**Real-Time Data Streaming Analytics for Fraud Detection**  
**Bootcamp Project - 4**

**Objective**

The objective of this project is to build a real-time fraud detection pipeline using Azure Stream Analytics (ASA) to analyze incoming transaction data. The goal is to detect and flag suspicious activities, enabling faster responses to potential fraud.

**Architecture**

A diagram of a transaction fraud detection

AI-generated content may be incorrect.

* **Azure Event Hub** (for ingesting streaming data)
* **Azure Stream Analytics (ASA)** (for real-time processing and filtering)
* **Azure Data Lake Storage (ADLS)** (for storing clean and validated data)
* **Azure Cosmos DB/Azure SQL Database** (for unique transactions and reference data)
* **Azure Synapse Analytics** (for reporting and analysis)
* **Power BI** (for visualization of flagged fraud cases - optional)

**Role Assumption**

As a Data Engineer on the fraud detection project, I was responsible for designing and implementing the Azure Stream Analytics (ASA) job, which processes real-time transaction data from Azure Event Hub. My role focused on writing and optimizing the Stream Analytics Query to detect fraudulent transactions efficiently by implementing below logic

* **High Amount Fraud:** Flagged transactions where amount > 5000.
* **Unusual Location Fraud:** Compared the transaction’s location with the user’s historical location from an **SQL reference table**.
* **Rapid Transaction Fraud:** Used **Tumbling Window** logic to detect 3+ transactions within **10 minutes** from the same user\_id and IP address.

**Resource Group**:

A screenshot of a computer

AI-generated content may be incorrect.

**Implementation Steps**

**1️. Data Ingestion**

* **Configured Azure Event Hub** to receive streaming transactions in **JSON format**.
* Transactions were pushed into the Event Hub, simulating a live data stream.

A screenshot of a computer

AI-generated content may be incorrect.

**2️. Stream Analytics Job**

* **Input Source**: Connected **Azure Event Hub** to **Azure Stream Analytics**.
* **Reference Data**: Used **Azure SQL Database** to store user location history.

A screenshot of a computer

AI-generated content may be incorrect.

* **Processing Logic**:
  + **Filtered out NULL or corrupt transactions**.
  + **Flagged High Amount Fraud**: Any transaction where amount > 5000.
  + **Identified Unusual Location Fraud** by joining transaction data with **Azure SQL Database** (lookup table containing users' historical locations).
  + **Detected Rapid Transaction Fraud**: Used a **Tumbling Window (10 minutes)** to check if a user performed **3+ transactions from the same IP address**.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

**Stream Analytics Query:**



**3️. Data Output**

* **Clean Data**: Valid transactions were sent to **ADLS (parquet format)**.
* **Unique Transactions**: Unique transactions were stored in **Azure Cosmos DB**.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

**4️. Data Analysis & Reporting**

* **Azure Synapse Analytics**: Created **External Tables** on **ADLS data** for querying flagged fraud transactions.
* **SQL Query**: Summarized flagged fraudulent transactions for reporting.
* **Power BI Dashboard**: Designed a **fraud monitoring dashboard** visualizing flagged transactions.

A screenshot of a computer

AI-generated content may be incorrect.

**DDL Script for External Table:**



A screenshot of a computer

AI-generated content may be incorrect.

**GitHub Link:** https://github.com/Arunkumar-Senthilkumar/Fraud-Detection---Stream-Analytics