Project Documentation Boot Camp 3

# Project Overview

The project involves developing a data warehousing solution using Azure Synapse Analytics for a financial institution to consolidate customer transaction data for business intelligence reporting.

# Tools and Technologies Used:

- Azure Synapse Analytics

- Azure Data Factory (ADF)

- Azure Data Lake Storage

- Power BI

# Detailed Implementation

## 1. Data Modeling:

I designed a data model for customer transactions that could accommodate the data collected from various transaction types and customer interactions.

## 2. Data Ingestion:

The raw transaction data is ingested into Azure Data Lake Storage using Azure Data Factory (ADF). This process was handled by a pipeline activity named 'Copy data1', which pulled data from a SQL Server source and wrote it into a Synapse SQL Pool.

A screenshot of a computer

AI-generated content may be incorrect.

## 3. Data Transformation:

Data transformation tasks were conducted in Azure Synapse using SQL pools. These transformations were managed through Notebook activities. Specifically, Notebook 3 handled the transformations for three critical tables: Customer, Transactions, and Accounts.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

## 4. Optimization:

The system was optimized for performance using indexing and partitioning strategies defined in your SQL scripts. These optimizations ensure efficient data retrieval and query execution, crucial for real-time reporting and analytics.

## 5. Reporting:

A Power BI report was created to provide insights into customer behaviors and transaction patterns. This report allows the financial institution to make data-driven decisions based on customer transaction trends.

A screenshot of a graph

AI-generated content may be incorrect.

## 6. Verification and Validation:

Data integrity was verified by comparing raw and transformed data. Additionally, query performance was validated using execution plans to ensure that the data retrieval processes were optimized.

# Deliverables:

- Data model schema.

- ADF pipeline configurations for data ingestion.

- Optimized Synapse SQL pool setup.

- Power BI report showcasing customer analytics.

# Project Completion Criteria:

- Successful integration of transaction data into Synapse.- Data warehouse optimized for quick query performance.- Comprehensive Power BI reporting reflecting accurate and insightful data analytics.