Streaming Layer Solution

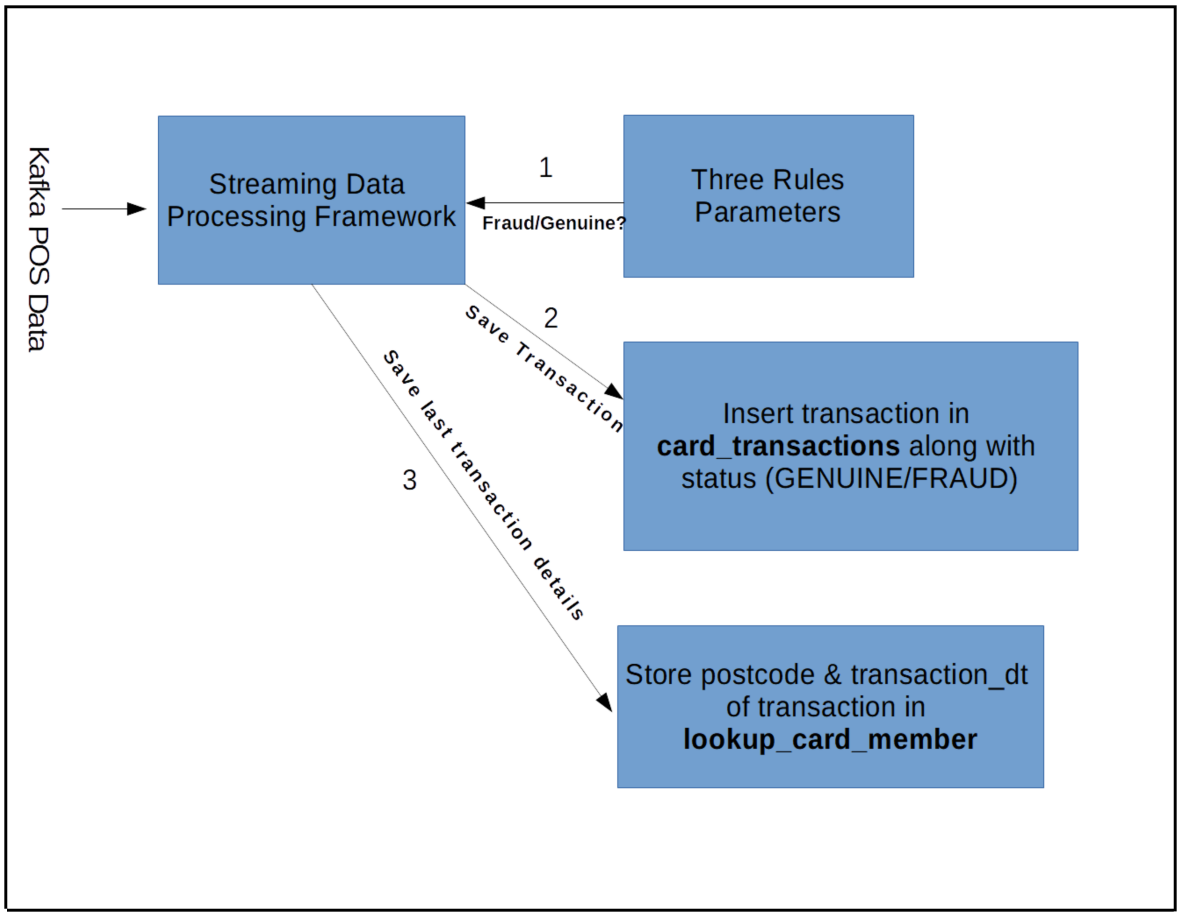
Capstone Project – Final Submission

Venkatesh Jagannathan

Explanation

Basic Steps

1. Connect to Kafka POS Data
2. Process the Stream
    1. De-Serialize Transaction Data from JSON to Java Object
    2. Check for Genuine
        1. Retrieve Card Rule Parameters (ucl, score, post code and last transaction date) from HBase
        2. Check transaction amount against UCL
        3. Get member score and compare against max score (i.e. 200)
        4. Calculate speed of travel between last and current transaction & compare if more than 4km/sec
        5. Set the status of the transaction based on above parameters validation
    3. Save Transaction
    4. Update Last transaction details of a card



The streaming layer starts by consuming incoming transaction stream from kafka topic where card transaction data is publish. The stream is later processed to understand the nature of transaction and classify and genuine or fraud. The result is printed to console & saved in database.

Stream Processing

In income stream contains JSON data which is de-serialized to transaction data that can be conveniently processed downstream. From the converted stream each transaction is check for genuineness using three parameters (I) UCL (II) Score (III) Last Transaction Post Code & Time. These details are fetched from database for each card from lookup table. One classified, the transaction is stored in the card_transaction table. Additionally the postcode and transaction datetime is stored against the card in the lookup table to be used back for validating next transaction.

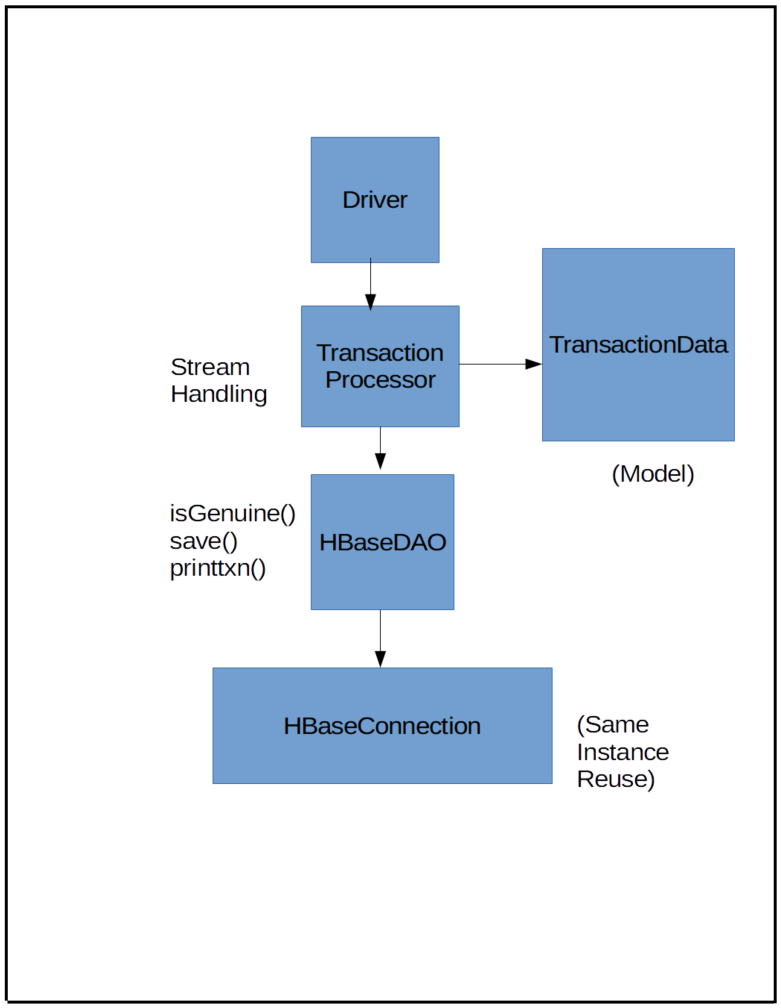
Genuineness Validation

a) **UCL**: The UCL defines the upper limit of the transaction. Hence if the transaction amount is greater than UCL for the card, the transaction is classified as fraud. Else it is genuine.
b) **Score**: Score is simply obtained from external system using batching layer and a simple check against a maximum value of 200 is done to classify if genuine of fraud.
c) **Last Transaction**: The Distance between the postcode of last transaction and currently validated transaction is calculated and compared against the time between these to transactions to know if it physically possible. If the speed required to travel between two locations exceeds maximum possible travel speed of 4km/sec, the transaction gets classified as fraud, else genuine.

Java Solution

Classes: -

- Driver: Entry point of the application. Subscribes to Kafka topic and streams card transaction data
- TransactionData: Models the transaction
- TransactionProcessor: Deserializes and Processes stream
- HbaseDAO: Uses connection to retrieve and save data
- HbaseConnection: Connects to Hbase database and return required HTable instance



Output

Maven Project & Jar: The source and compiled output

Console Output Log: The complete income transaction from topic is classified and printed on the console. The output is redirected to file and included in submission.

Final Card Transactions: After all the transactions are processed, the card_transactions table is exported to csv file and included in submission.