P. Hima Varshini
CCE-23031

Data Base Management System

Preliminary Report: Fraud Detection in Banking Transactions Using SQL

---

1. Introduction

With the rise of online banking and electronic transactions, fraudulent activities have become a growing concern in the financial sector. Fraudsters constantly devise new techniques to bypass traditional security measures, leading to increased financial losses and decreased trust in the banking system. This makes it imperative to develop automated and effective fraud detection systems that can identify suspicious activities in real-time.

Traditional fraud detection systems often rely on static, rule-based approaches, which may not be sufficient to address the dynamic and evolving nature of fraud. This project aims to develop an SQL-based system for detecting fraudulent banking transactions, leveraging relational databases to analyze patterns and identify anomalies in transactional data.

2. Requirements Definition

1. **User Management**:
    - Store user details such as ID, account number, email, phone number, and normal locations.
2. **Transaction Management**:
    - Record transactions with attributes: ID, amount, timestamp, location, and device/IP.
    - Allow querying based on transaction data (amount, location, etc.).
3. **Fraud Detection Logic**:
    - Flag transactions exceeding predefined thresholds (e.g., high-value, high-frequency).
    - Identify location mismatches and transactions from unknown devices/IPs.
    - Log reasons for flagged transactions (e.g., high transaction amount, suspicious location).
4. **Data Storage**:
    - Use a relational database (SQL) for transaction and flag data.
    - Store flagged transactions with reasons in a dedicated flags table.

ER Diagram and Relationship Tables

Entity-Relationship Details:

1. **Users Table**:
    - **Attributes**: UserID (PK), Name, AccountNumber, Email, NormalLocations (multivalued), PhoneNumber
    - **Relationship**: One-to-many with Transactions (A user can have many transactions).

2. **Transactions Table**:
    - **Attributes**: TransactionID (PK), UserID (FK), Amount, Timestamp, Location, DeviceIP
    - **Relationship**: Many-to-one with Users (A transaction belongs to one user) and one-to-one with Flags (A transaction may have one flagged entry).
3. **Flags Table**:
    - **Attributes**: FlagID (PK), TransactionID (FK), Reason, FlaggedTimestamp
    - **Relationship**: One-to-one with Transactions (Each transaction can have one flagged entry).

Table	Attribute	Description	Relationship
Users	UserID (PK)	Unique identifier for each user.	One-to-Many with Transactions (UserID)
	Name	Full name of the user.	
	AccountNumber	User's bank account number.	
	Email	User's email address.	
	NormalLocations	Common geographic locations where the user typically transacts.	
	PhoneNumber	User's contact number.	
Transactions	TransactionID (PK)	Unique identifier for each transaction.	One-to-One with Flags (TransactionID)
	UserID (FK)	Foreign key referencing UserID in the **Users** table.	One-to-Many with Users (UserID)
	Amount	Monetary value of the transaction.	
	Timestamp	Date and time of the transaction.	

Table	Attribute	Description	Relationship
	Location	Geographic location of the transaction.	
	DeviceIP	Device or IP address used for the transaction.	
Flags	FlagID (PK)	Unique identifier for each flagged entry.	
	TransactionID (FK)	Foreign key referencing TransactionID in the **Transactions** table.	One-to-One with Transactions (TransactionID)
	Reason	Explanation for why the transaction was flagged as suspicious.	
	FlaggedTimestamp	Date and time when the transaction was flagged.	

Fraud Detection Rules:

1. **High-Value Transactions**: Flag transactions where the amount exceeds a predefined threshold (e.g., 100,000).
2. **High-Frequency Transactions**: Identify users with multiple transactions within a short time period (e.g., more than 5 transactions in 10 minutes).
3. **Location Mismatch**: Flag transactions originating from locations outside the user's normal location.
4. **Device/IP Anomalies**: Detect transactions made from unknown devices or IP addresses.

Planned Results:

The expected outcomes of the fraud detection system include:

- **High-Value Transactions**: Identification of transactions that exceed a predefined threshold (e.g., 100,000), as these are typically high-risk and require scrutiny.
- **High-Frequency Transactions**: Detection of multiple transactions in a short period from the same user, which could indicate fraudulent behavior or account takeover.
- **Location Mismatches**: Transactions originating from locations that deviate from the user's normal location, which can be a strong indicator of fraudulent activity.

These detection patterns aim to flag suspicious transactions for further investigation, helping banks mitigate potential fraud.

Anticipated Challenges and Limitations

- **False Positives**: Overly strict rules may flag legitimate transactions as suspicious.
- **Scalability**: Handling large transaction volumes efficiently might pose challenges.
- **Data Limitations**:
    - The self-created dataset may not represent real-world complexities.
    - Simplified patterns or lack of diversity in data could impact fraud detection accuracy.
- **Validation**: Testing and validating the system thoroughly without real-world data may limit insights.