# Real-Time Traffic Congestion Analytics Pipeline

## 1. Introduction

This project aims to design and implement a real-time analytics pipeline for detecting and visualizing traffic congestion using big data technologies. By integrating streaming data sources (such as IoT sensors and GPS data), the system provides live insights into traffic flow and congestion patterns.

## 2. Objectives

• Collect and process real-time traffic data from multiple sources.

• Detect congestion patterns using distributed processing.

• Visualize congestion trends on an interactive dashboard.

• Support decision-making for urban traffic management.

## 3. System Architecture

The proposed architecture consists of the following components:  
1. Data Ingestion Layer: Uses Apache Kafka to stream data from sensors, GPS devices, and traffic cameras.  
2. Processing Layer: Utilizes Apache Spark Streaming for real-time analytics and aggregation.  
3. Storage Layer: Stores processed data in HDFS or Cassandra for scalability.  
4. Visualization Layer: Uses Grafana or Tableau to display real-time congestion heatmaps.

## 4. Data Flow

1. Traffic data from IoT devices and vehicles are published to Kafka topics.  
2. Spark Streaming consumes Kafka data, processes it in micro-batches, and applies congestion detection models.  
3. Processed data is stored in HDFS or Cassandra for querying.  
4. Visualization tools connect to the storage to display live analytics dashboards.

## 5. Technologies Used

• Apache Kafka – for data ingestion and stream handling.

• Apache Spark Streaming – for real-time processing and analytics.

• HDFS / Cassandra – for distributed data storage.

• Grafana / Tableau – for real-time visualization.

• Python / Scala – for implementation and data transformation.

## 6. Expected Outcomes

• A working real-time data pipeline that can detect and visualize traffic congestion.  
• Improved understanding of how distributed systems handle streaming data.  
• A scalable architecture applicable to smart city traffic management systems.

## 7. Conclusion

The Real-Time Traffic Congestion Analytics Pipeline demonstrates how big data technologies can be integrated to create intelligent and responsive systems. This project serves as a foundation for future smart city applications that rely on real-time data analytics.