Real-Time Traffic Analysis and Prediction

DATA SEEKERS

March 2025

1 Project Idea

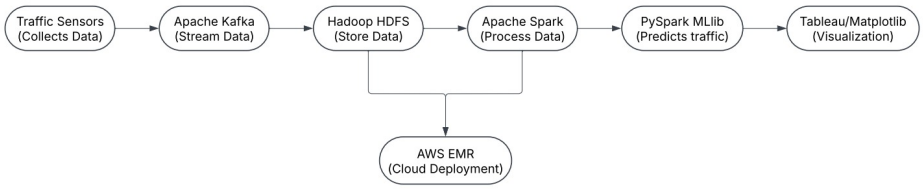
The objective of this project is to develop a real-time traffic monitoring and prediction system using big data technologies. The system will analyze historical and live traffic data from sensors placed in different regions to predict congestion patterns and optimize traffic flow management. This will be accomplished using large-scale data processing frameworks and machine learning models to improve urban transportation systems and reduce travel time.

2 Tools and Technologies

The following tools and technologies will be utilized in the project:

- **Data Storage**: Apache Hadoop (HDFS) for distributed data storage.
- **Streaming**: Apache Kafka for real-time traffic data ingestion.
- **Processing**: Apache Spark and PySpark for big data processing.
- **Machine Learning**: PySpark MLlib for traffic flow predictions.
- **Visualization**: Tableau and Matplotlib for graphical representation.
- **Cloud Services**: AWS EMR (Elastic MapReduce) for deployment.

3 High-Level Architecture / Methodology



4 Explanation of the Data Flow Diagram

- **Step 1**: Traffic sensor data is collected from various regions and ingested into the system via Apache Kafka.
- **Step 2**: Kafka streams the real-time traffic data to Apache Spark for processing.
- **Step 3**: Data is stored in Hadoop Distributed File System (HDFS) for historical analysis.
- **Step 4**: PySpark MLlib applies machine learning models to predict future traffic trends based on historical and real-time data.
- **Step 5**: The processed data is sent to visualization tools like Tableau for monitoring and insights.

- **Step 6**: The system provides predictive traffic insights to help optimize urban transportation planning.
- **Step 7**: AWS EMR is used to handle the scalability of Hadoop and Spark processing.

5 Project Goals

- Develop an efficient pipeline to process real-time traffic data.
- Implement machine learning models to predict congestion levels.
- Optimize traffic flow management based on predictive insights.
- Improve urban mobility through data-driven decision-making.
- Deploy the system on a scalable cloud platform (AWS EMR).