

# E-Commerce Analytics Platform

Real-Time Big Data Processing with Hadoop & Grafana

# The Challenge



## Massive Data Volume

Processing millions of orders, customer profiles, and clickstream events daily



## Real-Time Insights

Marketing and ops teams need insights in minutes, not days



## Cost Efficiency

Traditional BI tools cost \$100K+ annually



## Scalability

Must handle 10x traffic during Black Friday / holiday peaks

# Solution Architecture



## Technology Stack

# E-Commerce Data Sources

## Orders (50,000 records)

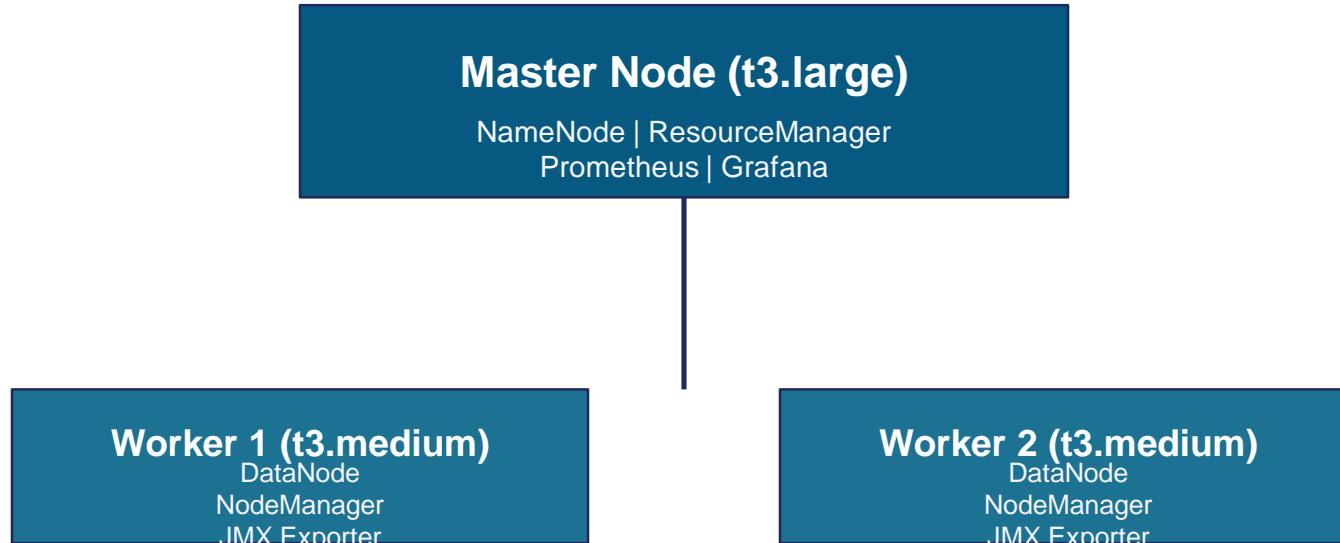
- Customer purchases
- Payment processing
- Multi-region sales
- Product performance
- Revenue tracking

## Customers (5,000 profiles)

- User demographics
- Lifetime value
- Tier segmentation
- Purchase history
- Behavior patterns

## Clickstream (200,000 events)

# Live Cluster Architecture



## Cluster Capacity

- Total: 16GB RAM, 6 vCPUs, 90GB Storage

# Real-Time Monitoring Dashboard

## Business Metrics

Total Orders Storage

Data Ingestion Rate

Analytics Jobs Running

Processing Capacity

## HDFS Metrics

Storage Used/Remaining

Live DataNodes

Block Replication Health

NameNode Performance

## YARN Metrics

Memory Allocation

vCore Utilization

Active NodeManagers

Job Queue Status

# E-Commerce Use Cases



## Sales Analytics

Revenue by region, product performance, trend analysis



## Customer Segmentation

Identify high-value customers, personalized marketing



## Conversion Optimization

Funnel analysis, A/B testing, cart abandonment



## Inventory Forecasting

Demand prediction, stock optimization, auto-reorder



## Fraud Detection

Real-time anomaly detection, payment security



## Recommendation Engine

Product suggestions, personalized experience

# Live Demo Scenarios

1

## Sales Performance by Region

? Which regions are performing best?

→ Run MapReduce job on 50K orders → Get revenue breakdown in 30 seconds

2

## Customer Segmentation Analysis

? How many high-value customers?

→ Analyze 5K customer profiles → Platinum/Gold/Silver/Bronze distribution

3

## Real-Time Order Processing

? Can we handle Black Friday load?

→ Monitor cluster under simulated load → Watch metrics update in Grafana

# Business Value & ROI

**\$137**

**Monthly Cost**

vs \$2-5K for managed services

**Minute**  
**S**  
Time to Insight

vs days with manual analysis

**10x**

**Scalability**

Handle peak loads easily

**\$300K+**

**Annual Savings**

Automation + efficiency

## Key Benefits

- ✓ Real-time insights for data-driven decisions
- ✓ Scale effortlessly during peak seasons
- ✓ Reduce infrastructure costs by 70-80%
- ✓ Eliminate manual analysis bottlenecks

# Ready to Transform Your E-Commerce Analytics?

## Next Steps:

1. Assess your current data volumes and requirements
2. Define success metrics and KPIs
3. Plan production environment deployment
4. Set up data ingestion pipelines