GAOUI Yanis, HUMEAU Gauthier, BOIZET Hugo, DESMONS Damien, PITRAT Clement

REAL-TIME FRAUD DETECTION

THE PROBLEM

- Card fraud losses : $32 billion (2023)
- Thousands of transactions per second
- Detect BEFORE completion (< 5 sec)
- Minimize false positives

LAMBDA ARCHITECTURE

Speed Layer

- Spark Streaming
- Real-time
- 2-3 sec latency

Batch Layer

- Spark Batch
- Historical data
- 100% accuracy

TECHNOLOGY STACK

- **Kafka** : High-performance message broker
- **Apache Spark** : Batch + Streaming processing
- **MongoDB** : Serving layer database
- **Kubernetes** : Orchestration + auto-scaling

FRAUD SCORING ALGORITHM

Score =

0.25×Country + 0.20×Category + 0.20×Amount + 0.10×Online + 0.10×Hour + 0.15×Velocity

Alert threshold: Score ⩾ 0.6

High-risk countries : 0.7-0.8

Crypto transactions : 0.85-0.9

KEY RESULTS

2-3 sec

Latency

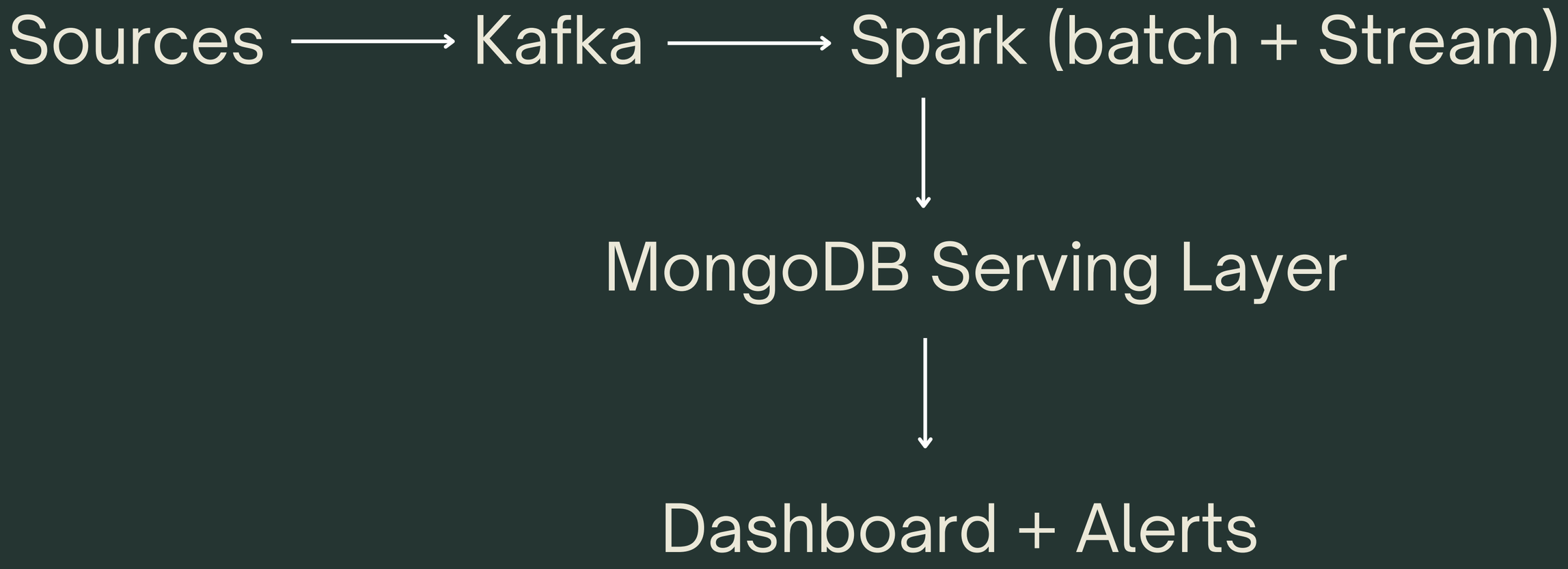
95%

Accuracy

10 TPS

Throughput

DATA FLOW


Sources
Kafka
Spark (batch + Stream)
MongoDB Serving Layer
Dashboard + Alerts

REAL-TIME DASHBOARD

- **4 KPIs** : Total transactions, Fraud count, Fraud rate
- Real-time chart (transactions/hour)
- World map (fraud by country)
- Breakdown by merchant category
- Fraud alert history

WHY LAMBDA?

- ML retraining on complete historical data
- Regulatory compliance (audit trail)
- Complex multi-day fraud patterns
- Batch accuracy surpasses stream approximations

FUTURE IMPROVEMENTS

- Cloud deployment (AWS/GCP)
- Advanced ML models (XGBoost, LSTM)
- Feature Store (Feast/Tecton)
- A/B testing for model deployment

THANKS FOR YOUR ATTENTION

Any questions about our Project ?