AI-Powered Credit Card Fraud Detection: Data Implementation

# 1. Data Pipeline Architecture

Data Sources:  
- Transactional Data (real-time stream)  
- External Data (known fraud databases, blacklisted merchants, regional risk scores)  
  
Data Ingestion:  
- Kafka / AWS Kinesis for real-time streaming transactions  
- Batch ingestion via S3 / Data Lake for historical training data

# 2. Data Storage

Raw Data Storage:  
- Amazon S3 / Azure Blob Storage for logs and transaction history  
  
Processed Data Storage:  
- Amazon Redshift / Azure Synapse for analytics-ready data  
- DynamoDB / MongoDB for low-latency transaction lookup  
  
Feature Store:  
- AWS SageMaker Feature Store or Feast

# 3. Data Processing & Feature Engineering

Streaming Data Processing:  
- Apache Flink / Spark Streaming for real-time feature extraction  
  
Key Features:  
- Transaction velocity, average transaction amount, location deviation, device/IP reputation scores, time since last transaction, merchant category risk score  
  
Batch Processing:  
- Historical aggregations for model retraining  
- ETL pipelines via Apache Airflow / AWS Glue

# 4. AI Model Design

Model Type:  
- Binary classification (Fraud / Legitimate)  
  
Possible Algorithms:  
- XGBoost / LightGBM, Deep Neural Networks, Isolation Forest / Autoencoder  
  
Model Features:  
- Real-time transaction metadata, aggregated customer behaviors, historical fraud patterns, external threat intelligence indicators

# 5. Real-Time Fraud Detection Service

Inference API:  
- Hosted on AWS SageMaker Endpoint / Azure ML Service  
  
Transaction Guard:  
- Flag or approve transactions based on fraud probability threshold  
  
Decision Logging:  
- Store predictions and actions taken for auditing and retraining

# 6. Model Monitoring and Retraining

Monitoring Metrics:  
- Prediction drift, data drift, fraud detection recall/precision, latency, throughput  
  
Model Retraining Workflow:  
- Scheduled batch retraining  
- Deploy new model via CI/CD pipeline

# API Spec Example

POST /predict\_fraud  
  
Request Body:  
{  
 "card\_token": "abcd1234",  
 "amount": 250.00,  
 "timestamp": "2025-05-20T10:30:00Z",  
 "merchant\_id": "store5678",  
 "device\_fingerprint": "xyz987",  
 "ip\_address": "192.168.0.1",  
 "geo\_location": "37.7749,-122.4194"  
}  
  
Response:  
{  
 "fraud\_probability": 0.87,  
 "action": "flagged"  
}