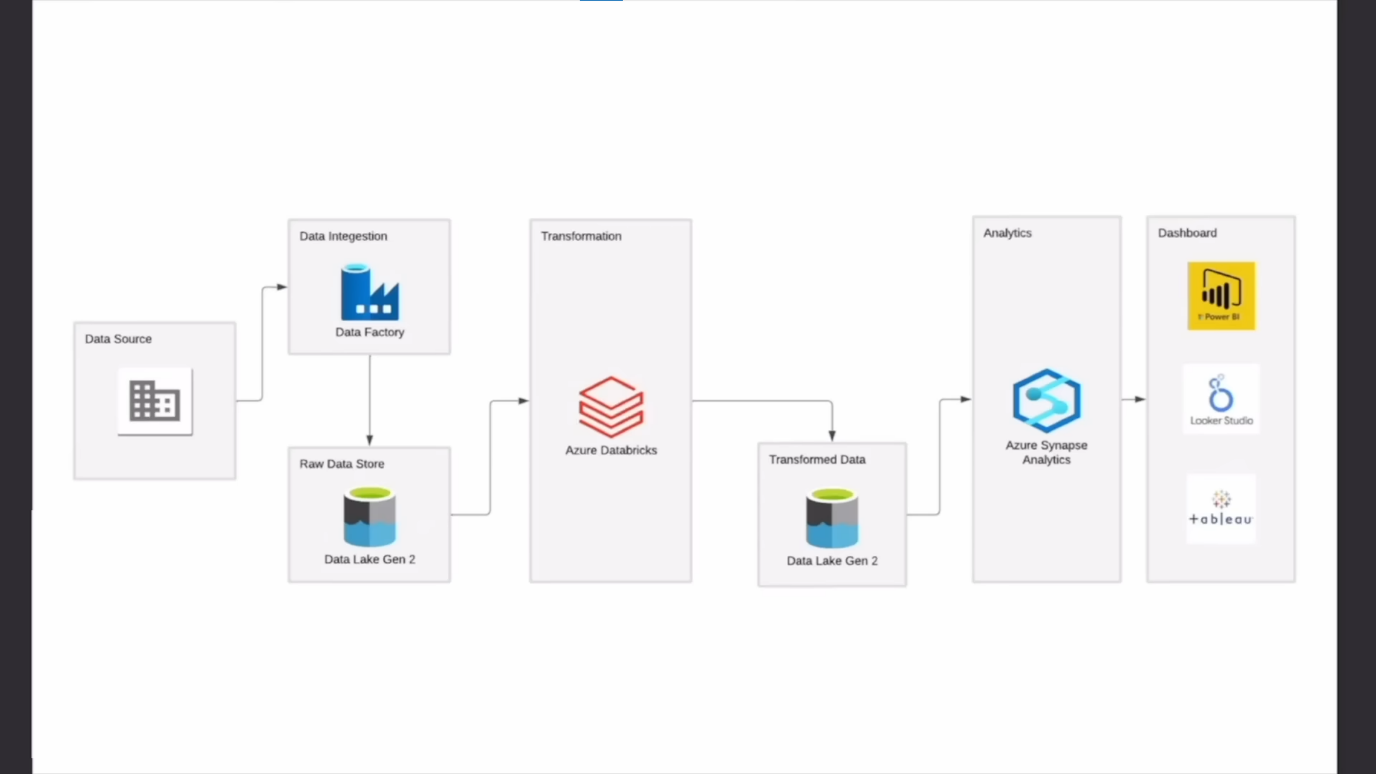
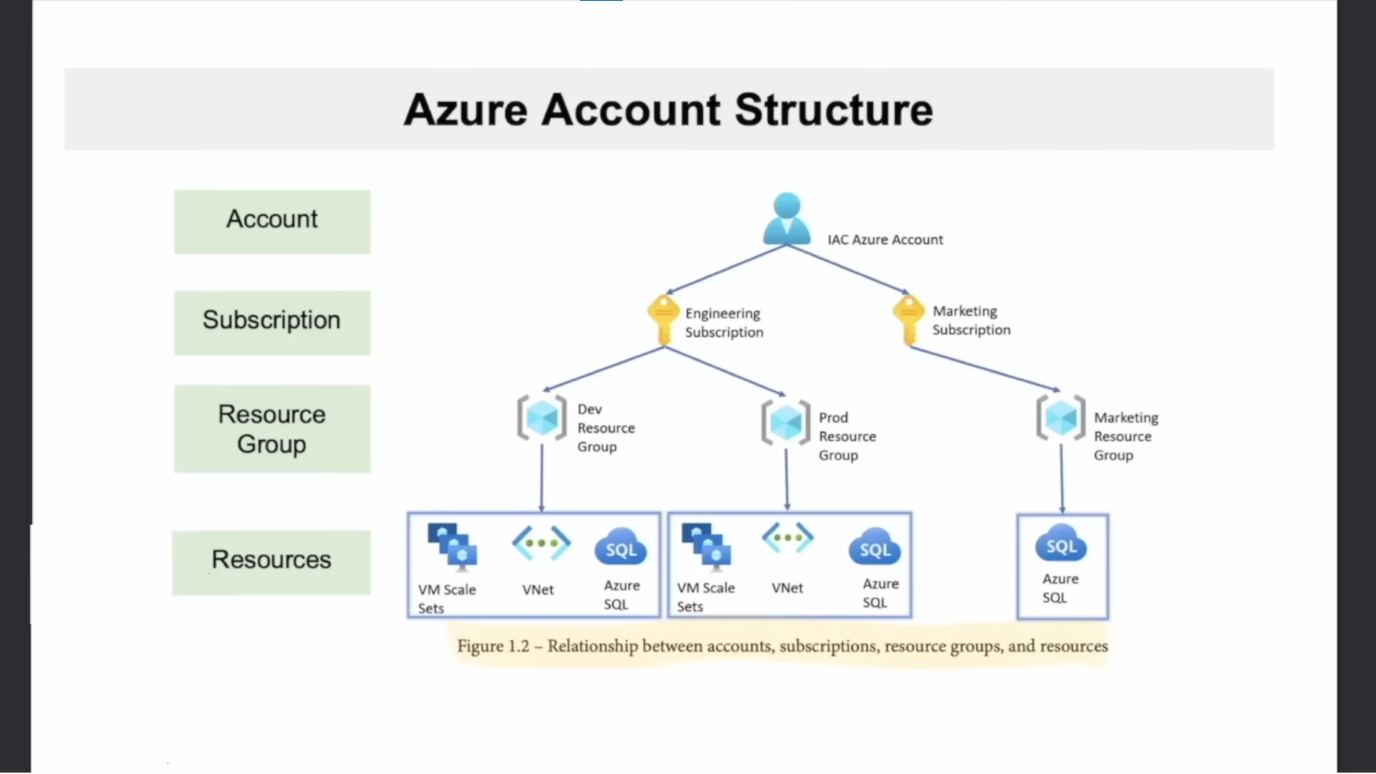
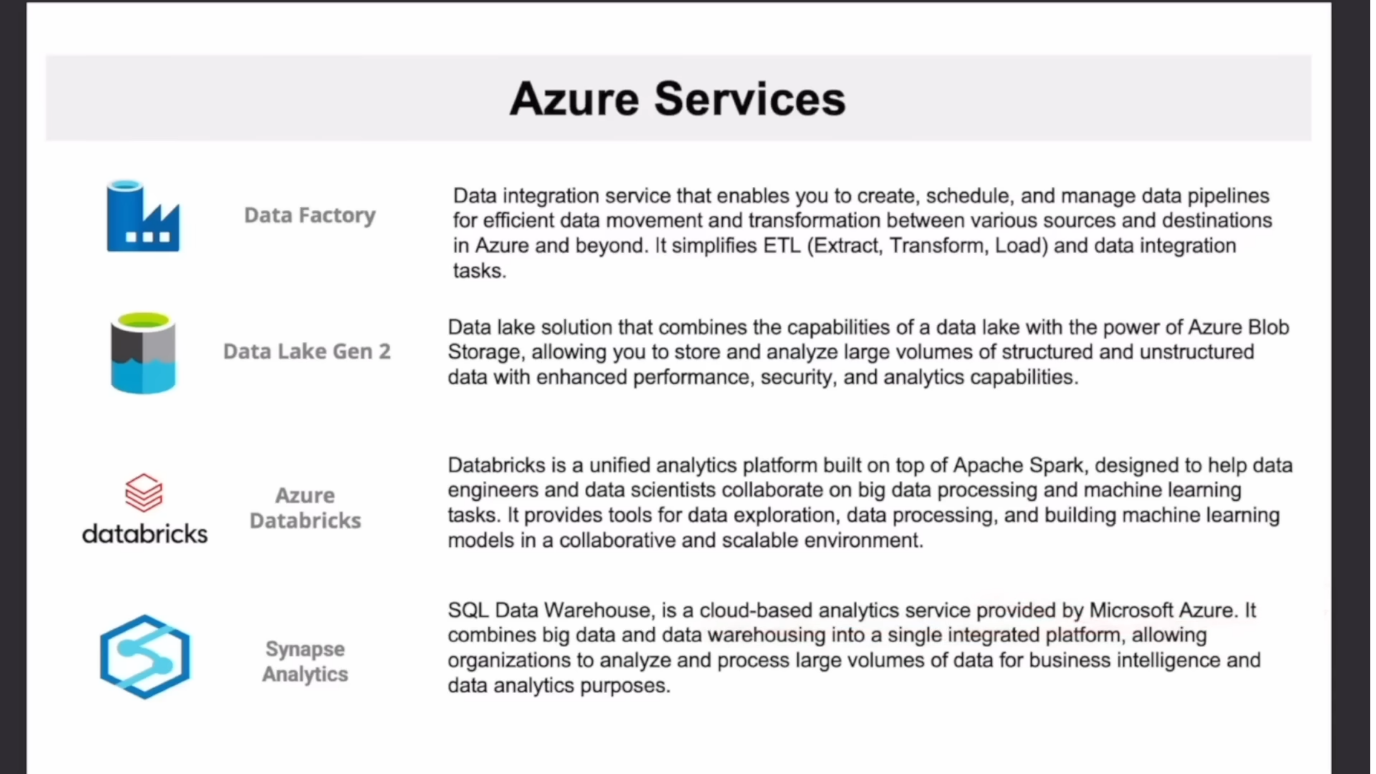
Overall architecture of the setup

# Azure Account Structure



Azure Services Used

# Overview of the steps followed in the project

* Create storage account
* Stored data into container (raw and transformed data)
* create a datafactory resource and assign the resource group
* create a linked service and select the source of data -> map the sink to data lake storage gen 2 (ADLS)
* Validate the pipepline
* Create databricks resource and the cluster to run the spark jobs
* Start notebook -> attach the cluster
* Create app registration in azure for databricks
* Application client id and tenant id & New secret id
* Use Key vault for credentials
* Make connection with spark notebook
* Go to container -> IAM and give permissions/role to the app
* Create synapse analytics workspace
* Goto data and use lake database -> create database and tables -> load the data(infer column)