

# Vehicular Traffic Violation Controlling System using Big Data Technologies

Priyanka M R

Department of Computer Science,  
4<sup>th</sup> Semester MTech, RNS Institute of Technology  
Bengaluru, India

Hemanth S

Professor, Department of Computer Science,  
RNS Institute of Technology,  
Bengaluru, India

**Abstract**—This paper presents how the big data tools and technologies can be utilized in controlling the traffic violations. Since the number of vehicles are increasing day-by-day, simultaneously the violations are also increasing. There is a need for controlling these violations. But when performing these actions manually, it would consume more time and all the details cannot be maintained in real time because of the large amounts of data. Thus automating is necessary. This is where big data comes into picture.

**Keywords**— Big Data, Traffic violations, Manual, Automating

## I. INTRODUCTION

In the past few years, there is a drastic change in the amount of data that is being generated. This large amount of data is often called as “Big Data”. Big data refers to the datasets which are huge. By using traditional methods, managing and processing of such data is tedious. Big data came into existence when the requirement of the processing of dataset exceed the capability of the traditional data management system (for example: RDBMS). The huge datasets are usually obtained from IOT devices, sensors, logs, mobile devices etc.

Big data has various characteristics such as volume, variety, velocity, veracity, validity, variability, visualization etc. Out of these, volume, variety and velocity are considered as 3V's of Big data. Volume refers to the large amount of data that is generated for every second. It is not the matter of terabytes but much more than that such as zeta bytes. Variety refers to that different types of data that can be processed using big data tools and technologies. It specifies that even unstructured data can be processed so there is no requirement that the data be structured. Next characteristics is velocity. It represents the speed at which the data is generated, analyzed, processed and also the speed at which it is visualized.

Big data tools and technologies are often complicated but working with big data results requires that the base has to be set up properly such that the administration, circulation and the flexibility of the data is maintained [1]. Moreover for real time big data processing, the framework and the implementation techniques are massive. The Hortonworks distribution platform is used here. HDP provides various services of which NiFi, Kafka, Storm etc are used.

Since there are a large number of vehicles increasing day-by-day, there is also increase in the number of violations occurring. The count of the traffic rules violated is increasing. Few of the commonly violated traffic rules are parking in no parking area, over speeding, signal jumping, rash driving, reverse direction, driving without license, drunken driving, carrying excess goods and so on. These violations must be reduced as they leads to hazardous situations.

For tracing of the vehicle, the GPS tracker is required. The GPS system is set up in the vehicle such that the details (vehicle id, latitude, longitude and time stamp) is generated for every 5 seconds representing the position of the vehicle. As the data is generated for every 5 seconds, this results in big data. Simultaneously all the vehicles' GPS system would be generating data. The violations must be handled in real time big data processing.

Of the violations listed previously, the most commonly occurring are no parking, over speeding, driving in one way direction. Hence this paper shows how these violations can be controlled using the big data tools and technologies.

## II. LITERATURE SURVEY

Jinesh, Mahendra and Chintan [1] proposed an overview of Big data. This overview describes how big data came into existence. It specifies how the data is generated in very large amounts. The datasets are continuously obtained from all logs in phones to military services. The paper describes that the term Big data is very vague and there is no proper explanation for the frameworks of Big data as the limits of Big Data are very dynamic. The traditional methods are not suitable. This is because the techniques in the traditional methods are not suitable for managing and processing such huge amounts of data upto zeta bytes and more. This paper describes the 5V's of Big Data: Volume, Variety, Velocity, Veracity and Value. These 5V's represents the characteristics of Big Data.

Xiaojng Zhao [2] has described the various tools and technologies for Big Data. Hadoop is the most commonly used tool. Hadoop makes use of Map Reduce concept for processing. Map Reduce, HDFS, YARN services are introduced here. The work flow of the Map Reduce is shown in this paper which describes the mapper and reducer side. However Map Reduce concept is suitable only for the batch files.

Dan Omar, Anna [3] provided the basics of Big Data. They also described the platforms and frameworks suitable for big data. It introduced the concept of NoSQL databases. This paper gives the strengths and weakness of the various frameworks so that an optimal framework is chosen. Apache Hive, Apache Pig, Apache Hadoop, Apache Storm, HDFS(Hadoop Distributed File System), HBASE and services are detailed. Apache Storm is suitable for real time processing. The strength of the storm is that it is fast, reliable, supports multiple common programming language.

Wei-Hsiu Wang and Woo-Trong hin [4] described the emerging technology that is Big Data. This paper suggests Hadoop Big Data technology for key-value database and also explains the architecture. Major vendor of Big Data are introduced. IBM, Oracle, Microsoft are some of the major vendors of Big Data.

Real time Big Data has various challenges [5]. Jing hin, Ping Wong, Zhigo Zheng and Shengli Sun have conducted analysis and produced a results on the differences and challenges between big Data and real time big data. Data collection in big data system is obtained from the mobile terminals, tablets, computers and other terminals and is often stored in the cache. Whereas for real time big data, data collection has much more complexity, it requires the synchronization of the systems. Similarly data analytics, data security, data management performance also differ for big data and real time big data systems. This paper explains the framework for real time big data processing. The service that is suitable for real time big data is Storm.

Surshanov [6] examined various services of big data and concluded that apache storm is optional for real time data. Stream processing is designed in order to analyze and process the data which is generated in real time. In storm, this stream processing system is represented as a single program. In this system, the data sets are split into a number of streams and then sent to the storm topology which consists of spouts and bolts. Storm is mainly used because using storm, it is easy to fetch real time data, fast, fault tolerant, reliability and scalability.

Kafka[7] is a scalable, messaging system, with its core architecture as a distributed commit log. This paper tells that kafka was initially used by the LinkedIn corporation for the processing of the messages.

### III. PROPOSED ARCHITECTURE

In this paper, we propose the traffic violation controlling system that is automated using the big data tools and technologies. Fig.1. represents the system architecture for the traffic violation controlling system. As shown in the diagram, the GPS tracking system is installed on all vehicles. The GPS application is built such that the details (vehicle id, latitude, longitude, time, etc..) are generated for every 5 seconds. The output from the GPS is a text file. The text files from all the vehicles are sent to a single directory in the system. The text files that are obtained from all vehicles would be huge containing many latitude and longitude values. Hence big data technologies are chosen for the processing of these data.

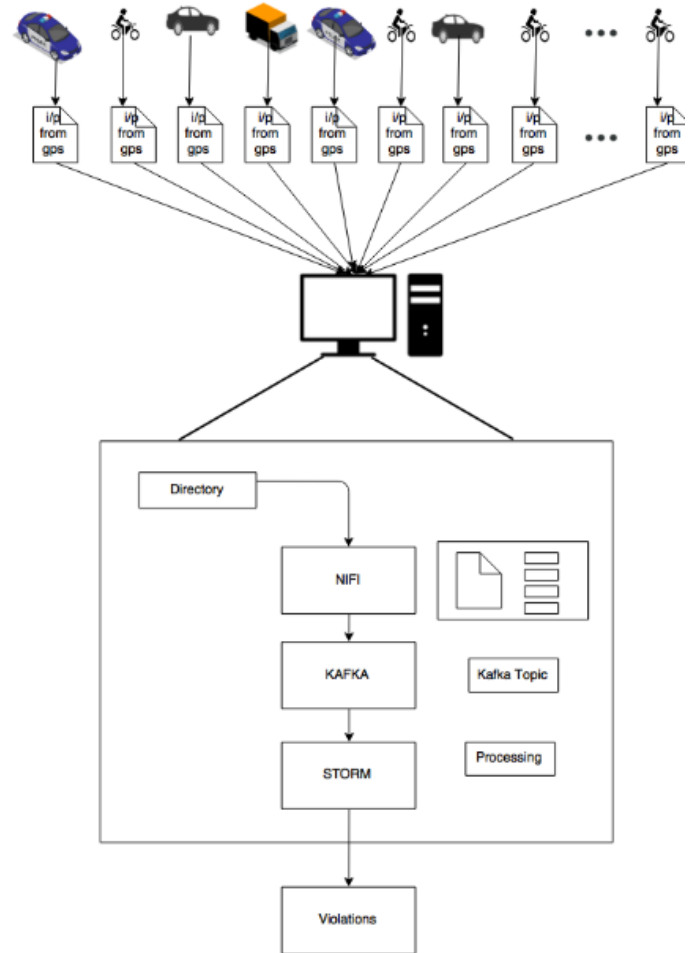


Fig. 1. System Architecture

From here starts the usage of the big data technologies such as NiFi, Kafka and Storm. These services are provided by Hortonworks Distribution Platform, a platform which provides various services of big data at a single place.

NiFi is a tool in HDP which is used for managing the flow of data between different systems. Here, NiFi is setup such that it picks up all the text files from the specified directory. NiFi consists of many different types of processors. While building the dataflow these processors are used to specify the further processing of the data. the processes which are used here are ExecuteProcess, SplitText, UpdateAttribute, RouteOnContent, MergeContent, PutFile and PutKafka. The ExecuteProcess would execute the shell script specified in it. Here, it is to read the files. After the files are read, next is the SplitText, the files are split based on the number of lines. This leads to the generation of the flow files. Here each flow file contains one line because the files are split based on one file so that each line would be processed. The UpdateAttribute processor assigns a unique UUID for every flow file. Then the Flow files are merged and sent to the PutKafka processor. This processor writes the data to the Kafka Topic.

Once the messages are in Kafka topic, Kafka picks these messages for message streaming. Kafka is also a tool provided by HDP. Kafka is used because it does message streaming between its producer and consumer. Here NiFi acts as the Kafka producer. The streams are then sent to the storm for the logical processing.

Storm is the tool provided by HDP for the logical processing of the data. This is place where the actual logic is written. Storm topology is comprised of spouts and bolts. Spouts pick streams of data from the kafka consumer and sends it to the bolt. Bolts perform the actual processing of the data streams. Bolts write the processing results to the worker log. Thus the violations in the traffic are listed.

#### IV. IMPLEMENTATION

Initially each road is represented as a polygon. Therefore to differentiate between the roads, parking areas and signals respectively road polygon, parking polygon and signal polygons are created using Google My Maps.

##### A. Parking Module

This module describes the various factors and conditions to be considered in order to identify the violations that occur related to parking. By considering the consecutive latitude and longitude, we need to decide whether the vehicle is moving or stationary. If the consecutive latitude and longitude values are same for a certain period of time, then the vehicle is stationary, else the vehicle is moving. Based on the latitude and longitude we need to decide to which polygon the vehicle lies. Whether the vehicle lies in the road polygon, parking polygon or signal polygon must be decided. Based on the respective polygon further factors must be taken into consideration. If the vehicle is in parking polygon then the vehicle is parked. Then it is necessary to check whether the two wheeler and four wheeler vehicles are parked properly. There is a count variable, which is used to have a track on the number of minutes the vehicle is parked or stopped. If the vehicle is in no parking area during a signal then it is considered as stopped, not parked.

##### B. Over Speed Detection Module

If the result of the polygon detection is a road polygon, then other parameters such as over speeding, rash driving, driving in one way, drunken driving can be considered. In this module lets consider the over speeding parameter. In order to detect the over speeding vehicles, it is necessary to have the details of speed, which is obtained from the GPS system. For every road polygon a maximum speed limit is set. The speed of the vehicle and the speed of the polygon are compared to decide whether the vehicle is speeding or not.

##### C. Reverse Direction Detection Module

If the result of the polygon detection is a road polygon, then, driving in one way can be considered. In this module this is considered. For every road a direction is assigned. The direction of the vehicle is computed using the bearing angle. Using bearing angle, the direction of the obtained, this is compared with direction specified in the polygon.

#### V. APPLICATIONS

The traffic violation controlling system developed using big data provides various useful applications. Since everything is automated, the man power required would be very less. This system would take care of all the violations. Once the violations are identified, the violations would be notified to the person driving the vehicle. On the other hand this would also reduce the number of accidents. This can be done because the person would be notified in seconds as soon as he violates the rules.

#### ACKNOWLEDGMENT

I would like to make use of this wonderful opportunity to thank my parents for their constant support and motivation. I would like to thank my internal guide Prof. Hemanth S, Department of Computer Science at RNS Institute of Technology for their guidance in successfully undertaking the project. I would also like to thank our beloved Dr. G T Raju, who is the professor, dean and HOD of Department of Computer Science for the encouragement and support. Finally I would also like to thank my teaching and non-teaching staff for providing us wonderful teaching and their blessing to move in a right path.

#### CONCLUSION

The existing traffic violation controlling system involves more man power for performing functionalities such as listing the vehicles that violated the rules. Moreover these functionalities cannot be performed all at once in real time. Thus automating the violation by controlling system is important. This is done by using big data concepts such as NiFi, Kafka and Storm. The proposed system is designed such that it recognizes three traffic violations: parking, over speeding, driving in one way directions. As a future enhancement, the other traffic violations can be considered.

#### REFERENCES

- [1] Mahendra S. Patil , Jinesh K Kandar, Chintan B. Khatri, " Big Data Overview", International Journal of Engineering Research and Technology (IJERT), volume 3 issue 7, July 2014.
- [2] Xiaojing Zhao, " An evolution for optimal big data processing with hadoop", University of Dresden, September 3, 2014.
- [3] Dan Omar, Anna Ohlsson, "A guide in the Big Data Jungle", Blekinge Institute of Technology, 2015.
- [4] Woo-Trong-hin, Wei-Hsiu Weng, " A Scenario Analysis of Big Data Technology Portfolio Planning", International Journal of Engineering Research and Technology (IJERT) vol 2, issue 3, May 2013.
- [5] Ping Wang, Zhigao Zheng, Shengli Sun and Jing hui, " Real Time Big Data Processing Frameworks, Challenges and Solutions", Applied Mathematics and Information Science, vol. 4, 3169-3190, 2015.
- [6] S Surshanov, "Using Apache Storm for Big Data", IITU , Kazakhstan, January 10, 2015.
- [7] Guozhang Wang , Joel Koshy , Sriram Subramanian , Kartik, "Building a Replicated Logging System with Apache Kafka", LinkedIn Corporation.