

**Data Source**

1.Using GIT as our primary data source.

**Azure Data Factory**

1.It is a Data Pipeline

2.It provides integration service that integrates data from various sources

**Azure Data Lake Solution**

1.It is an object storage

2.Stores structured and unstructured data.

**Azure Databricks**

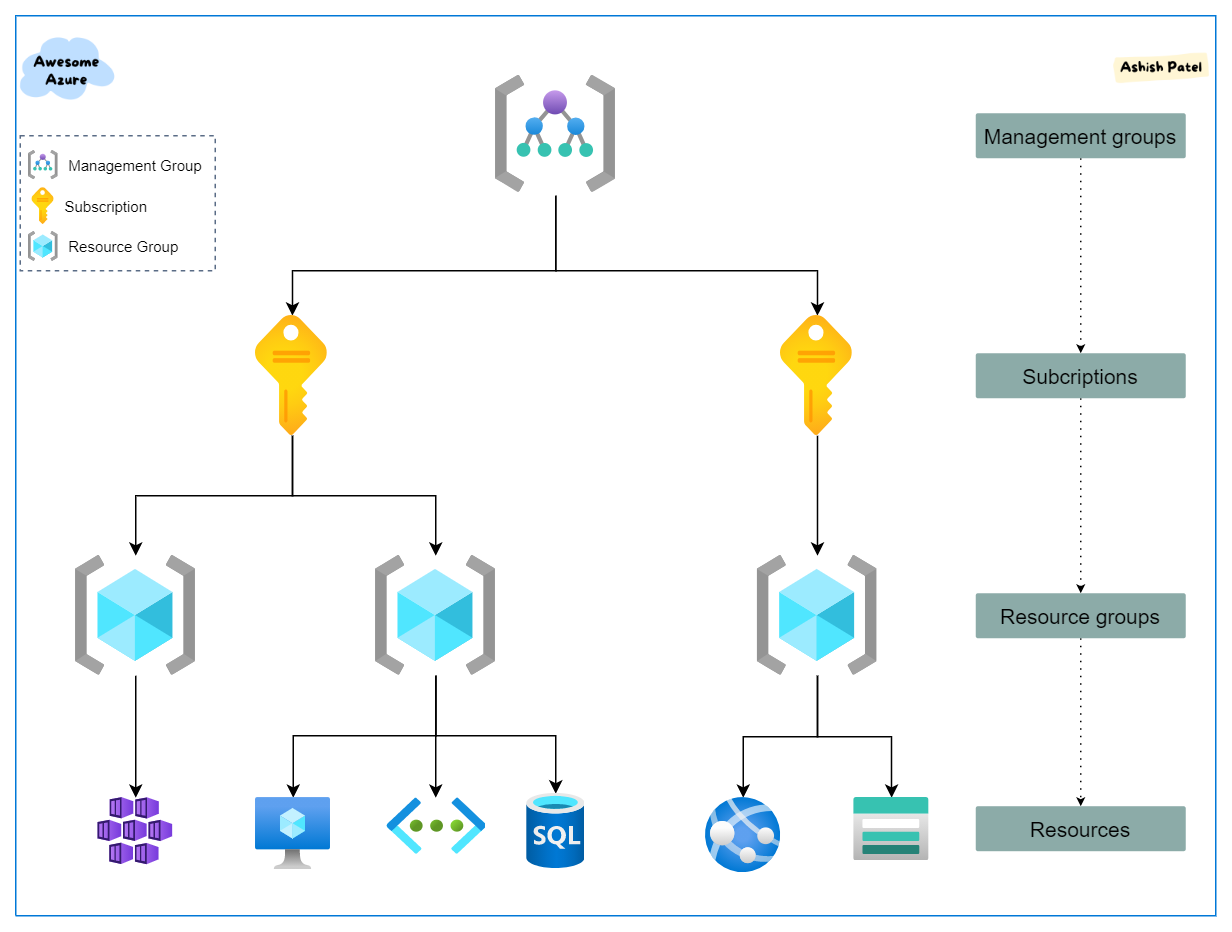
It is an analytical platform that provides an environment to transform the data using the power of Apache Spark.

**Azure Synapse Analytics**

1.Cloud sql datawarehouse

2.Analysing data using queries.

**Azure Account Structure**

****

**STEPS for ETL**

1.Extract data from GIT.

2.Extraction is executed using ADF Pipeline

3.The extracted raw data will be stored in Azure Data Lake.

4.The raw data is transformed using Pyspark in Azure Data Bricks.

5.Transformed data is again stored in Azure Data Lake.

6.Target tables are created In Azure Synapse Analytics and transformed data are loaded to perform queries and analysis.

**STEPS for Dashboarding(playing with csv and parquet format)**

1.Extract the raw csv file from data lake using ADF pipeline in Azure Synapse Analytics.

2.Data pipeline transforms source data type and writes back to data lake in parquet form

3.Create spark database and point spark to parquet file.

4.Query spark table using server less sql

5.PowerBi reads from server less sql and projects graph.