Big Data In Transportation - Annotated Bibliography

Team: Artjola Meli, Shradha Godse

- *Choi Dojin, Song Seokil, Kim Bosung, Bae Insu, (2015) 'Processing Moving Objects and Traffic Events based on Spark Streaming'*

The authors propose an advanced method to manage and analyze real-time data streams of vehicle locations and traffic events using Spark Streaming. The core objective is to enable real-time sharing of information about vehicles, pedestrians, and traffic events through the WEVING service, a social driver assistance system. This system employs a grid indexing technique to organize data by time and location, facilitating efficient continuous range queries (CRQ). Traditional single-server models are inadequate for handling such volumes, especially in real-time applications. The proposed method involves distributing data processing across multiple servers using Spark and Spark Streaming, leveraging their in-memory capabilities to enhance speed and fault tolerance. The system indexes vehicle locations and traffic events dynamically, grouping similar traffic events to avoid redundancy and ensure meaningful data sharing.

Experiments conducted using a traffic dataset modified to simulate real-world conditions validate the method's efficiency. The system, implemented on a Spark Streaming-based in-memory distributed processing platform, successfully grouped 100,000 records of event data into 735 distinct events, demonstrating its capability to handle high data volumes and complex queries effectively.

- *Lian Yangi, Zhang Guoqing Lee Jaeyoung, Huang Helai, (2020),'Review on big data applications in safety research of intelligent transportation systems and connected/automated vehicles'*

The paper provides a comprehensive review of how Big Data is leveraged to enhance traffic safety within Intelligent Transportation Systems (ITS) and Connected/Automated Vehicles (CAV). The study explores various applications, including crash detection and prediction, identification of crash hotspots, driving behavior analysis, and the discovery of factors contributing to crashes. The paper highlights the potential of Big Data analytics to overcome traditional data limitations, offering deeper insights and more proactive traffic safety measures. The authors also discuss the challenges and future directions for Big Data safety analytics, emphasizing the need for systematic approaches to handle the complexity and volume of data in this field.

- *Kumar M., Singh C., (2017), 'Building Data Streaming Applications with Apache Kafka', Packt Publishing, Limited*

The book serves as an in-depth guide for designing and deploying enterprise-grade streaming applications using Apache Kafka. It begins with a detailed introduction to messaging systems and the principles behind them, then delves into Kafka's architecture, including its producers, consumers, and integration with various big data tools like Apache Spark and Apache Storm. The book also covers advanced topics such as capacity planning, security, and best practices for using Kafka in ETL pipelines and big data applications, making it a comprehensive resource for both beginners and experienced practitioners in the field of data streaming.

- *Faisal Rafiq Khan, (2021) 'Apache kafka with real-time data streaming'*

The paper provides an in-depth examination of Apache Kafka, a leading framework for real-time data streaming in Big Data analytics. It highlights Kafka's architecture and its role in establishing scalable, durable, and low-latency data stream pipelines between systems or applications. Kafka's distributed platform efficiently

processes large volumes of data with high throughput, making it ideal for organizations requiring reliable and swift data handling. The authors discuss Kafka's ability to run as a cluster on multiple servers, with streams of records categorized into topics, and the role of brokers and ZooKeeper in managing and coordinating data flow within the Kafka cluster.

The paper also explores Kafka's application in real-time streaming, emphasizing its fault-tolerant, durable data storage and low-latency processing capabilities. Kafka's versatility and integration with tools like Apache Storm, HBase, and Spark enable it to handle billions of messages per day, making it a robust choice for enterprises. The discussion includes Kafka's deployment in notable companies like LinkedIn, Netflix, Yahoo, and Uber, demonstrating its effectiveness in supporting mission-critical, real-time applications. The conclusion underscores Kafka's significance in modern data architectures, highlighting its reliability, scalability, and integration capabilities for handling vast amounts of streaming data.

- *Ahmed Hassebo, Mohamed Tealab, (2023) 'Global Models of Smart Cities and Potential IoT Applications: A Review'*

The paper offers a comprehensive evaluation of smart city models and the integration of Internet of Things (IoT) technologies in urban settings. As cities face increasing urbanization, the deployment of IoT applications becomes essential in addressing urban challenges and enhancing sustainability. The paper proposes eight global models for smart cities, providing frameworks for city planners to design and implement IoT solutions effectively. These models cover various sectors, including transportation, energy management, waste management, public safety, and healthcare, highlighting the potential of IoT to optimize infrastructure, resource allocation, and overall quality of life.

The authors emphasize the necessity of collaboration between governments, industry stakeholders, and citizens to address challenges such as privacy, security, and interoperability. By examining smart city initiatives worldwide, the paper identifies

common themes and key components for success, advocating for a holistic approach that integrates technology, governance, and citizen engagement. The review also acknowledges the need for responsible and equitable implementation of IoT technologies, urging policymakers and urban planners to work together to overcome these challenges and ensure sustainable urban development. The integration of advanced technologies like artificial intelligence and 5G is expected to further enhance the capabilities of smart cities, driving innovation and improving urban living conditions.

- *Christos Katrakazas, ConstantinosAntoniou, NataliaSobrino Vazquez, Ilias Trochidi, Stratos Arampatzis, (2020) 'Big Data and Emerging Transportation Challenges: Findings from the NOESIS project'*

The paper explores the critical factors necessary for the successful implementation of Big Data technologies in transportation. The NOESIS project, funded by the EU Horizon 2020 program, aims to identify key challenges and opportunities in leveraging Big Data to improve various aspects of transportation, including planning, operations, management, and maintenance. Through a comprehensive literature review and expert validation, the paper identifies 13 major challenges that can benefit from Big Data applications, ranging from environmental impacts to traffic management and road safety. The authors also propose a "Learning framework" and a "Value Capture" mechanism to assess the socioeconomic benefits of Big Data investments in transportation, aiming to enhance the value generated from these technologies.

It also highlights the significance of understanding the specific areas within transportation that can be optimized using Big Data, such as connected and autonomous vehicles, freight logistics, and infrastructure maintenance. By addressing issues related to data ownership, privacy, and quality, the NOESIS project sets the foundation for future research and practical implementations. The findings contribute to the broader understanding of how Big Data can transform transportation systems, offering insights into developing effective strategies and policies to harness its full potential.