Big Data Analytics Architecture for Real-Time Traffic Control

Syam Sundar Herle
School of Informatics and Computing, IN 47408, U.S.A.
Bloomington, Indiana
syampara@umail.iu.edu

ABSTRACT

The advent of Big Data has triggered disruptive changes in many fields including Intelligent Transportation Systems (ITS). The emerging digital devices have opened unique opportunities to enhance the performance of the ITS. There is crucial need to develop new tools and system to keep up with the Big Data advent. In this paper we will discuss how integrating Big Data with ITS will help in Real-Time traffic control and we will explore some of the existing system which helps in real-time traffic control. Furthermore, we will propose a flexible architecture based on distributed computing platform for real-time traffic control. We will use Kafka, a state-of-the-art Big Data tool for building data pipelines and stream processing.

KEYWORDS

i523, hid219, Big Data, kafka, Real-Time traffic control, Intelligent Transportation System

1 INTRODUCTION

According to [11] "the volume and speed at which data are generated, processed and stored is unprecedented". Big Data is the process of gathering management and analysis of data to reveal hidden patterns. The advent of Big Data have changed many fields ranging from urban planning to vehicle safety. In Big data approaches, the challenge is not alone confined to collecting data but also to exploit the collected data to draw valuable information . Given the magnitude of massive stream of real-time data push the limits of the current storage and processing capabilities. The current statistical model cannot be applied on the crowding data streams, as some these unstructured data [7] are context-based from the internet.

Most of the current system rely on ad hoc architecture solutions [2] [6], which are focused on satisfying predefined goals like predicting traffic flow and are hard to extend to different application. In response, this paper proposes a comprehensive architecture for Big Data for real-time traffic control, which is flexible in ways that it can accommodate diverse set of data sources and different ITS applications, particularly decision support for real-time traffic control.The proposed architecture is based on Kafka, as the data analysis using Kafka is reliable and scalable.

2 BACKGROUND

Real-time traffic control systems are composed of two main components: observation of the situation and implementation of the selected control strategy. A local system analyzes the real-time input data where they are integrated and process to identify the situation. Once a threshold is exceeded, one of the predefined strategies is implemented to optimize the controller objective function. Feedback loop and Model Predictive Control (MPC)[8] are the most common traffic control approaches which are mainly single objective and require purposely-sensed data. Data driven approaches for Big Data in ITS can be divided in three main category,

- Urban planning : Planning public transportation network using studies on travel demand and mobility pattern estimation using mobile location data[8] or call details records [9].
- Transportation operation : services in this domain either focus on decision-making support systems for traffic operations. For example travel time prediction[3], traffic incident and anomaly detection[13].
- Safety : exploring the critical situations arising from the design of the infrastructure has been studied by analyzing trajectories extracted from video data[12].

Big Data analytics approaches scale with respect to the amount and speed of data that needs to be analyzed by relying on a set of storage and computing machines called cluster [10]. Each cluster operates on its own, locally stored, set of data (map function) the results from individual machines are then aggregated (reduce function). Different Big Data analytics tools have emerged like Hadoop Distributed File System (HDFS) or a NoSQL database, which are for batch analytics. On the other side lie tools applying stream analytics, which is preferable when low-latency [5] data-driven decisions are needed. Important tools in this category include Flink, Kafka, and Spark Streaming.

3 RELATED WORK

Some of the research study on developed architecture for Big Data analytics in ITS are "Sipresk" created by Khazaei et al. [4] which is a cluster based platform that is built on Godzilla conceptual framework [10] and validated to estimate the average speed and the congested sections of a highway. And another study, by Xia et al. [2] have employed Hadoop distributed computing platform with MapReduce parallel processing to forecast near-future traffic flow.

From both the studies, we can see that there is scarce literature in applying Big Data approach in ITS and none of the approaches focus on Big Data stream processing.

4 PROPOSED BIG DATA ANALYTICS ARCHITECTURE FOR REAL-TIME TRAFFIC CONTROL

When developing a platform for data analytics in transport system we need to address many queries some of the nature of queries are,

- Descriptive and predictive queries
- Periodic and non-periodic
- Real-time and non-real time
- single and multiple user

4.1 REQUIREMENTS

In order to address above queries, an architecture for traffic control that relies on Big Data analytics has a number of requirements namely,

- Support analysis of data in streaming mode and analysis of historical data in batch mode.
- Provide an easy way to specify a data analytics query and its triggering policy.
- Provide an easy way to plug-in the analysis of different data sources, even as they become available.
- Provide intuitive mechanisms to considering multiple data sources in answering a single query.
- Provide an easy way to plug-in advanced data analysis (Machine learning algorithm).
- Large number of data sources and consumers and scale linearly with these numbers.
- Hardware faults by continuous operation and without loss of data.

4.2 Overall Architecture

In order to satisfy the above requirements, we came up with the overall architecture depicted in Fig. 1 In it, the different ITS actors (i.e. drivers, detectors, actuators, operators, etc.) act either as publishers or subscribers to Kafka topics. Kafka is used as the layer that decouples publishers and subscribers from the analytics engine.

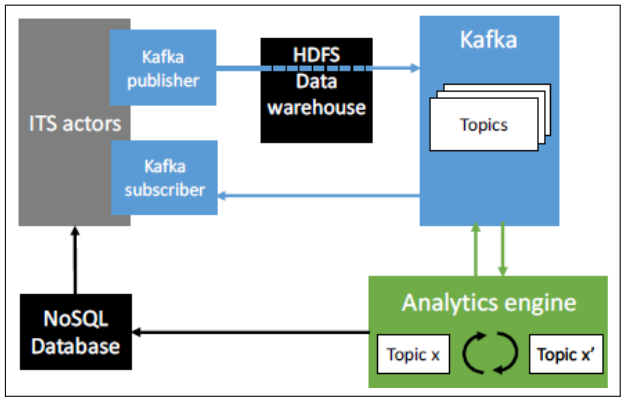

Figure 1: Architecture of the proposed platform.

Once a publisher publishes a new data item, this gets sent to Kafka and also saved in a Hadoop Distributed File System (HDFS) data warehouse for posterior analysis. The analytics engine gets input from all the publishing topics and performs data analysis. The results of the data analysis may trigger changes that are (i) published to one or more subscriber topics, and (ii) logged in a NoSQL database for posterior analysis of the findings (e.g. in order to determine the accuracy/recall of a predictive model mined from the incoming data). Once a change is published, it is picked up by the ITS actors that listen to the particular subscriber topic; they are ultimately responsible for enacting the change in the ITS.

4.3 Prototype of Big Data platform for traffic control

In the described platform, Kafka has the role of the communication medium between the traffic system with its sensors, like loop detectors and actuators like traffic lights and the data analysis module. A number of Kafka topics represent the different types of incoming data from the traffic system like mean speed from loop detectors, vehicle speed and position data from on-board GPS devices. There are no assumption on the format of the data, but in prototype platform the paper have used JSON format.A special Kafka topic, represented as change provider in the platform is responsible for delivering the changes in the form of Kafka messages that should be enacted in the traffic actors. Such changes are the results of data analytics engine.

The data analytics engine performs the analysis and control logic's defined by each consumer into vary from simple feedback loop to sophisticated machine-learning algorithms. Moreover, users can customize the time intervals for receiving the outcome of the analytics engine. As data come in, they are being processed via user-specified reducer functions. These functions are specific to each topic.

5 SIMULATION CASE STUDY

In the traffic control problem, the controller receives the average density from loop detectors on a cross section of a three-lane freeway and decides whether the hard shoulder should be opened or closed. Due to safety reasons, an operator observes the section via Surveillance camera to detect obstacles or stopping vehicles on the hard shoulder. The paper studies the hard shoulder opening system on a 3 km segment of A9 freeway in the north of Munich. This section of the freeway has been used as a digital test bed to assess the performance different types of sensors e.g. camera, which fits to Big Data definition in term of volume, velocity and variety suitable the characteristics of a Big Data analytics.

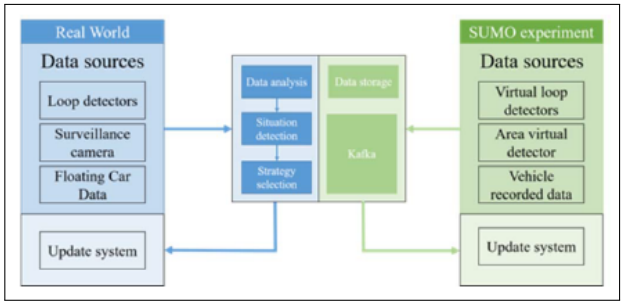

Figure 2: Cross-comparison of SUMO experiment with real-world situation.

Since we are interested in high-level ITS architecture and proof of concept for smooth operation of the proposed platform, without losing generality, we have modeled this section in SUMO [1], a microscopic traffic simulation. In order to achieve a realistic representation of the reality, we use virtual detectors in SUMO, each corresponding to an existing sensor. An area detector is placed over the hard shoulder to represent the surveillance camera, virtual loop

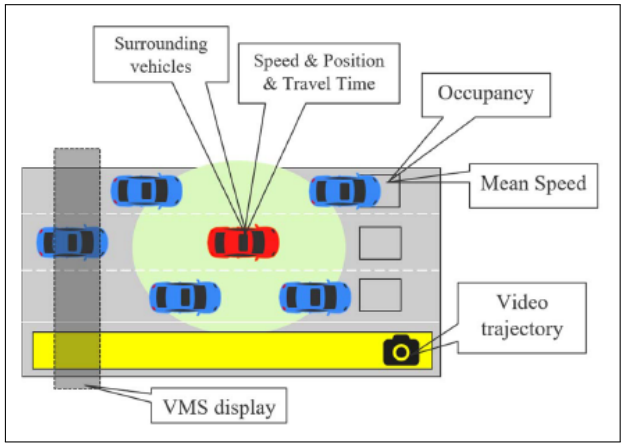

Figure 3: Topics from Kafka publishers in SUMO experiment.

detectors that measure mean speed and occupancy and floating car data that provide momentary speed and position as well as travel time along the section. Fig. 2 depicts a comparison of the real-world with the SUMO experiment and Fig. 3 illustrates the Kafka publishers and subscribers together with the corresponding published topics.

6 CONCLUSION

In this work, we proposed a comprehensive and flexible architecture for real-time traffic control based on Big Data analytics. The architecture is based on systematic analysis of the requirements of the domain. The proposed architecture has been partly reified in a platform employing Kafka. It has been put to action in operating a feedback control loop to open or close hard shoulder of a freeway. The main limitation of the study was lack of access to real-world data. Although using simulation in traffic studies is common, data generated in SUMO are well-structured, valid and do not require data quality and plausibility checks. We recommend to consider these essential issues in future research.

ACKNOWLEDGEMENTS

This work was done as part of the course "I523: Big Data Applications and Analytic" at Indiana University during Spring 2017. Many thanks to Professor Gregor von Laszewski at Indiana University Bloomington for their academic as well as professional guidance. We would also like to thank Associate Instructors for their help and support during the course.

REFERENCES

[1] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker. 2012. "Recent Development and Applications of SUMO - Simulation of Urban MObility". (2012), 128–132 pages.
[2] D. Xia, B. Wang, H. Li, Y. Li, and Z. Zhang. 2016. "A distributed spatial– temporal weighted model on MapReduce for short-term traffic flow forecasting". (2016), 246–263 pages.
[3] H. Chen and H. A. Rakha. 2014. "Real-time travel time prediction using particle filtering with a non-explicit state-transition model". (2014), 112-126 pages.
[4] H. Khazaei, S. Zareian, R. Veleda, and M. Litoiu. 2016. "Sipresk: a big data analytic platform for smart transportation". (2016), 419–430 pages.
[5] J. Kreps, N. Narkhede, J. Rao, and others. 2011. "A distributed messaging system for log processing". (2011), 7 pages.
[6] J. Yu, F. Jiang, and T. Zhu. 2013. "RTIC-C: A big data system for massive traffic information mining,". (2013), 246–263 pages.
[7] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen. 2011. "Datadriven intelligent transportation systems: A survey,". (2011), 1624–1639 pages.
[8] L. D. Baskar, B. De Schutter, and H. Hellendoorn. 2016. "Model-based predictive traffic control for intelligent vehicles: Dynamic speed limits and dynamic lane allocation". (2016), 246–263 pages.
[9] M. S. Iqbal, C. F. Choudhury, P. Wang, and M. C. GonzÃąlez. 2014. "Development of origin–destination matrices using mobile phone call data". (2014), 63–74 pages.
[10] M. Shtern, R. Mian, M. Litoiu, S. Zareian, H. Abdelgawad, and A. Tizghadam. 2014. "Towards a multi-cluster analytical engine for transportation data". (2014), 249–257 pages.
[11] OECD/ITF 2015. *"Big Data and Transport: Understanding and assessing options"*. OECD/ITF. Available at http://www.itfoecd.org/sites/default/files/docs/15cpb_bigdata_0.pdf,Accessed: 2017-3-26.
[12] P. St-Aubin, N. Saunier, and L. Miranda-Moreno. 2015. "Large-scale automated proactive road safety analysis using video data". (2015), 363-379 pages.
[13] S. Chawla, Y. Zheng, and J. Hu. 2012. "Inferring the root cause in road traffic anomalies,âĂİ in Data Mining (ICDM)". (2012), 141–150 pages.