Financial Fraud Detection System

Introduction and Project Overview:

The objective of this project is to implement a system that can detect fraudulent activities in financial transactions using SQL queries. Fraudulent transactions often involve behaviors that deviate from normal patterns, such as unusually large withdrawals, multiple transactions in a short time span, or an abnormal increase in spending within a given time frame. Detecting such behaviors can help identify fraud early, potentially saving businesses and users from significant losses.

The project is driven by real world-inspired financial data with cardholders, credit cards, merchants and transactions information. SQL enables analysis of the datasets to carry out queries on these large sets and raises alerts based on fraud-prone transactions. In addition, alerts will be made automatically depending upon criteria defined while raising it towards authorities concerned.

Main Characteristics of the Project:

- Detection of Suspicious Transactions:
  The system will flag transactions that are over a specified monetary threshold, such as $500; these may represent unusually large withdrawals and may be part of a fraudulent transaction. These types of transactions are labeled as suspicious and require further investigation.
  Detecting Patterns of Frequent or Repeated Transactions
  Fraud often occurs in bursts where multiple transactions are made within a very short time window. This could be indicative of a fraudster trying to quickly deplete a victim's funds. This project looks for patterns of frequent transactions within a narrow time frame (e.g., multiple transactions within 30 minutes).

- Giving Alerts Regarding Unusual Spending Behavior:
  The system tracks the spending behavior of users over time, identifying sudden surges in transaction amounts or total spending. A sharp increase in spending may indicate fraud, especially if the behavior is inconsistent with the user's usual transaction history. Giving Alerts Regarding Unusual Spending Behavior.

- Efficient Data Management and Querying:
  To enable the detection of fraudulent activities, the project utilizes advanced SQL techniques including:

Stored Procedures: The flagging of suspicious transactions on predefined thresholds is automated.
Functions: Calculation of user spending patterns for comparison with normal behavior
Views: The creation of reusable views that aggregate fraud alerts and suspicious transactions.
Common Table Expressions (CTEs): This is used in simplifying complex queries, like identifying users who made suspicious withdrawals or those who have rapid transaction behaviors.
SQL Querying for Fraud Detection

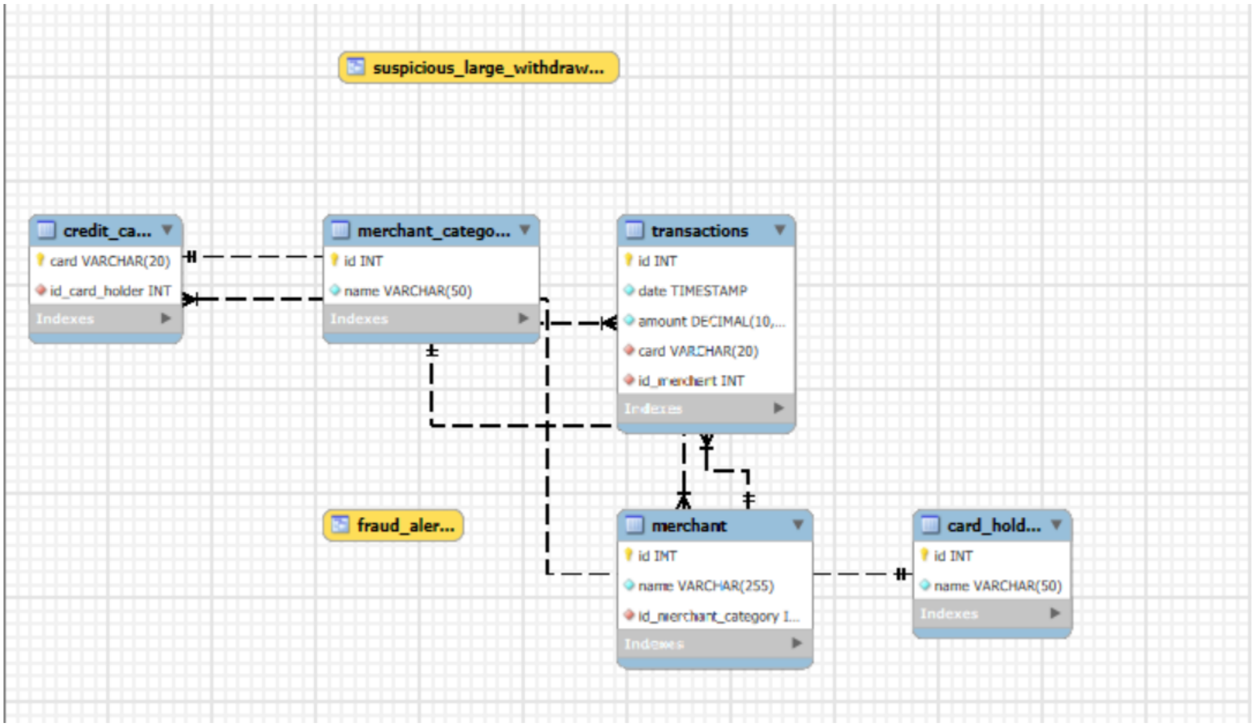
- Conditional logic for normal or suspicious transactions based on several thresholds.

Requirements and Problem understanding:

- Tables :
  The database schema consists of the following key tables:
  card_holder: Contains the details of the users (cardholders) including their name, contact details, and account balance.
  credit_card: Stores the credit card details linked to each cardholder.
  transaction: Holds the transaction details such as the amount, date, card used, merchant, etc.
  merchant: Contains the details of merchants where transactions occur.
  Link: Fraud_detection Project
- Schema :
  Below is the data model that outlines the relationships between the tables:
  card_holder (id, name, etc.)
  credit_card (card, id_card_holder, etc.)
  transaction (id, card, id_merchant, amount, date, etc.)
  merchant (id, name, etc.).
  Link: Fraud_detection Project

- Data Relationships :
  The relationships between the tables are as follows:
  The card_holder table is related to the credit_card table through the id_card_holder column.
  The credit_card table is linked to the transaction table via the card column.
  The transaction table connects to the merchant table using the id_merchant column.

- ER Diagram:



Data Generation Process:

The dataset used for this project was sourced from the Fraud Detection SQL GitHub repository, which provides a synthetic dataset designed for educational purposes in credit card fraud detection. The dataset is also available on Kaggle. Its generation involved creating fictional yet realistic data points to simulate real-world credit card transactions. The process includes randomizing transaction details, injecting anomalies to mimic fraudulent behavior, and ensuring logical dependencies across different data tables. This synthetic data allows for effective training and testing of fraud detection models.

Link: Fraud_detection Project

Scenarios or Business requirements (SQL Queries):

Following is the link for Business requirements and the corresponding SQL Queries.

Link: Fraud_detection Project

Conclusion:

We employed the data in the Fraud Detection SQL GitHub repository to construct an entire Financial Fraud Detection System. Here we built a database and carried out numerous SQL queries that aimed to scrutinize financial transactions, activities, accounts, and details related to merchants. Such queries went as simple as the list of active users with their balances and could go all the way up to more complex transactions to find suspect activity or excessive transfer activity. Further, we used window functions, CTEs, views, stored procedures, and functions to optimize the detection process. This was a comprehensive analysis framework that simulated real-world applications of financial fraud detection and, therefore, provided important insights into data management and anomaly detection.

Team Members:

1. Vaishnavi Shinde
2. Vishakha Shinde
3. Kaushal Borkar
4. Rajeev Aken
5. Suyash Dixit