IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, November 2025
Impact Factor: 7.67

A Scalable Big Data Framework for Real Time Traffic Monitoring System

Sanskruti Walke[1], Lavanya Yenguptla[2], Shrdhaa Wadkar[3], Prof. Medhe. S. S[4]
Degree Students, Department of Computer Engineering[1,2,3]
Faculty, Department of Computer Engineering, Chas, Ahamadnagar, India
Adsul Technical Campus, Chas, Ahmadnagar, India

Abstract: *Traditional traffic management systems struggle to handle the massive volume and high velocity of data generated by modern urban environments, including GPS sensors, traffic control devices, and car embedded systems. This paper introduces a scalable and real-time intelligent transportation system based on a big data framework to address these challenges. The system is architected to process large-scale stream data continuously, enabling the real-time detection of critical traffic events such as incidents, crashes, and congestion.*

The proposed architecture is built on a distributed computing platform, leveraging components from the Hadoop ecosystem like the Hadoop Distributed File System (HDFS) for scalable storage and MapReduce for processing large historical datasets. The real-time stream processing layer, which can be implemented using technologies such as Apache Kafka or IBM InfoSphere Streams, allows for immediate data analysis and rapid response to dynamic traffic conditions. This layer incorporates operators for filtering, correlating, and aggregating data streams to identify abnormal events and traffic jams based on metrics like vehicle speed and density.

Experimental results demonstrate the system's ability to gather and manage large-scale data from numerous sensors efficiently and with high fault tolerance, processing hundreds of thousands of tuples per second while consuming minimal hardware resources. The framework's capacity to provide instant, high-precision information to end-users (via web consoles and smartphones) facilitates effective decision-making for traffic management and urban planning. The system effectively satisfies the 4V criteria of big data, offering a robust solution for enhancing urban mobility and road safety.

Keywords: Traditional traffic management systems

I. INTRODUCTION

Implement Data Quality and Veracity Checks: Integrate automated data validation and cleaning steps at the ingestion layer to ensure the accuracy and reliability of the traffic information.

Implement Robust Security Measures: Use encryption for data at rest and in transit, enforce strict access controls, and perform regular security monitoring.

Monitor and Optimize Continuously: Track key performance indicators (KPIs) like throughput and latency, and use automated monitoring tools to identify bottlenecks and optimize resource allocation.

II. NEED FOR SYSTEM

Handle big data challenges: The framework must address the "4 Vs" of big data:

Volume: Processing the massive amount of data generated by countless sensors, cameras, GPSdevices.

Velocity: Analyzing data in real-time as it streams in to enable immediate action

Variety: Integrating and processing data from many different sources, such as road sensors, weather stations, and mobile devices.

Veracity: Ensuring the data's quality and reliability for accurate analysis.

Support real-time decision-making: A scalable system enables dynamic traffic management, such as adjusting signal timings, diverting traffic during incidents, and providing drivers with real-time route information to prevent jams.

Ensure high availability and fault tolerance: The system must be resilient and continuously available, even if some components fail. A scalable framework can be designed for high availability, ensuring that traffic monitoring and management are never interrupted.

Enable future growth: Investing in a scalable system avoids the financial burden of replacing entire systems as traffic volumes increase. It allows for easier expansion and addition of new functionalities or data sources.

Optimize resource utilization: By using scalable cloud-based or distributed systems, cities can use computing resources more efficiently, which can lead to cost savings.

Provide insights for improvement: A big data framework allows for the storage and analysis of historical data to identify long-term trends, optimize infrastructure, and improve urban mobility over time.

III. SCOPE OF SYSTEM

The scope of work for a scalable big data framework for a real-time traffic monitoring system includes data ingestion, stream processing, data storage, big data analytics, and user visualization. The primary objective is to collect large-scale, real-time data from diverse sources to effectively monitor traffic flow, detect incidents (crashes, congestion), and provide actionable insights for improved traffic management.

1. Data Ingestion & Sources
2. Stream Processing & Analysis
3. Data Storage & Management
4. Analytics & Insight Generation
5. User Interface & Application Layer

The scope extends to achieving several broad objectives:

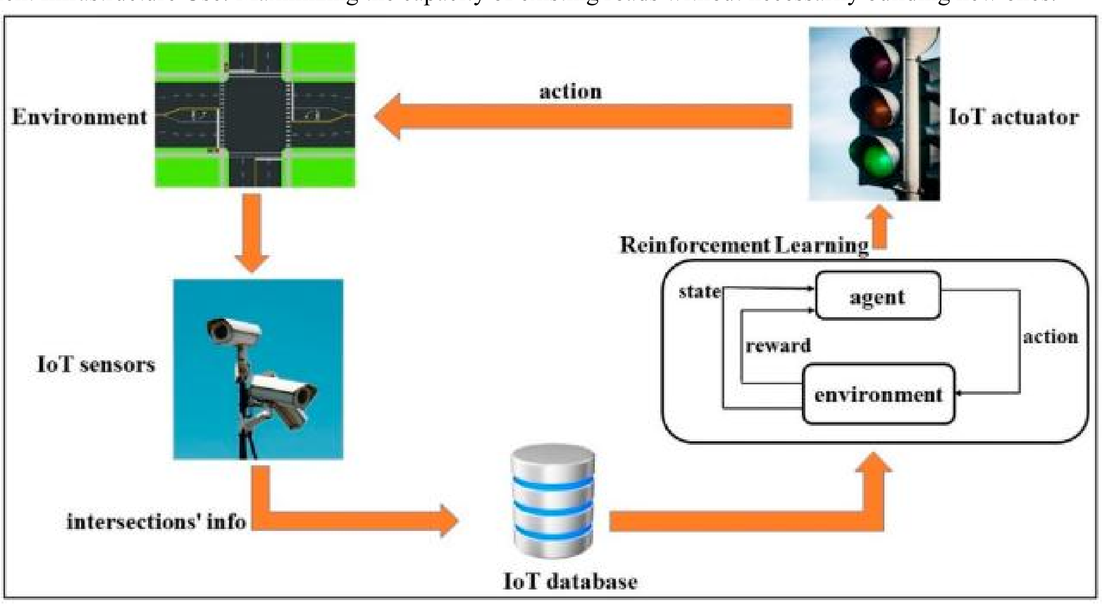
Improved Traffic Flow: Minimizing congestion and reducing overall travel times.

Enhanced Safety: Reducing the number of accidents by alerting drivers to hazards and managing speed limits.

Prioritization of Emergency Vehicles: Providing immediate clearance for ambulances, police, and fire engines to improve response times.

Environmental Impact Reduction: Decreasing fuel consumption and vehicle emissions by reducing idle time and optimizing routes.

Efficient Infrastructure Use: Maximizing the capacity of existing roads without necessarily building new ones.



IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 3, November 2025

Impact Factor: 7.67

IV. CONCLUSION

A scalable big data framework for real-time traffic monitoring successfully uses modern technologies like Apache Kafka, Spark, and machine learning to improve urban mobility and safety. The key conclusions are that these systems can process massive data streams efficiently, provide real-time actionable insights, and are more cost-effective and adaptable than traditional methods.

REFERENCES

[1]. Data Ingestion: Apache Kafka is the preferred solution for high-throughput, fault-tolerant message queuing.

[2]. Stream Processing: Apache link and Apache Storm are commonly used for low-latency, real-time analytics.

[3]. Batch Processing: Apache Spark and Hadoop Map Reduce are used for processing large volumes of historical traffic data.

DOI: 10.48175/568