# ****RETAIL STORE AZURE PROJECT****

**BRONZE LAYER (RAW DATA INGESTION)**

**AZURE SERVICES:**

1. **SQL Database Creation:**
   * Established an Azure SQL Database and created individual tables corresponding to each dataset — *Store, Transaction, Customer,* and *Product.*
2. **Storage Account Setup:**
   * Configured Azure Storage Accounts and created dedicated containers for each file type: *store, transaction, customer,* and *product.*
3. **Data Ingestion via Data Factory:**
   * Implemented **Copy Activities** in Azure Data Factory (ADF) for each dataset.
   * Defined **Source** as *Azure SQL Database* and **Sink** as *Azure Data Lake Storage Gen2* (ADLS Gen2), storing data in **Parquet** format for optimized performance.
4. **API-based Data Extraction:**
   * For *Customer Data*, performed Copy Activity using an **HTTP connector** as the source (JSON format API response) and stored the output in **Parquet** format in ADLS.

**DATABRICKS:**

1. **Storage Mounting:**
   * Mounted the Azure Data Lake Storage (ADLS Gen2) within **Databricks** using PySpark for seamless data access.
2. **Dataframe Creation:**
   * Loaded all four datasets — *Store, Transaction, Customer,* and *Product* — into Spark DataFrames using the spark.read.parquet() function, as the files were stored in Parquet format.

**SILVER LAYER (DATA CLEANING AND TRANSFORMATION)**

**DATABRICKS:**

1. **Data Cleaning:**
   * Performed comprehensive data cleaning by:
     + Converting columns to their appropriate data types.
     + Removing null values and duplicate records to ensure data quality and integrity.
2. **Data Integration:**
   * Created a **consolidated Silver dataset** by joining all four DataFrames using their respective ID relationships (primarily from the *Transaction* dataset).
3. **Data Storage in Delta Format:**
   * Stored the transformed Silver dataset in **Delta** format in ADLS Gen2 for efficient querying and version control.
4. **SQL Table Conversion:**
   * Registered the Silver DataFrame as a **SQL Table** within Databricks for further analysis and querying.

**GOLD LAYER (BUSINESS AGGREGATION AND ANALYTICS)**

**DATABRICKS:**

1. **Business-Level Aggregation:**
   * Performed grouping and aggregation based on business requirements to derive key metrics, including:
     + **Total Quantity Sold**
     + **Total Sales Amount**
     + **Number of Transactions**
     + **Average Transaction Value (ATV)**
2. **Gold Dataset Storage:**
   * Stored the finalized Gold dataset in ADLS Gen2 in Delta format for downstream analytics and reporting.
3. **SQL Table Registration:**
   * Converted the Gold DataFrame into a **SQL Table** in Databricks for easy integration with visualization tools.

**POWER BI DASHBOARD (VISUALIZATION AND INSIGHTS)**

1. **KPI Card Visuals:**
   * Created interactive card visuals for key performance indicators:
     + *Total Sales*
     + *Total Quantity Sold*
     + *Number of Transactions*
2. **Chart Visualizations:**
   * Designed analytical charts to uncover business trends:
     + *Total Sales by Month*
     + *Total Sales by Store*
     + *Total Sales by Product*
     + *Total Quantity Sold by Category*
3. **Dynamic Filtering:**
   * Implemented a **Location-based Slicer** to enable dynamic filtering across all visuals and enhance analytical interactivity.

**SUMMARY:**

The **Retail Store Azure Project** demonstrates an end-to-end data analytics solution built using **Azure SQL Database, Data Factory, Data bricks, and Power BI** following the **Medallion Architecture (Bronze–Silver–Gold)**. Data was ingested from SQL tables and APIs into **Azure Data Lake Storage Gen2** in Parquet format (Bronze Layer), cleaned and transformed using **PySpark in Data bricks** to create a unified and structured **Silver dataset**, and further aggregated into key business metrics such as Total Sales, Quantity Sold, Number of Transactions, and Average Transaction Value in the **Gold Layer**. Finally, the refined data was visualized in **Power BI** through interactive dashboards and KPI cards, providing actionable insights into store performance, product sales, and customer behaviour.