Financial Transaction Monitoring

Group Members

1. **Abhinaya M.**
2. **Lakshan S J.**
3. **Pavithra Sri S.**
4. **Sachin Sahadev Singh**

Problem Statement

Financial Transaction Monitoring

ETL Problem

• Ingest real-time transactions from Event Hubs into Azure Databricks Structured Streaming.

• Clean and flag potential issues, store processed stream in Delta Lake.

• Maintain infra-as-code (IaC) using Azure DevOps ARM/Bicep templates for Event Hub & Databricks setup.

ML Problem

• Use Isolation Forest or Autoencoder to detect fraudulent transactions.

• Alert integration using Azure Logic Apps + Power BI dashboards.

Procedure

1	2	3	4	5
Load Dataset From Kaggle (https://www.kaggle.com/code/benroshan/transaction-fraud-detection/input)	Introduction: End-to-End Fraud Detection Platform	Step 1 & 2: Data Acquisition & Cloud Storage	Step 3: Ingest into Databricks – Bronze Table	Step 4: Clean & Transform – Silver Table
6	**7**	**8**	**9**	**10**
Step 5: Feature Engineering – Gold Table	Step 6: Build & Register ML Model	Step 7: Orchestration with Apache Airflow	Step 8: Real-time Alerts via Azure Logic Apps	Step 9: Visualize Fraud Trends in Power BI

Step 1 & 2: Data Acquisition & Cloud Storage

1

Source Raw Transaction Data: Obtain relevant financial transaction data from Kaggle for initial analysis and development.

2

Secure Cloud Data Lake: Upload dataset to Azure Storage Blob, establishing a scalable, accessible data source.

3

Foundation for Analytics: Centralized cloud storage provides a robust base for all subsequent data processing steps.

Step 3: Ingest into Databricks – Bronze Table

1

Raw Data Ingestion Layer: Read raw CSV from Azure Blob into Databricks, creating the Bronze layer Delta table.

2

Immutable Data Capture: Bronze table acts as an unchangeable, versioned record of all raw input data.

3

Foundation for Delta Lake: This is the crucial first step in building a reliable and scalable Delta Lake architecture.

Step 4: Clean & Transform – Silver Table

Enhance Data Quality: Clean missing values, cast data types, and standardize formats for better analysis.

Prepare for Analysis: Derive new features like 'hour from timestamp' and 'amount band' for deeper insights.

Governed Data Layer: The Silver layer provides a refined, trustworthy dataset for downstream applications.

Step 5: Feature Engineering – Gold Table

ML-Ready Data Creation: Enrich the Silver data with advanced features for machine learning models.

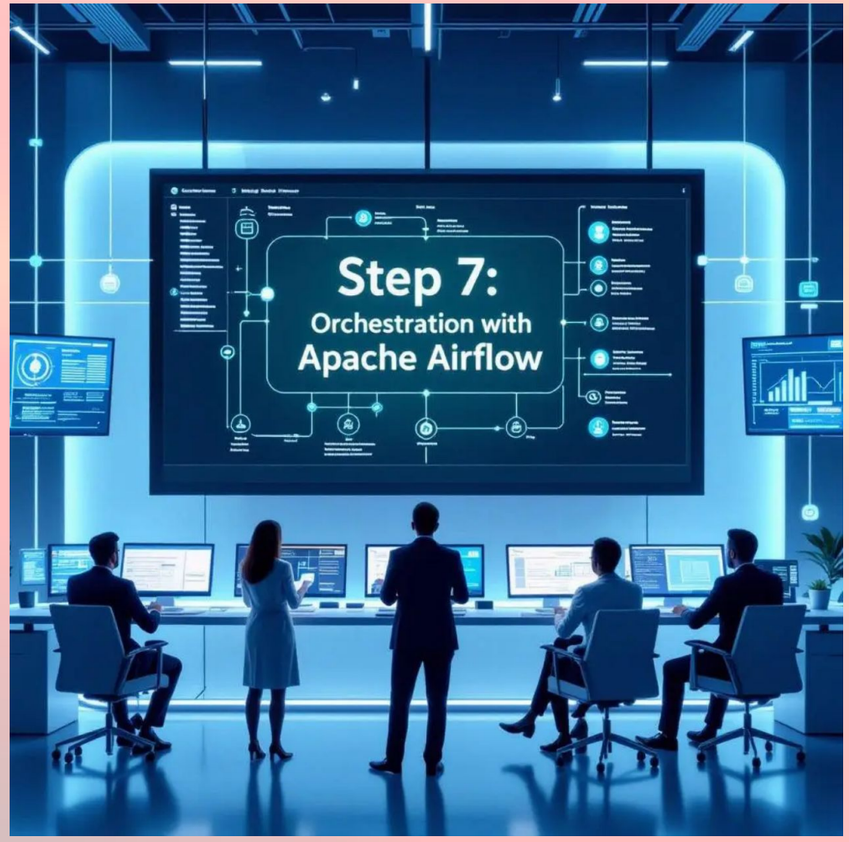
Business Insight Layer: Add 'is_suspicious' flags and 'fraud_risk_score' for actionable business intelligence.

Optimized for Consumption: The Gold table is optimized for direct use by ML models and business intelligence dashboards.

Step 6: Build & Register ML Model

1. **Detect Anomalies with ML:** Train an Isolation Forest model on the Silver data to identify fraudulent transactions.
2. **MLOps with MLflow:** Log model experiments and versions using MLflow for reproducibility and management.
3. **Enhance Fraud Detection:** Machine learning adds an intelligent layer for more accurate and proactive fraud identification.

Step 7: Orchestration with Apache Airflow



Automate Data Pipelines: Orchestrate all steps (Bronze to Gold, ML) as a DAG using Apache Airflow.

Reliable Workflow Management: Use DatabricksSubmitRunOperator to trigger notebooks, ensuring smooth, scheduled execution.

Continuous Processing: Airflow ensures the fraud detection pipeline runs consistently, keeping data fresh.

Step 8: Real-time Alerts via Azure Logic Apps

1

Instant Fraud Notification: Trigger alerts (email/Teams) from Azure Logic Apps upon fraud detection.

2

Configurable Alert Channels: Flexibly send notifications to various platforms based on defined rules.

3

Accelerate Response Time: Automated alerts enable rapid response to minimize potential fraud impact.

Step 9: Visualize Fraud Trends in Power BI

Interactive Fraud Dashboard: Connect Power BI to the Gold Delta table for comprehensive fraud monitoring.

Key Metric Visualization: Visualize fraud trends, suspicious amounts, and risk by merchant and region.

Data-Driven Decisions: Empower analysts with actionable insights for strategic fraud prevention.

Thank You