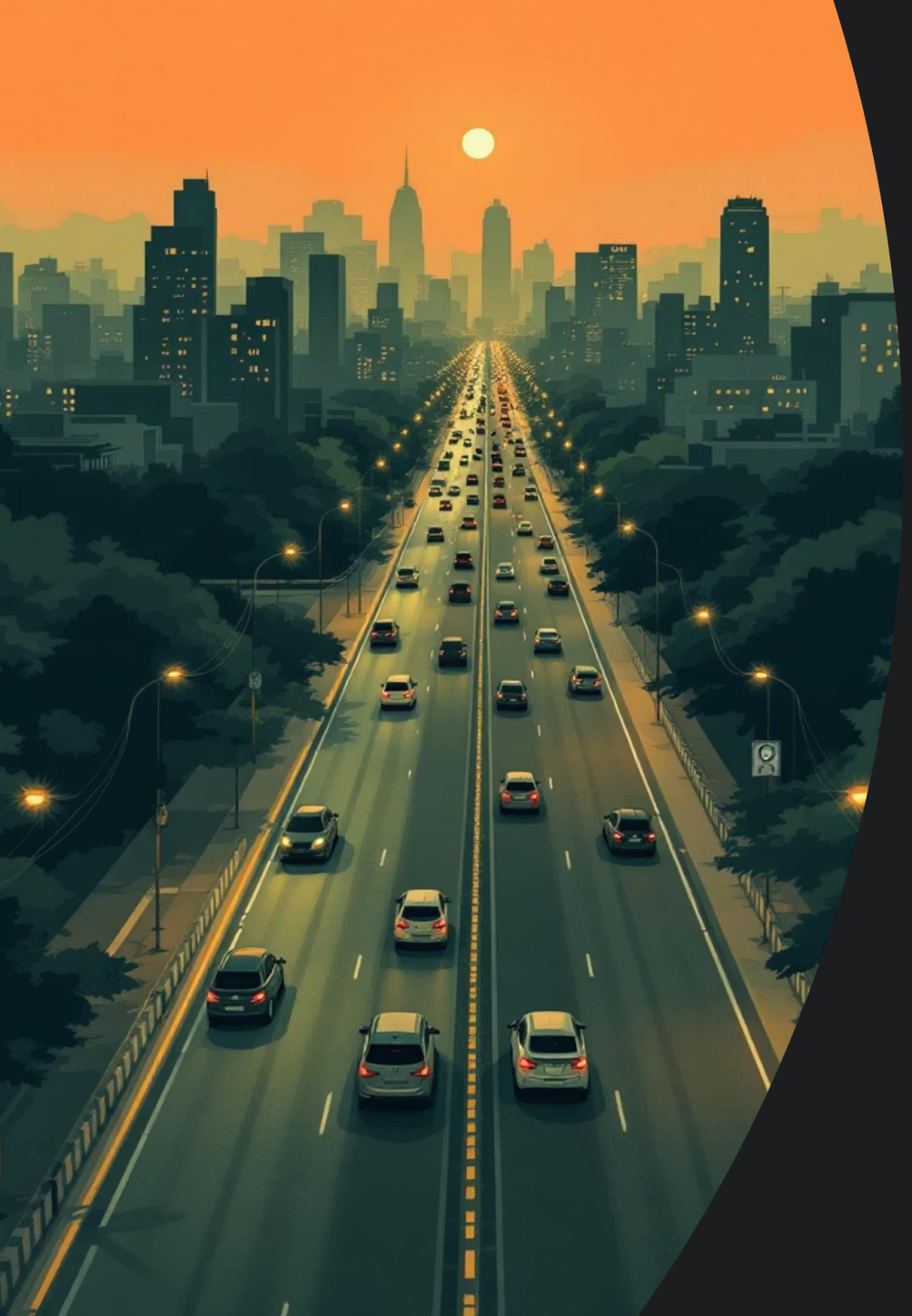
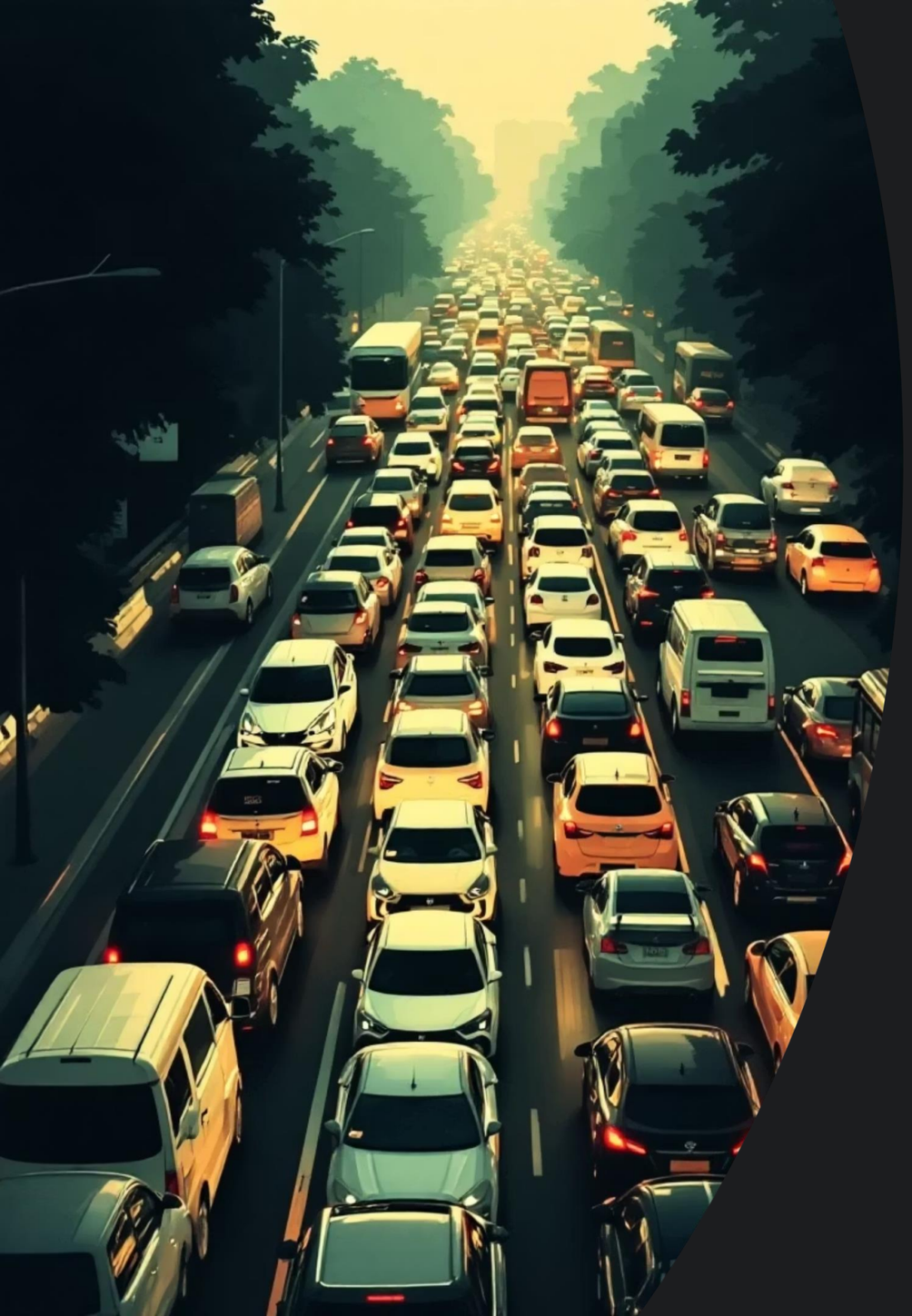


Urban Traffic Congestion Monitoring

A Data Engineering Solution for Smarter Cities

TYAIEC-2A5



The Urban Traffic Challenge in India

India's major urban centres face severe congestion, with average vehicle speeds dropping below 20 km/h during peak hours in cities like Mumbai and Delhi. This congestion results in substantial economic losses estimated at billions annually through wasted fuel, lost productivity, and time delays.

Economic Impact

Billions in annual losses from wasted fuel and lost productivity

Environmental Cost

Increased pollution and respiratory diseases linked to traffic jams

Infrastructure Gap

Need for scalable, cost-effective real-time monitoring solutions

Why Traditional Traffic Monitoring Falls Short

High Infrastructure Costs

Physical sensors like induction loops, CCTV systems, and radar require substantial capital investment and ongoing maintenance expenses that strain city budgets.

Limited Geographic Coverage

Traditional networks cannot scale across rapidly expanding urban areas and informal settlements, leaving large zones without monitoring capability.

Data Silos & Latency Issues

Fragmented systems lack integration, creating delays that prevent timely traffic management decisions when congestion occurs.

Current Reality: Most Indian cities lack integrated, real-time traffic data platforms connecting all monitoring sources.

Introducing CityPulse

A Modern Data Engineering Approach

CityPulse is an open-source, containerized pipeline that simulates real-time urban traffic data without requiring physical sensors. This innovative approach eliminates infrastructure costs whilst providing comprehensive, scalable traffic monitoring for Indian cities.

Apache Kafka

Real-time data ingestion

Spark Streaming

Live data processing

Docker

Containerized deployment

ML Models

Predictive analytics

System Architecture Overview

Data Ingestion & Processing

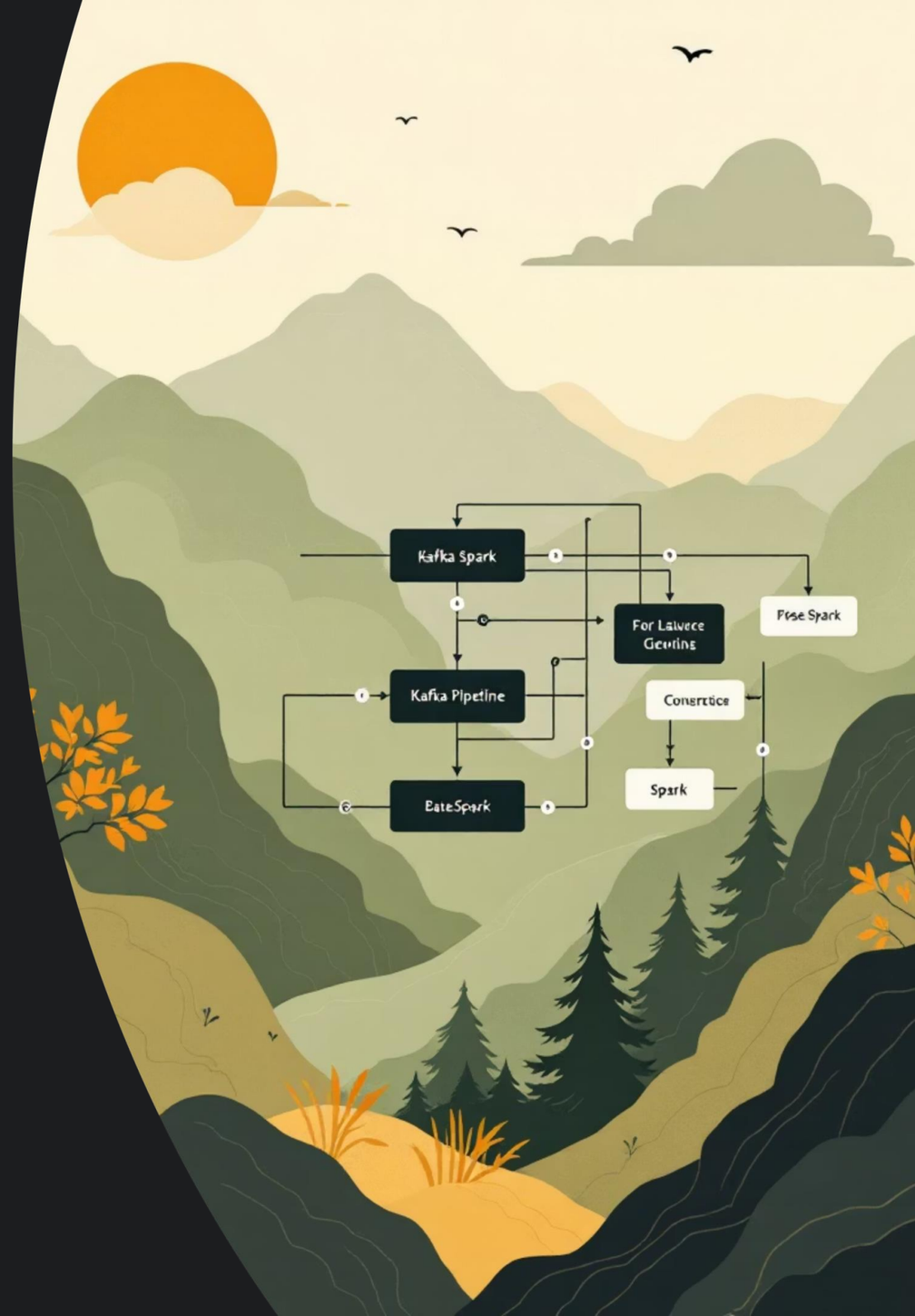
A Kafka cluster ingests over 11 million traffic records daily, streaming vehicle telemetry, GPS coordinates, and weather conditions in real-time.

Apache Spark Structured Streaming performs live transformations and detects anomalies as data arrives.

Storage & Intelligence

Temporary buffers optimise writes to a centralized data warehouse, ensuring efficient data persistence.

Machine learning modules predict congestion patterns from refined data, whilst Flask backend with React frontend powers interactive dashboards.



Data Sources & Pipeline Details

01

Synthetic Data Generation

Simulates real-world urban traffic signals including vehicle counts, speeds, GPS coordinates, and dynamic weather conditions across multiple zones.

02

Historical Integration

Incorporates government traffic volume data from existing sensors, creating a hybrid approach that blends synthetic and real-world sources.

03

Real-Time Comparison

Enables continuous comparison of current traffic patterns against historical averages, identifying deviations and emerging congestion.

04

Geospatial Processing

PostgreSQL with PostGIS extensions store location data, enabling location-based queries and interactive mapping of traffic patterns across the city.

Machine Learning for Congestion Prediction

CityPulse employs lightweight supervised learning models trained on 177 carefully engineered features including time-of-day, location, road type, weather, and recent traffic patterns. These models enable proactive traffic management rather than reactive responses.

Volume Prediction

Forecasts short-term traffic volume changes

Anomaly Detection

Identifies unusual patterns indicating congestion build-up

Route Guidance

Suggests alternative routes to reduce congestion

Proactive Signals

Enables adaptive traffic signal control optimisation

Technical Approach: Linear regression with Spark MLlib optimised for streaming data, ensuring minimal latency in predictions.

Real-World Impact & Use Cases

1 Hotspot Identification

City planners identify chronic congestion points and optimise traffic flow dynamically using data-driven insights rather than assumptions.

2 Demand Management

Supports adaptive tolling, carpool incentives, and congestion pricing during peak hours to reduce vehicle volumes and smooth traffic.

3 Emergency Response

Predicts and mitigates gridlocks, enabling faster emergency vehicle routing and reducing response times for critical incidents.

4 Environmental Benefits

Smoothing traffic flow reduces pollution emissions and improves urban air quality, supporting public health and sustainability goals.

Challenges & Future Directions

Current Challenges

- Integrating heterogeneous data sources including cameras, GPS, weather APIs, and event calendars
- Scaling architecture for megacities with millions of vehicles and complex, multi-layered road networks
- Detecting rare anomalous events like accidents and roadworks with AI-driven systems

Future Expansion

- Multi-modal transport integration: buses, metro systems, auto-rickshaws, and pedestrian flows
- Holistic urban mobility insights combining all transportation modes
- Advanced anomaly detection for unusual patterns and emergency scenarios
- Deeper predictive capabilities for long-term urban planning

Towards Smarter, Safer Indian Cities

Data-Driven Foundation

Urban traffic congestion monitoring through advanced data engineering is critical for sustainable urban growth and liveable cities.

Proven Solution

CityPulse demonstrates a scalable, cost-effective model specifically designed for India's unique urban challenges and resource constraints.

Actionable Insights

Combining real-time data pipelines with machine learning unlocks actionable traffic management insights for city authorities.

Call to Action

Invest in data-driven urban mobility solutions today to transform India's cities for tomorrow and beyond.