**High-Level Design (HLD)**

**Objective**

To design a robust system capable of detecting fraudulent credit card transactions in real-time, minimizing false positives, and ensuring high accuracy.

### **Components**

1. **Data Sources**

* + Transaction Data: Real-time transaction data including amount, merchant, location, time, etc.
  + Historical Data: Past transaction data for training and validating models.
  + External Data: Additional data sources such as geolocation, IP address, device information.

2. **Data Ingestion Layer**

* APIs to collect real-time transaction data.
* ETL (Extract, Transform, Load) processes for historical data.

3. **Data Storage**

* Relational Databases (e.g., PostgreSQL) for structured data.
* Data Lakes (e.g., Amazon S3) for large volumes of historical data.

4. **Feature Engineering**

* Feature extraction from raw data.
* Creation of derived features such as transaction frequency, average spend, location patterns, etc.

5. **Machine Learning Models**

* Supervised learning models (e.g., Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, Neural Networks).
* Anomaly detection models (e.g., Autoencoders, Isolation Forests).

6. **Model Training and Evaluation**

* Training pipeline for model development.
* Cross-validation and performance metrics (precision, recall, F1-score, ROC-AUC).