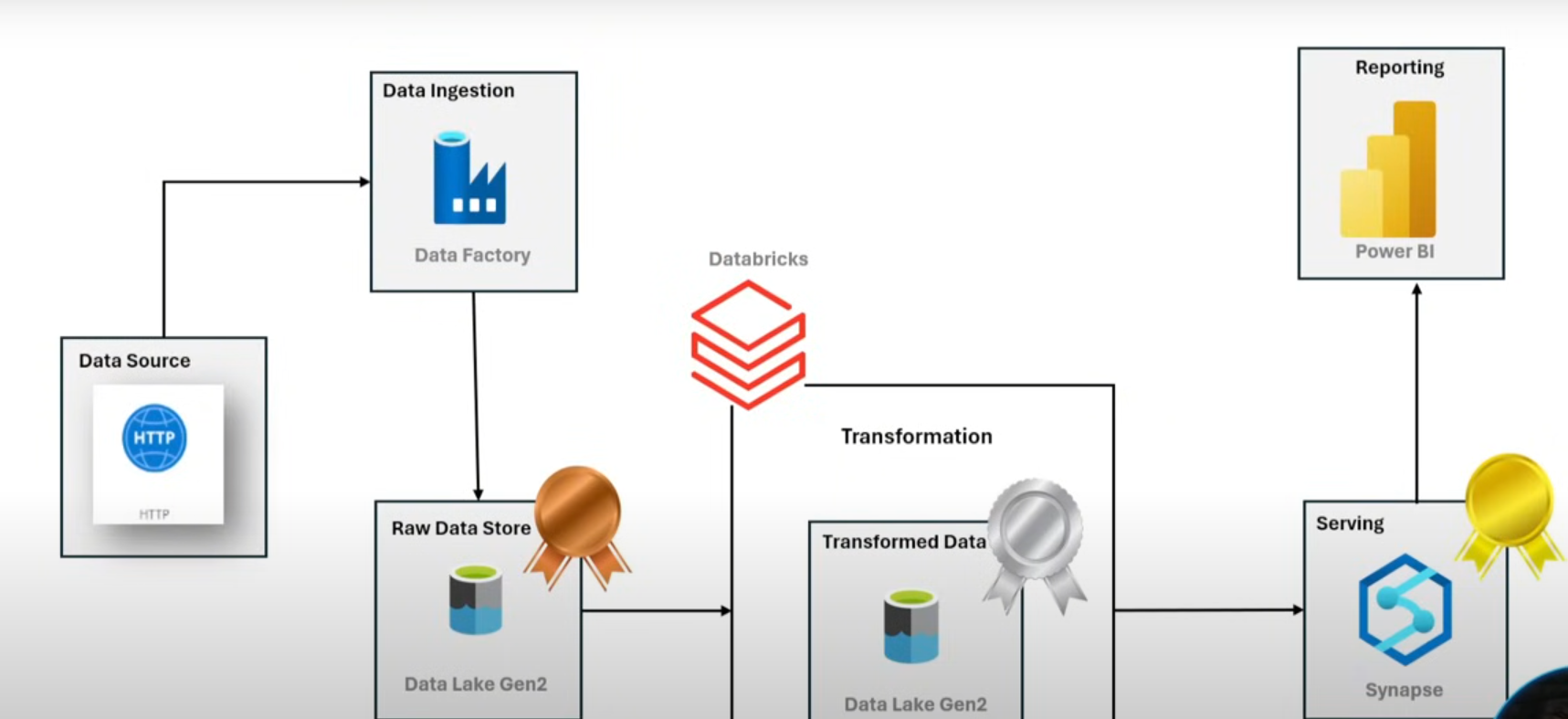
Data Architecture:



Data source is fetching from API

Azure Data Factory orchestration tool is used to pull source data and will land data to bronze layer and also does below activities:

* Build static, dynamic pipelines
* parameters
* loops
* lookup

Medallion architecture:

Will make data traverse through 3 different zones

* raw
* transform
* serving

or also calls as:

* bronze
* silver
* gold

Bronze or raw layer:

Will keep data as it is that is available in source

Silver or transform later:

Pick data from bronze layer and push data to silver layer with some transformations

Gold or serving layer:

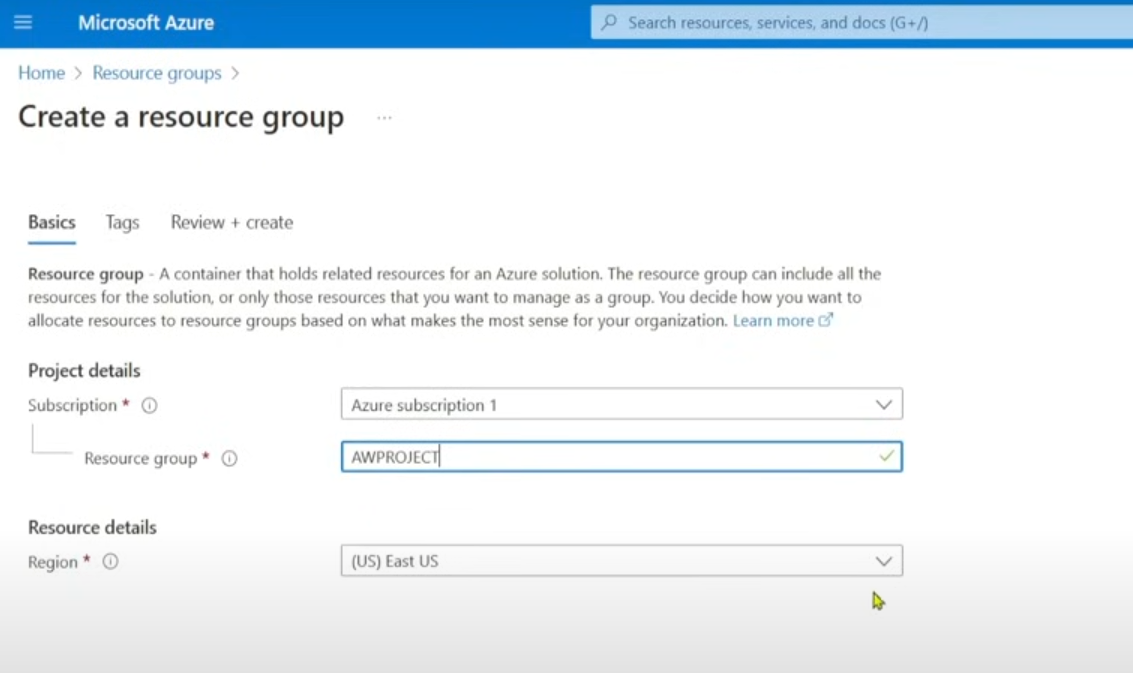
Serve data to stakeholders like : analysts, data scientists, etc

Use powerbi tool to create dashboards

Dataset link:

[Adventure Works](https://www.kaggle.com/datasets/ukveteran/adventure-works)

Creating a resource group:



To categorize any resources or resource groups, use tags. Not required for now:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Create first resource: datalake (storage account)

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Redundancy : by default, it will keep GRS (geo redundant storage) because it replicates data within different regions

But change to LRS to store data in local data centre

By default storage will create blob storage but to create datalake, tick ‘enable hierarchial namespace’:

A screenshot of a computer

Description automatically generated

Difference between blob and datalake:

Containers will be there in both blob and datalake

In blob, we cannot store hierarchy of folders which can be possible in datalake

Create a data factory:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

After home tab, author tab where pipelines will be created.

Below that, monitor tab is there, used to monitor the pipelines like: status, if it fails what is the reason

Managed tab is below monitor. It manages repositories, github connections, devops connections, linked services

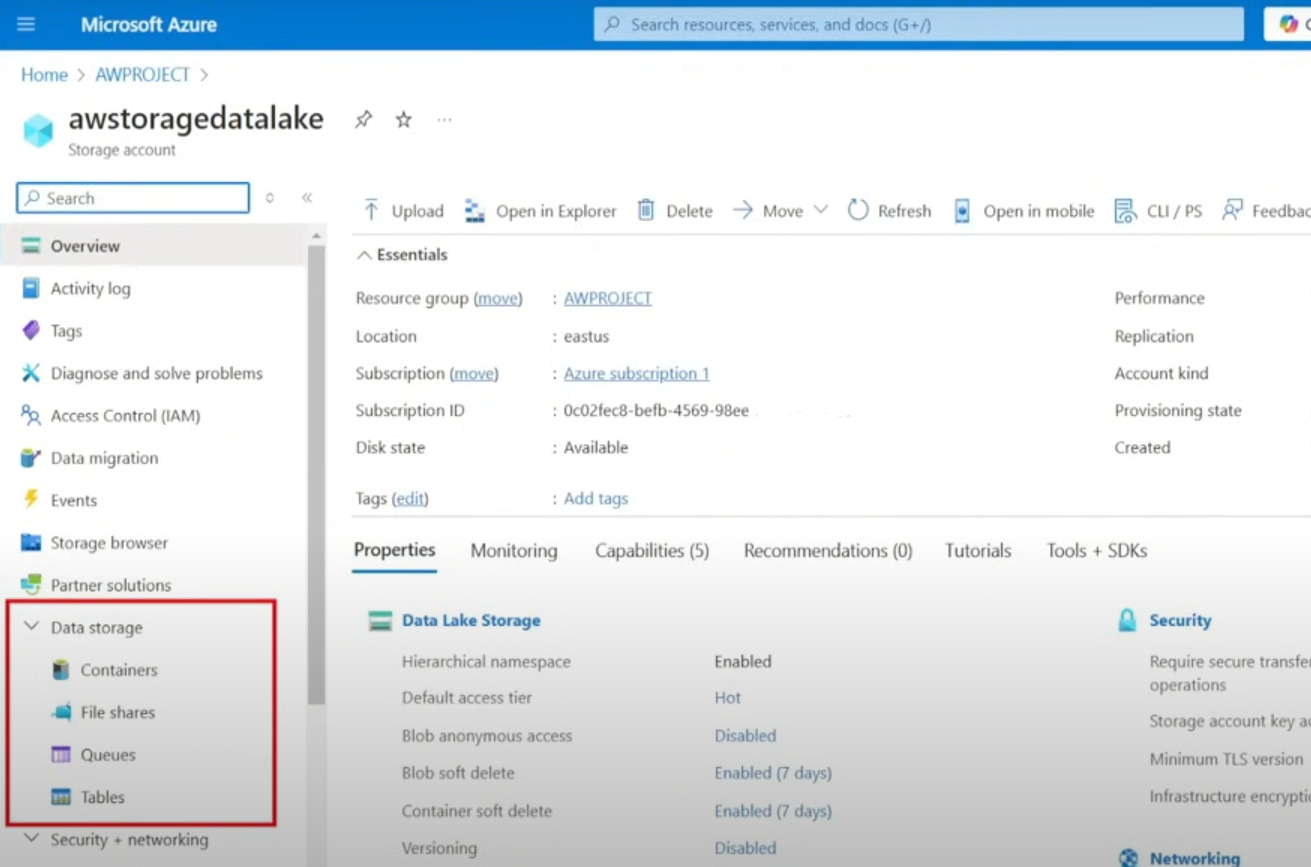
Learning resources is the last tab where resources, documentations, take some datasets available.

A screenshot of a computer

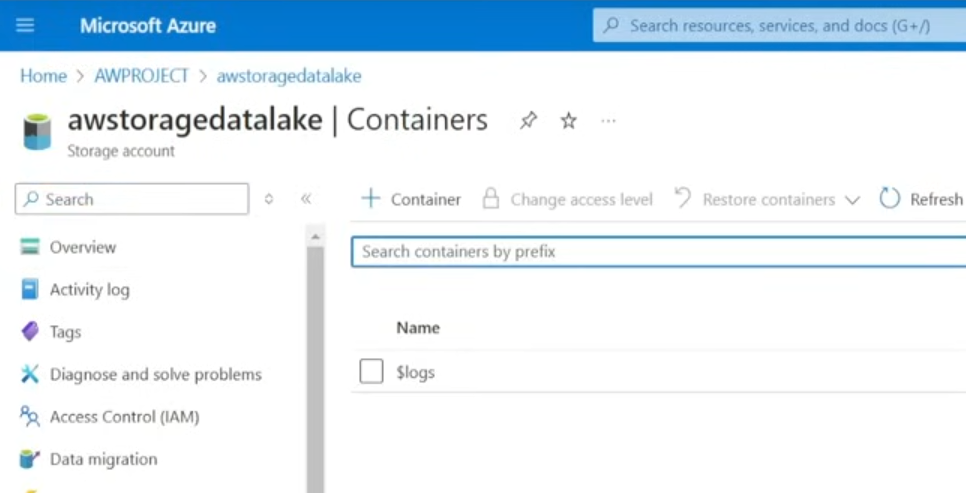
Description automatically generated

Above tabs are used to pick ADF building blocks such as : pipelines, activities, functions, notebooks etc.

Go to home, open storage account created:



Select containers:



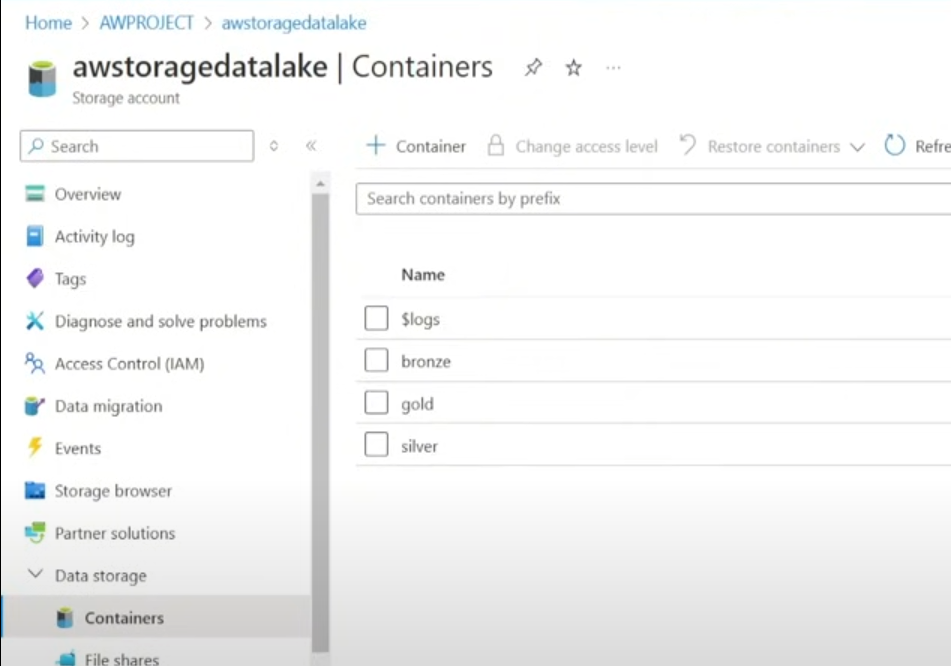
Create 3 containers for : bronze, silver, gold

Click ‘+ Container’ to create a container. Mention name as below and click ‘Create’:

A screenshot of a computer

Description automatically generated

Likewise,create for silver, gold:



Load the data to bronze layer by creating linked service

Create a pipeline:

We need source and destination to create a pipeline.

A screenshot of a computer

Description automatically generated

Create a name for pipeline and click on properties, it will hide the naming tab

A screenshot of a computer

Description automatically generated

Copy activity:

Click ‘move and transform’ then drag and drop ‘copy data’ and rename ‘Name’

A screenshot of a computer

Description automatically generated

‘CopyRawData’ will load data from source and place in destination

In Github URL having data

Create a linked service to create a connection to push data to datalake

‘CopyRawData’ needs to build connection between source and destination

These connections called as linked service

Create static and dynamic pipeline

Github link is having 2 URLs: base and relative URLs

<https://raw.githubusercontent.com/anshlamba03/Adventure-Works-Data-Engineering-Project/refs/heads/main/Data/AdventureWorks_Calendar.csv>

yellow color: base URL

remaining is relative URL

create linked service using 2 ways

1st way:

A screenshot of a computer

Description automatically generated

Click on CopyRawData and source, new

2nd way: (recommended)

A screenshot of a computer

Description automatically generated

After selecting http, click on continue

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Click test connection