## **Customer 360 Data Integration**

Bootcamp Project - 3

### **Overview**

A retail business wants to build a unified Customer 360 view by integrating data from multiple sources, including online transactions, in-store purchases, customer service interactions, and loyalty programs. This project uses a mix of fact and dimension tables to ensure a clean, scalable structure.

### **Architecture**

* **Tool**: [Draw.io](http://draw.io/) to design the architecture diagram.

**Role**

Assume the role of Data Engineers responsible for implementing an end-to-end data pipeline in Azure.

### **Tools and Technologies**

* **Azure Synapse Analytics**
* **Azure Data Lake Storage (ADLS)**
* **Azure SQL Database**
* **Power BI**

### **Step-by-Step Process**

#### **Step 1: Ingest Data**

* Use Azure Synapse Analytics pipelines to ingest data from multiple sources (online, in-store, customer service, loyalty programs) into the **raw** container in ADLS.

#### **Step 2: Define Staging (Curated) Schema**

* Create DDLs for a **staging (silver)** database in Azure Synapse Analytics to clean and standardize the data.

#### **Step 3: Define Analytics (Gold) Schema**

* Create DDLs for an **analytics (gold)** database in Azure SQL Database for aggregated, business-ready data.

#### **Step 4: Load Curated Data**

* Processed data moves to the **staging** database, Azure SQL DB.

#### **Step 5: Load Analytics Data**

* Aggregated, summarized data(below tables) are loaded into the **analytics** database in Azure SQL Database.
* View 1 - for **Average Order Value (AOV)**
* SUM(Amount) / COUNT(OrderID) per product, category, and location.
* View 2 - for Segment customers based on total spend, purchase frequency, and loyalty tier (**LoyaltyAccounts.TierLevel**).
  + Example: "High-Value Customers" (Top 10% spenders), "One-Time Buyers," "Loyalty Champions."
* View 3 - for Analyze **DateTime** to find peak days and times in-store vs. online.
* View 4 - for Number of interactions and resolution success rates per agent (**ResolutionStatus**).

#### **Step 6: Build a Power BI Dashboard**

* Develop a Power BI dashboard for Customer 360 insights, combining data from all sources.
* Publish the final report to Fabric Workspace.

### **Ways to Verify and Validate**

* **Ingestion Validation:** Ensure all data sources are correctly ingested into ADLS.
* **Data Integrity Checks:** Verify data consistency and relationships across tables.
* **Dashboard Validation:** Compare Power BI dashboard outputs with sample data for accuracy.

### **Deliverables**

* **Ingested Data:** Raw data from all sources in ADLS.
* **Curated Data:** Cleaned data stored in Azure Synapse (Silver layer).
* **Aggregated Data:** Business-ready, summarized data stored in Azure SQL Database (Gold layer).
* **Power BI Dashboard:** Visual representation of Customer 360 insights.

### **Project Completion Criteria**

* Successful integration of all customer data sources.
* Clean, consolidated customer profiles stored in Azure SQL.
* Power BI dashboard provides actionable Customer 360 insights.

Mentees should upload all relevant details to GitHub, including:

* **SQL Scripts**: DDL (schema creation) and DML (sample data) scripts.
* **Pipeline JSON**: Exported JSON of Azure Synapse Analytics pipelines.
* **Documentation**: A README file or a .doc file explaining the project steps, architecture, and key considerations.

**Sample DDL Scripts**

|  |
| --- |
| CREATE TABLE Customers (  CustomerID INT PRIMARY KEY,  Name VARCHAR(100),  Email VARCHAR(100),  Address VARCHAR(255) );  CREATE TABLE Products (  ProductID INT PRIMARY KEY,  Name VARCHAR(100),  Category VARCHAR(50),  Price DECIMAL(10, 2) );  CREATE TABLE OnlineTransactions (  OrderID INT PRIMARY KEY,  CustomerID INT,  ProductID INT,  DateTime DATETIME,  PaymentMethod VARCHAR(50),  Amount DECIMAL(10, 2),  Status VARCHAR(20),  FOREIGN KEY (CustomerID) REFERENCES Customers(CustomerID),  FOREIGN KEY (ProductID) REFERENCES Products(ProductID) );  CREATE TABLE Stores (  StoreID INT PRIMARY KEY,  Location VARCHAR(100),  Manager VARCHAR(100),  OpenHours VARCHAR(50) );  CREATE TABLE InStoreTransactions (  TransactionID INT PRIMARY KEY,  CustomerID INT,  StoreID INT,  DateTime DATETIME,  Amount DECIMAL(10, 2),  PaymentMethod VARCHAR(50),  FOREIGN KEY (CustomerID) REFERENCES Customers(CustomerID),  FOREIGN KEY (StoreID) REFERENCES Stores(StoreID) );  CREATE TABLE Agents (  AgentID INT PRIMARY KEY,  Name VARCHAR(100),  Department VARCHAR(50),  Shift VARCHAR(50) );  CREATE TABLE CustomerServiceInteractions (  InteractionID INT PRIMARY KEY,  CustomerID INT,  DateTime DATETIME,  AgentID INT,  IssueType VARCHAR(50),  ResolutionStatus VARCHAR(50),  FOREIGN KEY (CustomerID) REFERENCES Customers(CustomerID),  FOREIGN KEY (AgentID) REFERENCES Agents(AgentID) );  CREATE TABLE LoyaltyAccounts (  LoyaltyID INT PRIMARY KEY,  CustomerID INT,  PointsEarned INT,  TierLevel VARCHAR(20),  JoinDate DATE,  FOREIGN KEY (CustomerID) REFERENCES Customers(CustomerID) );  CREATE TABLE LoyaltyTransactions (  LoyaltyID INT,  DateTime DATETIME,  PointsChange INT,  Reason VARCHAR(100),  PRIMARY KEY (LoyaltyID, DateTime),  FOREIGN KEY (LoyaltyID) REFERENCES LoyaltyAccounts(LoyaltyID) ); |

**Reference Dataset** For realistic sample data, refer to [Kaggle’s Customer 360 Data](https://www.kaggle.com/datasets/varunkumari/customer-360-data).