

## Assignment – 1

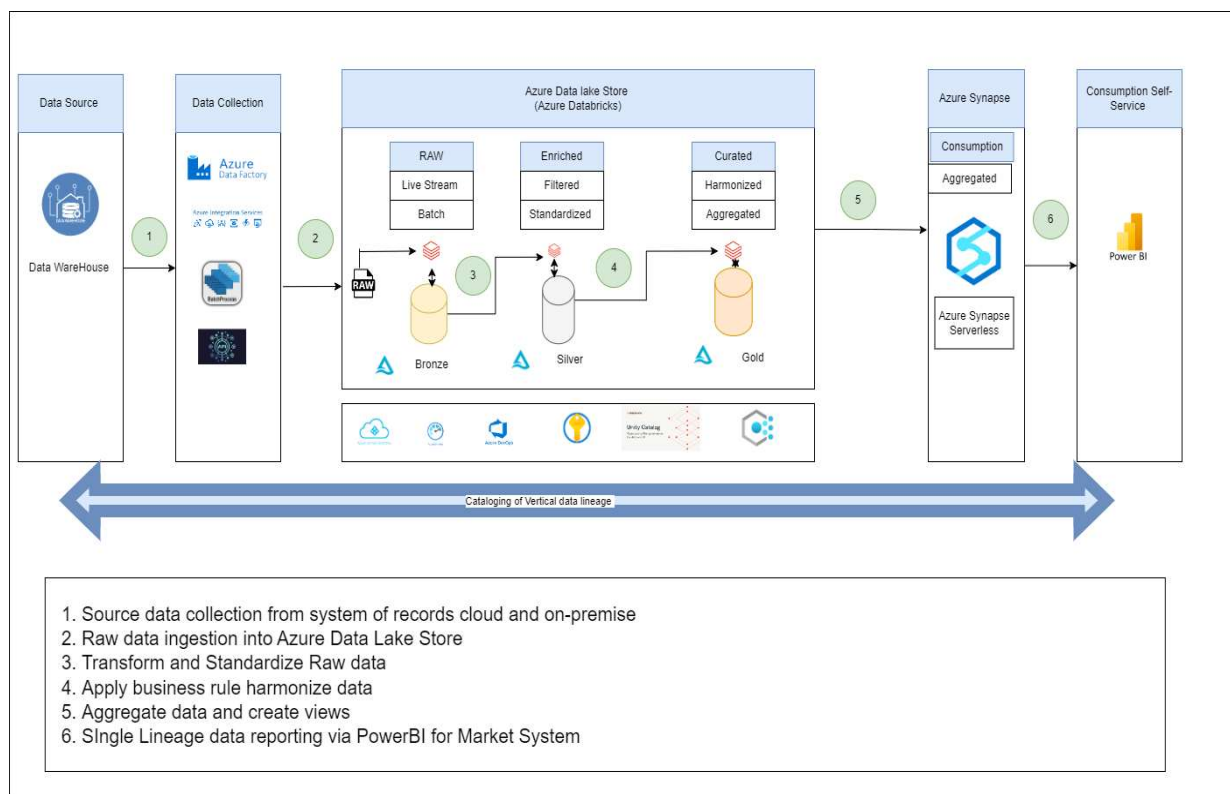
### Requirement :

- Marketing System to ingest Data from Cetralized storage into Designed Storage
- The solution should be scalable and future Proof
- The solution should be compatible with both On-Premise and Cloud
- Data set should be readily accesible for dashboarding

### Assumptions :

- The current Datawarehouse setup is based on Star Schema /Snowflake Schema
- The datawarehouse tables are having unique identfier to identify change in the record
- The Cetralized store team is using Azure as one of the products ,if there is change in Cloud provider the architecture will change accordingly

### Architecture :



**Note :** If above architecture is not visible ,png has been attached in the Git and Folder .

### Integration Runtime :

- The compute infrastructure Integration Runtime(IR) is used by ADF pipelines to copy the data across different networks

### Azure Data factory :

- Orchestrator for data pipeline activities
- Data ingestion from Datawarehouse to Bronze

### Databricks and Unity Catalog :

- Databricks with Spark ,Delta tables ,DLT and UC catalog will be used for data processing and Data lineage

- **Data lake Landing and Standardised Layer (bronze) :** The source data will ingested (pull jdbc/file read etc and Push ) as raw format /parquet .No Data Quality ,transformation and business rules will be applied .
- **Curated data persistence (Silver) :** The standard data from bronze layer will be curated by applying enrichment and data quality rules and stored in silver container ,The data will be stored in Delta lake format .Maintained as SCD 2 and can be accessed by SQL Pools .ACID properties will be used to maintain scd type 2 .
- **Publised data persistence(Gold) in ADLS Gen2:** Apply the business rules and aggregation,maintained as SCD 2 table ,Used for Analytical data model .

### **Synapse Server less SQL Pool :**

Serverless sql pool will conatin gold layer data and will be used for Consumption from Power BI and other analytical tool ,This will ve also be the repository to store the metadata tables required for ingestion .

### **Azure Key vault :**

All the secret will be managed in the Key Vault

Day - 0 Strategy :

The pipline will load data from source warehouse to read data without data filter ,Fact tables can be loaded without filter of data ,for Dim table full load will be done

Day - 1 Strategy : Increment data filter will be used to load the data from source datawarehouse to Azure containers ,full load for Dimension table .

Backfill : In case of missing data backfill can be used with the incremental load to fill missing dates