Name:- Prathamesh Kodgirwar

Int-23

Project Title:
Finance Bank Data Processing and Dashboarding Using Microsoft Fabric

---

Objective:
To design and implement a scalable, end-to-end data pipeline leveraging SQL Server, Azure Data Factory, and Microsoft Fabric, structured using the **Medallion Architecture**. The goal was to efficiently ingest, transform, and analyze banking financial data and present actionable insights through a dynamic Power BI dashboard.


Tools & Technologies Used:

- **SQL Server / Azure SQL Database** – Initial data storage and transformation

- **Azure Data Factory** – Data ingestion and movement

- **Microsoft Fabric** – Data Lake (OneLake), Notebooks (PySpark & Spark SQL), Warehouse, Delta Lake tables

- **Power BI** – Visualization and business intelligence layer


Step-by-Step Workflow:

1. Data Ingestion (Bronze Layer)

- **Source**: Excel files containing historical and transactional financial banking data.

- **Preprocessing**: Uploaded and mapped Excel data into SQL Server staging tables.

- **Process**: Leveraged **Azure Data Factory (ADF)** Copy Data activity to extract from SQL Server and ingest into **Microsoft Fabric OneLake** as **Parquet** files.

- **Storage**: Raw data stored in the Bronze layer under a logical folder structure in OneLake
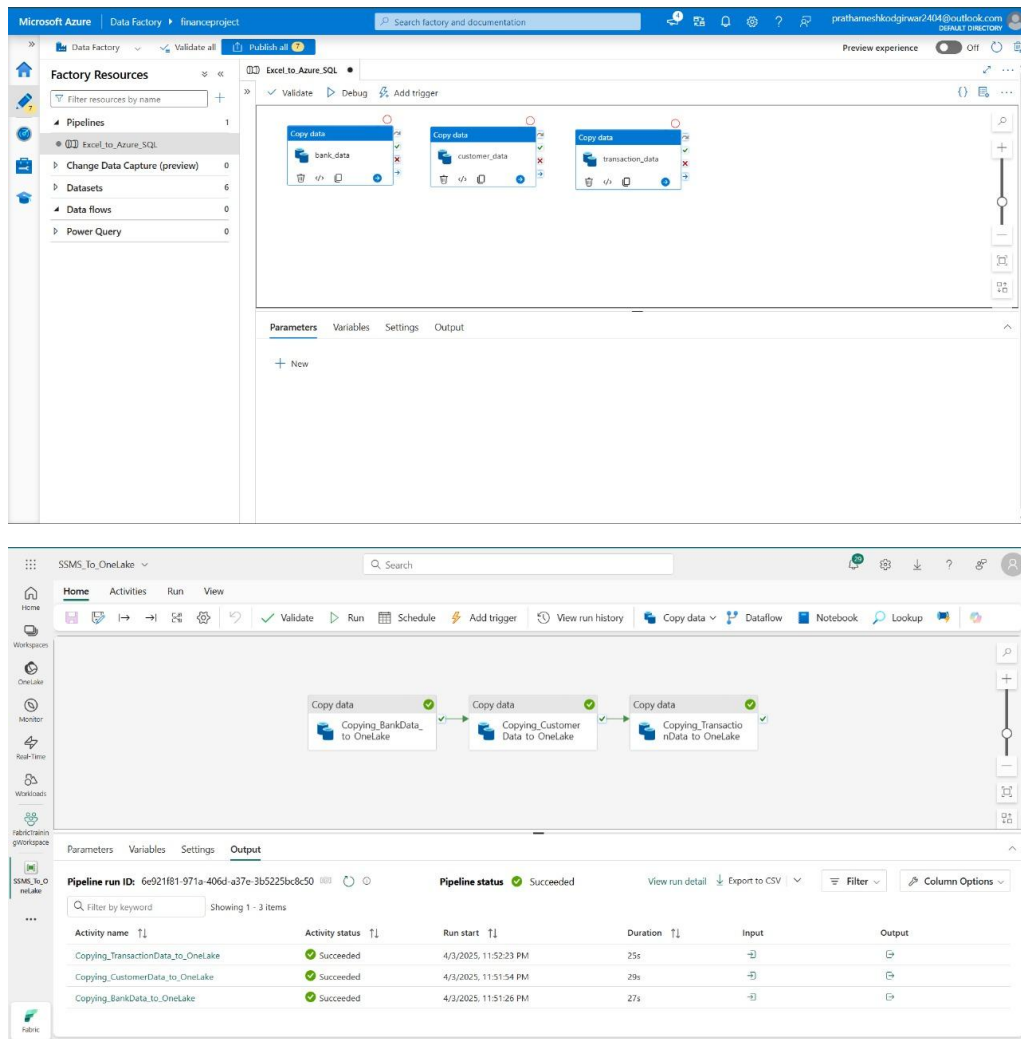
2. Data Validation & Structuring (Silver Layer)

- **Tools Used**: Microsoft Fabric **Notebooks** using **PySpark** and **Spark SQL**.

- **Key Tasks**:

    o Detection and handling of **null values** and **missing records**.

    o **De-duplication** to ensure data integrity.

    o Standardization of schema through **column renaming**, **data type alignment**, and **format consistency**.

- **Output**: Cleaned and structured data written as **Delta Tables** in the Silver layer.

---
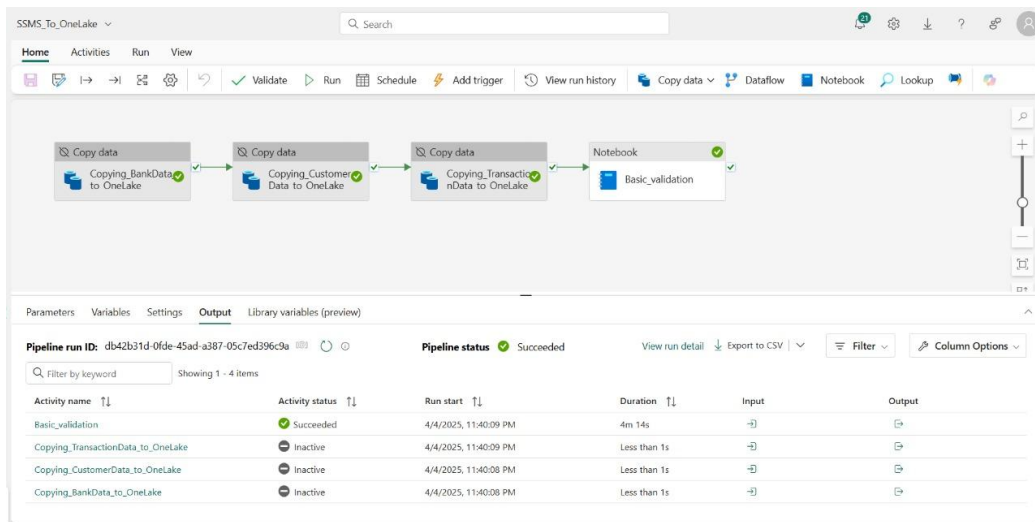
3. Data Transformation & Enrichment (Gold Layer)

- Created transformation logic using **additional Fabric Notebooks**.

- **Enhancements Included**:

  - Creation of **derived/aggregated columns** for KPIs and business metrics.

  - Implementation of **business rules** (e.g., account status logic, customer segmentation).

  - Mapped and merged related datasets (customer, accounts, transactions).

- **Output**: Final analytical datasets loaded into **Microsoft Fabric Warehouse** for optimized querying and seamless Power BI integration.

---
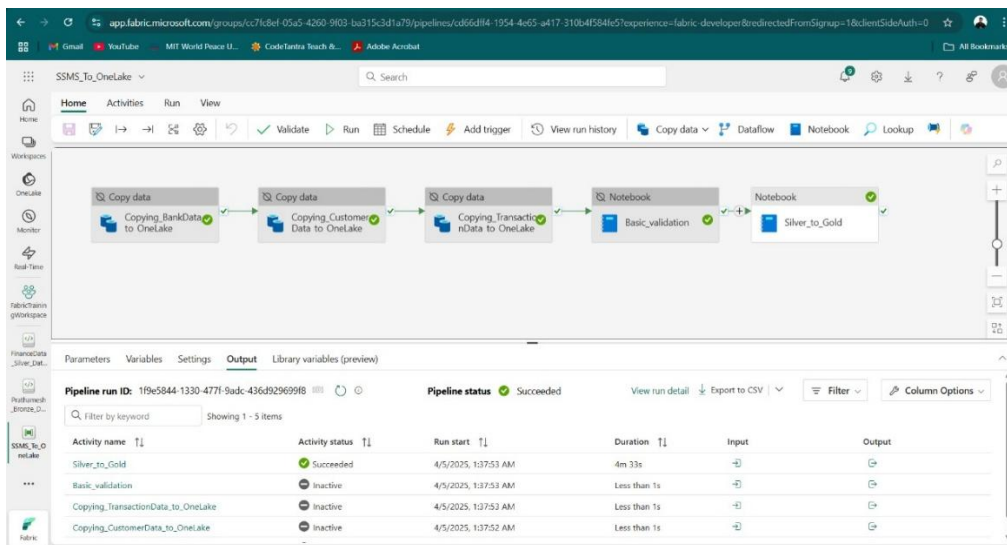
4. Data Visualization

- **Tool Used**: Power BI (connected directly to Fabric Warehouse).

- **Dashboard Highlights**:

  o **Key financial metrics** (e.g., account balances, total deposits/withdrawals).

  o **Time-series trends** (monthly, quarterly, and yearly views).

  o **Segmentation analysis** (by customer type, region, and account category).

  o Monitoring of **data quality issues**, including nulls, anomalies, and processing errors.
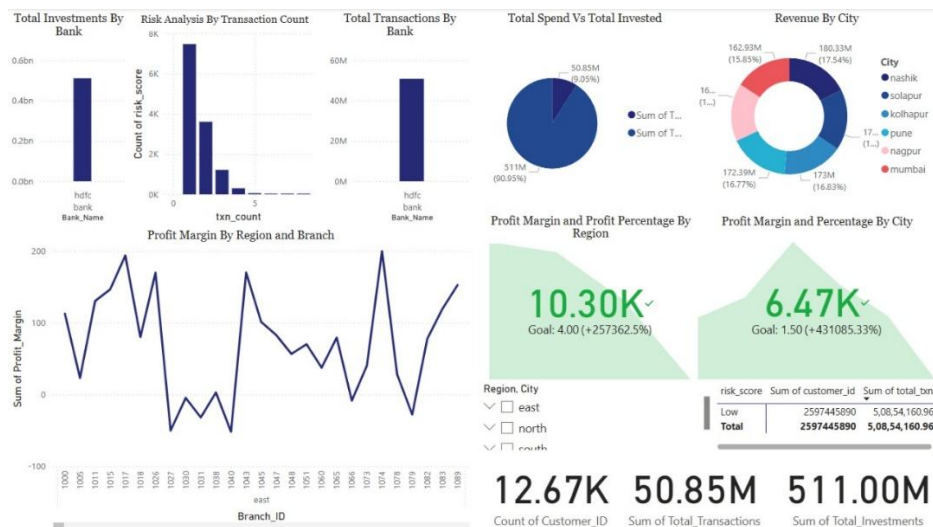
Architecture Followed: Medallion Architecture

Layer	Purpose	Storage	Tools Used
Bronze	Raw data ingestion	OneLake (Parquet)	Azure Data Factory
Silver	Cleaned & structured data	Delta Tables	Fabric Notebooks (PySpark/SQL)
Gold	Analytical, business-ready	Fabric Warehouse	Notebooks + Power BI

Challenges Faced:

- **Data type mismatches** during migration from Excel to SQL Server.

- Efficiently processing **large and inconsistent Excel files** with formatting issues.

- Managing **schema evolution** in Delta tables while maintaining data quality and lineage.

- Ensuring optimal performance during **Power BI direct queries** from Fabric Warehouse.

Outcome:

- Developed a **fully automated, scalable**, and **cloud-native** data pipeline.

- Delivered **clean, reliable, and analytics-ready** financial datasets.

- Enabled stakeholders to interact with a **live Power BI dashboard** providing real-time business intelligence, operational insights, and decision-making support.