Application of Streaming Data with Azure Data Factory in Vehicle Service Application

Dr. Sayed Abulhasan Quadri
SECAB Engineering College Vijayapura, India.

Dr. Senan Ali Abd
Al Hikma University College, Baghdad, Iraq.

Dr. Sayed Abdulhayan
P.A.College of Engineering, Mangalore, India.

Abstract
Cloud based data integration and cost optimizing has made great potential in real-time application such as traffic management. Predicting road traffic and suggesting instantly to select the less congested path would help the users in various traffic jam situations. A system is proposed using Azure Data Factory and Apache Kafka. Apache Kafka a favoured event streaming stage is utilized to gather, operate and accumulate pouring events data. It is capable of scaling to manage billions of streamed occurrences per minute. An attempt has been made that helps to find cost effective means in vehicle service application.

Introduction

Automobile traffic increased as population growth surges in the world. As a result environmental outcomes and people's health and protection concerns are more important than ever. Whereas there are no suggestions that the enormous automobile numbers will recede anytime soon, tackling the problem-inducing components, such as increased traffic and overcrowd, are among community's best stakes to avoid these outcomes. Thus, there is acute need of traffic management system to solve the delay and congestion problems.

Generally, anticipation of road vehicle traffic is done with support of Apache Hadoop and sensors. Commonly, utilized sensors are of two types, namely Passive Infra-red sensor and Ultrasonic sensor. These sensors are deployed in path of Traffic Signal area instead of attaching to vehicles. Obtained sensor data is collected and operated by using Apache Hadoop. Hadoop is a consignment refining design, so the anticipation of road traffic is occurred with the aid of previous actual data. Chief delinquent with present design is, it is not cheap as the deployment of sensors is required at every bypath of traffic signal and cannot envisage the real-life congestion or overcrowding. In this project, Azure Data Factory and Apache Kafka are utilized to address the traffic congestion problem. The project also emphasizes a smart route for the vehicle which is added value in respect of economic transportation.

Azure Data Factory and Apache Kafka

Azure Data Factory is Azure's cloud ETL (Extract transform and load) provision for scaled server less information amalgamation and information conversion. It provides a codeless UI (user interface) for instinctual crafting and single-pane-of-glass policing and administration. We can correspondingly mount and carry present SSIS (SQL Server Integration Services) bundles to Azure and execute these with complete concord in ADF (Azure Data factory. SSIS Runtime

Integration bids a copiously planned service, so we need not required to apprehend about substructure administration. [2]

Apache Kafka is an freely available dispensed task computing system utilized by numerous of companies for excellent rendering data pipelines, computing analytics, data incorporation, and crucial applications. It is a freely available platform maintained by the Apache Software Foundation written in Scala and Java Programming languages. It is intended to furnish a single, increasing-output, low-delay system for managing inputs of real-time data. Kafka could tie-up to other systems (for information export/import) via Kafka Connect, and provides the Kafka Streams libraries for computing stream applications. Kafka routines a zero-one grounded TCP-based protocol that is managed for accuracy and precision. Kafka depends on a "message set" schema which is grouped naturally avoiding roundtrip networking overhead. This leads to bigger packets of network, greater sized sequential disk tasks, continuously located memory blocks [...] which makes Kafka into a bursty and heavy stream of unsequenced messages. [3]

Literature review

Although several researchers have worked on the vehicular traffic management domain employing various technologies, brief highlights are presented in the review.

Zhang [4] has worked on Waze to resolve issues pertaining to traffic management. Waze is a subsidiary software application of Google that offers satellite navigation services. Real-time traffic data is indispensable to a diversity of applied applications. To get real time traffic information, various traffic perceiving appliances, such as infrastructure-mounted sensors, cameras, and loop detectors, have been fitted on roadway interconnects. However, traffic interventions have sought subsidiary information sources to observe traffic, owing to the exuberant cost of installation and maintenance in conventional data aggregation methods. Nowadays, crowd sourced congestion data has become easy available and is extensively considered to have excessive potential in smart transportation platform. Application Waze is a crowd sourcing real-time traffic software application that empowers users to disseminate real-time congestion data.

Azure platform has been suggested by Rawat, S., and Narain [5] to address congestion-related issues. The orchestration layer plays a key role in connecting the various environments and operationalizing the workflow in any Big Data or advanced analytics solution. The comprehensive solution might entail moving raw (uncooked) data from various sources to an Azure staging/sink store, running a series of rich transform jobs (ELT) on the raw data, and then producing the valuable insights that can be shared using reporting tools and kept in a data storeroom for quick access. The extract-transform-load (ETL) service offered by Microsoft Azure is called Azure Data Factory.

IoT gadget was created by Chong and Ng [6] for traffic management system. They made use of the IoT cloud server from Microsoft. In order to alleviate the traffic congestion, the fixed cycle traffic light system (TLS) was initially installed at street intersections. Fixed cycle TLS, however, is unable to handle a sharp increase in the number of registered vehicles. This work anticipates the creation of Internet of Things (IoT) traffic management system manoeuvres. Real-time traffic movement is captured by an Intel Edison, which connects to a Microsoft Azure IoT cloud server. The cloud server assigns priorities to each road based on its current level of traffic. Then, a dynamic active method is used to estimate the green light phase time (GLPT). According to the results of the simulation, dynamic cycle TLS reduces the length of the line and the wait time at the intersection of two roads by, respectively, 68 and 67 percent. Additionally, a traffic brigadier will find it easier to observe real-time traffic flow with the help of an observing programme.

AutoQML, a Cloud-Based Automated Circuit Architecture Search Framework, was introduced by Gómez et al. [7]. In order to optimally learn from and generalise from a source (input) dataset, conventional machine learning algorithms' erudition process was correctly tuned by hyperparameters that required to be customised. As a potential use for quantum computing that could soon offer a quantum advantage, quantum machine learning (QML) has been gaining popularity. However, quantum variations of conventional machine learning algorithms are familiar with numerous extra parameters and circuit differences that have their own complexities in tuning. The fundamental steps for automated quantum machine learning were made in this endeavour (AutoQML). A specific formulation of the issue was advised, followed by the advancement of a classical-quantum hybrid cloud architecture that permits parallelized hyperparameter discovery and model training.

Automated guided vehicle cost-effective methods were developed by Bejarano et al (AGVs). They used cloud-based features to implement a remote regulator mechanism to fix network issues with long and short-distance communication. A further streamlined solution is promised by remote monitoring and control of manufacturing equipment, which enables less expensive operational systems to run with no maintenance. However, the distance in terms of geography between a plant and its control centre may have an effect on the network connections' Quality-of-Service restrictions, which could stall the inclusive control process. They conducted a series of measurements across varying distances for the remote control, operation, and modelling of Automated Guided Vehicles (AGVs), which are frequently used in modern production environments, to gain a better understanding of these potential challenges and their consequences. They presented three communication patterns that reflected local, remote, and cloud-based service usage in order to complete these assessments. By using these patterns, they were able to link the Microsoft Azure cloud in the Netherlands, the VxLab at RMIT Technical University in Australia, and the plant of the future at Aalto University in Finland. They were able to use this to quantify crucial Quality of Service networking characteristics for data transfer over small, average, and extremely long distances. Empirical findings were shown to be pleasing and compelling.

IoT-enabled traffic management systems have been well discussed by Devon Smyth [9]. Systems for traffic management that are IoT-enabled essentially address such difficult problems at the source. Smart traffic management systems are built on networks of various sensors, particularly visual ones that transmit information about traffic, the environment, and potential dangers on high-traffic roads. With this knowledge, active traffic control systems can function without the restrictions of predefined timing interludes and instead fine-tune traffic light timing categorizations to direct real-time traffic problems in a more effective and safe manner. With the help of this suggested technology, communities might respond to accidents quickly, streamline traffic during rush hour, and ultimately reduce pollution by using more gas efficiently. Although almost any large metropolis might profit from this technology, a small network substructure is need to enable quick data transfer from the sensors. Instead of more shared solutions like adding roads and tracks, sophisticated traffic management systems are a more affordable and more sustainable method to reduce congestion in such densely populated cities using this pre-existing infrastructure. A growing sound network setup to enable smart traffic management systems may be one of the most effective ways to address both pollution and safety concerns, even as many of these cities throughout the world try to solve their congestion issues and avoid their natural repercussions.

There are numerous studies done by researchers in this area, and in the majority of them, road traffic predictions were made using sensors and the Apache Hadoop framework. Two types of sensors are typically used: passive infrared and ultrasonic sensors. These sensors are arranged in the lane of the traffic signal area and were not attached to any vehicles. Apache Hadoop is used to store and process the sensor-collected data. Hadoop is a system for consignment processing; therefore previous data was used to predict future traffic patterns. The main flaws with the current system are that it cannot forecast real-time traffic congestion and is not cost-effective because sensors must be installed in every traffic light lane.

Proposed Methodology and Implementation

A system is proposed using Apache Kafka and Azure Data Factory which helps to find cost effective means in vehicle service application.

The conceptual model to collect and process the data is shown in Figure 1. The collaborative technique of pipelining and streaming in Azure is well utilized. Data pipeline is a compound process of components which are interrelated with each other, the outcome of one component is given as input to the other component. The starting point of the data pipeline where input is fed is termed source and the end point where output is received is called a sink. All additional nodes in the pipeline are referred as transitional nodes. Each module in the data pipeline processes the data by executing some feats. Conferring to the obligation of any organization the proposed model can be used by any organization for evolving any data application.

The complete cycle consists of data collection, data ingestion, pre-processing, reception and data labelling. The conceptual model shown below is fully robotic in which tracking is accomplished all over the pipeline. Because of mitigation strategies the data pipeline is practically fault tolerant as the strategies elevate the belongings of liabilities. The squads can demand data from any place in the pipeline conferring to the requirements.

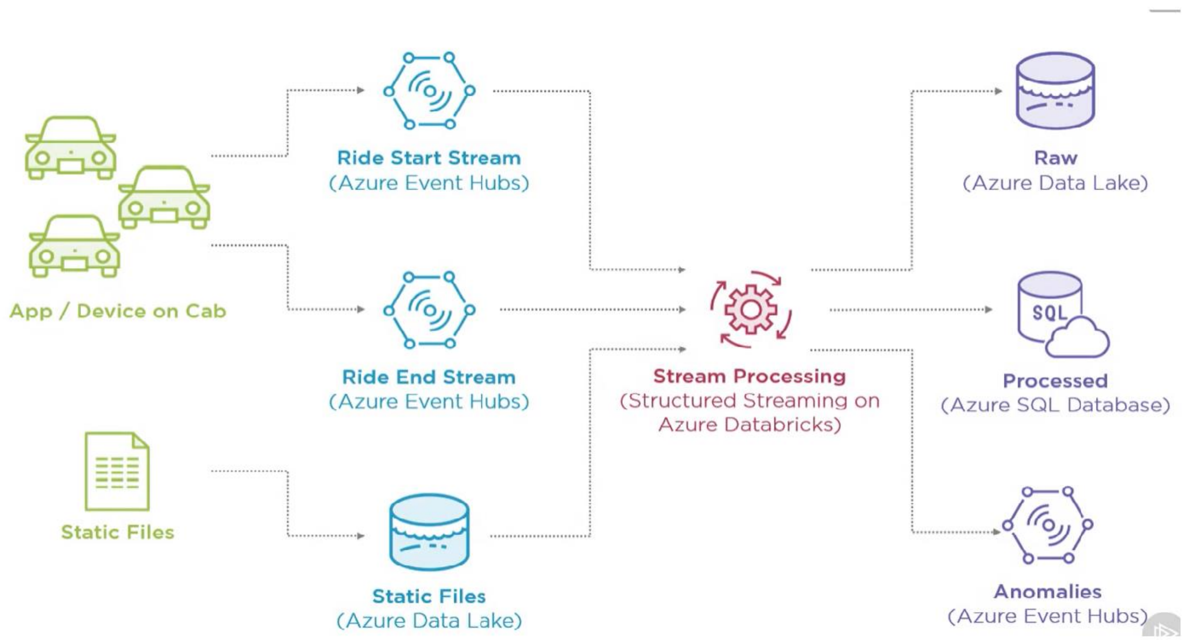

Figure1. Architecture of proposed system

Results

The Spark streaming engine obtains real-time streaming data from Kafka and processes it using robust distributed datasets in accordance with the suggested approach depicted in Figure 1. The Cassandra database is used to predict the entire amount of traffic as it is processed. Spring Boot is used to transform the database's extracted data into the final product. Results are shown in the monitoring dash wide as graphs and charts. This demonstrates the monitoring of road traffic in real time. It consists of three parts: a stream processor, a message system, and a collector for monitoring persistence. The network's data packets can be collected using Collector. It is adept at gathering both incoming and outgoing packets. The system has numerous collectors for gathering data from various switches. The data that is acquired from the collector is progressed by the stream processor. The important protocol then shifts to the stream processor to lower the input data once the collector first begins processing the detained packets.

The message system acts as a suspension bridge for data transmission between the stream processor and the collector. A suspension tie-up between the stream processor and collector is known as a message system, and it is used to design data transmission. By maintaining the congestion window, the sender can prevent a link between themselves and the receiver from becoming overloaded with traffic.

The congestion graphs of existing system and proposed system are compared and shown in Figure 2. The cost analysis of the same are depicted in Figure 3. It can be clearly comprehended that the proposed system employing integration of Apache Kafka and Azure data factory could provide best results in reducing congestion and cost.

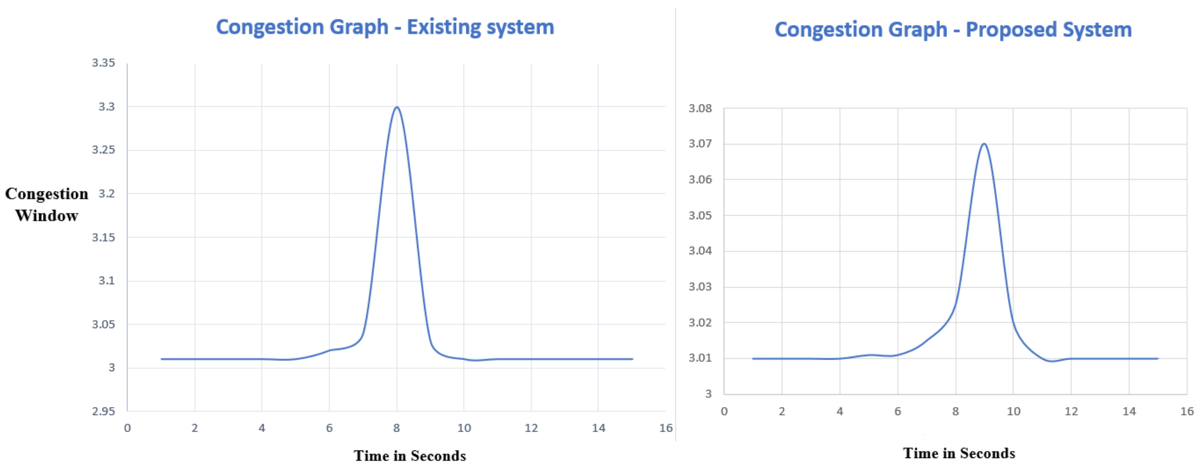

Figure 2

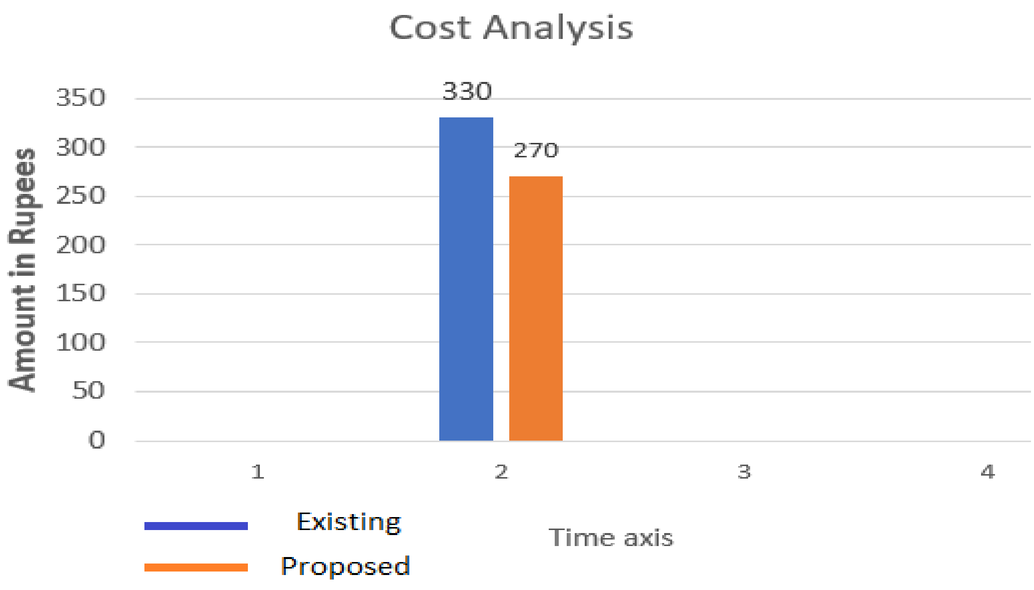

Figure 3

Conclusion

Real time traffic streaming data can be monitored across the globe with the usage of latest cloud computing services. The proposed work integrated Apache Kafka and Azure Data Factory to process data collected from vehicle management system. The potential of pipeline data streaming provided instant information at the dashboard end about least congestion route, thus reducing the cost parameter.

References

[1] https://hadoop.apache.org/

[2] https://docs.microsoft.com/en-us/azure/data-factory/#:~:text=Azure%20Data%20Factory%20is%20Azure%27s,with%20full%20compatibility%20in%20ADF.

[3] https://Kafka.apache.org/

[4] Zhang, Zhihua, "Exploring the Potentials of Using Crowdsourced Waze Data in Traffic Management: Characteristics and Reliability. " PhD diss., University of Tennessee, 2020. https://trace.tennessee.edu/utk_graddiss/6899

[5] Rawat, S., Narain, A. (2019). Introduction to Azure Data Factory. In: Understanding Azure Data Factory. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-4122-6_2

[6] H. F. Chong and D. W. K. Ng, "Development of IoT device for traffic management system," 2016 IEEE Student Conference on Research and Development (SCOReD), 2016, pp. 1-6, doi: 10.1109/SCORED.2016.7810059.

[7] Gómez, Raúl Berganza and O'Meara, Corey and Cortiana, Giorgio and Mendl, Christian B. and Bernabé-Moreno, Juan, Proceeding of IEEE ICSA 2022, FEB 2022 pages 1-8.

[8] Bejarano, R. Pääkkönen, J. Olaf Blech, I. Peake, P. Herrmann and V. Vyatkin, "Assessing Long Distance Communication Alternatives for the Remote Control of AGVs," 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), 2020, pp. 69-76, doi: 10.1109/ETFA46521.2020.9211954.

[9] Devon Smyth, Technical report, 2021, University of Pittsburgh https://sites.pitt.edu/~dts43/content/IoTTMpdf.pdf