International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 5 539 - 542

Big Data Approach for Secure Traffic Data Analytics using Hadoop

Koushalya Bijjaragi1, Poonam Tijare2
1 M.Tech Student, Computer Science and Engineering, CMR Institute of Technology,
132, AECS Layout, IT Park Road, Kundalahalli, Bangalore 560037
30koushalyab@gmail.com
2Assistant Professor, Computer Science and Engineering, CMR Institute of Technology,
132, AECS Layout, IT Park Road, Kundalahalli, Bangalore 560037
poonam.v@cmrit.ac.in

***Abstract*—** As the volume of traffic is increasing day by day, it gets difficult to store and process such huge sets of data using traditional software. A cluster of storage devices is needed to store such huge amounts of data and also a parallel computing model for analyzing those huge inputs of data. Hadoop is one such framework that provides reliable cluster of storage facility, which stores huge data in a distributed manner using a special file system, called Hadoop Distributed File System and provides efficient parallel processing feature through MapReduce framework. Using Map Reduce the filtered traffic data can be fetched easily, to provide end user with traffic analysis and giving useful predictions.

***Keywords*—** *HDFS: Hadoop Distributed File System, Map Reduce*

*****

1. INTRODUCTION

Traffic congestion is one of the uncontrollable problems faced almost every day in urban environments. The reasons for which could be attributed to increasing growth in the number of vehicles, lanes having improper plan or lacking a pre-plan, major junctions and traffic signals, commonly used roads for easy access to other roads, roads having no dividers, unavoidable long routes, bottlenecks and worn out roads and so on. Apart from these challenges faced in the current infrastructure, due to other constraints like space, economic or political, the engineers involved in road construction are obliged to construct better, well laid road networks and flyovers in neutralizing the problems of traffic jams in the urban scenario. Hence, the number of risks and challenges involved in stabilizing the traffic flow increase if the plan made to deal with it is baseless. Also the renovations made impact the existing map and would possibly lead to unwanted loops or perhaps new spots for traffic jams would come into the picture if addressing the possibilities of such issues in the near future is not heeded in a tactful manner.

One of the most beneficial applications of the systems employed in traffic control is the improved ability to control the road network traffic. Tracking methods are used to capture the position and/or location of massive number of mobile objects. With the help of that tracked information, analysis and prediction of traffic density in a given network is enhanced. This renders valuable information for controlling traffic flow, prediction of congestion and reducing the number of accidents in that network.

The field of big data for resolving the above questions provides a new technical approach. Big data applied to road traffic analysis has the following advantages:

1. Traffic management system that uses big data technology can handle vast amounts of complex and diverse data. Big data has resolved three major problems: data storage, data analysis and data management. Hadoop is born with the ability to handle massive amounts of data where data is segmented and is stored on different nodes. A large task is divided into small tasks, and is processed in a MapReduce model. At the same time, the system stability and fault tolerance is important.

2. Big data can improve the efficiency of transportation industry largely. Transportation industry, involving many aspects of work, need to handle massive amounts of data each day, needs more controlled mode of application and has a great deal of equipment. In the aspect of improving transport efficiency, improving the threshold capacity of the road network, adjusting traffic demands, big data technology has obvious advantages.

Heterogeneous nature of accident data is the major problem in its analysis. This heterogeneity should be reflected on analysis of the data else, few relationships that exist between the data might remain unseen.

The term big data classifies the particular forms of data sets comprising formless data which will be taken from the technical computing application layers.

2. EXISITNG SYSTEM

At present, traffic information is commonly shared between different traffic agencies by means of voice or data communications. For such data communication between management centers, a common language and a frame of reference is required. There are many possible ways to analyze traffic data like manually counting each vehicle or through image processing, video analytics and so on. Manual counting involves a team to count each vehicle travelling on the road and update the statistics and a separate team is required to analyze these results of statistics.

2.1 Limitations of the Existing System

- Requires very huge man power.
- Time consuming and cumbersome.
- Create unmanageable traffic jams.

- If there is no traffic, the team personnel still need to wait at junctions.
- Lack of traffic details to users: Present traffic systems fail to provide traffic information including busy/idle time and number of accidents happening.
- The expense involved in maintaining the field components used is also high.
- Setting up the framework initially is also difficult.

3. PROPOSED SYSTEM

A dataset which consists of information about traffic is created. Map Reduce is used for processing to get the secure analysis of traffic data. Based on the analysis, prediction of traffic is made to show at what time it will be high and low in a day for a particular area. Prediction is also made for which month will have most number of accidents. Authentication is provided using signature, hence user security is ensured. Encryption algorithm is used to protect data, as it is sent across network.

3.1 Hadoop Distributed File System – HDFS

A cluster in HDFS consists of a single name node which is a master server machine and is responsible for managing the entire file system. It also provides an access to the file system as requested by the clients. Each cluster also consists of multiple data nodes, and each data node in turn consists of blocks of split data. Data distribution to these data nodes is maintained by the master node. The main operations of data nodes include file read and write.

Hadoop adopts the concept of rack awareness, which mainly helps in storing data into a rack and finding its location in the cluster. This means that a Hadoop Adminstrator can clearly define what chunks of data should be saved on a particular rack. This way, it is possible to mitigate the data loss if a rack fails entirely. Network performance is also improved because data replication is done on various racks of different machines.

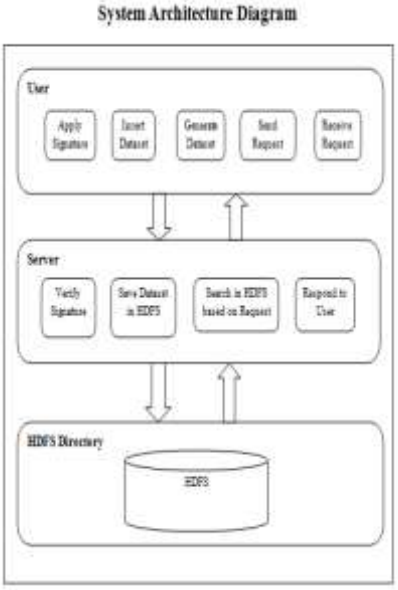
3.2 MapReduce

MapReduce is a software framework which was introduced by Google to carry out parallel processing on large data sets. This large data set is distributed over a large number of machines present in a cluster. For quick access, each machine computes and stores the data locally, this in turn contributes to distributed parallel processing. Such a computation involves two parts – Map and Reduce. In the Map phase, data nodes take raw input data and produce intermediate data based on the type of computation and then that data is stored locally. In Reduce phase, intermediate outputs from map phase is fetched by the nodes and then it is combined to derive final output that is stored in HDFS. Name node with its prior knowledge on the data distribution, tries to assign the task to a particular node based on the locality of data. Developers can write custom map and reduce functions suitable to the application and the MapReduce function then takes care of distributing and parallelizing tasks across a rack on commodity hardware in the cluster underneath.

Inter-machine communication is managed by the framework, thus programmers only have to focus on actual map-reduce functions. Hadoop uses this framework to analyze large datasets distributed over HDFS because of its unbeatable fault tolerance, reliability, distributed and parallel computing features.

3.3 System Architecture

Architecture design is a diagram which represents the basic structure of the entire project. It includes the various components that are a part of the project and how the components are connected. It also shows the actions performed by each component. Figure 3.1 shows the architecture design where the user needs to first sign in to communicate with the server. The user is provided with options to generate a dataset and also to manually insert an entry into the dataset.



Figs 3.1 System Architecture

The other operations of server include storing the dataset into HDFS and fetching the matching data from the directory as requested by the user. It does so by sending the request to HDFS directory where the Map and Reduce phase takes place according to the input request data.

The HDFS directory, when finishes the reduce phase, sends the result back to the server where the data is encrypted and sent across network. Once the user's signature is verified, decryption takes place and the requested information is sent to the user. The user on receiving the result of requested query, can analyze the data easily.

3.4 Signature Generation Algorithm

Input: Public key (A,B,h) , system parameters , message M
Output: Generate a valid group signature on M
Select random numbers a, roM, roR, mus,mux, mueprime, mut, muE
Computes the following values

E0 = g * roE
E1 = h+(h1*roE)
E2 = h+(h2*roE)
ACOM (A*(a2^rom) mod n).mod n
s = (Eprime+ke)* roM
BCOM (B *(w^ roR mod l) mod l

t = Eprime * roR

V0 = g*muE
V1 = (g * mux+(h1 * muE)
V2 = (g*mux))+(h2* muE)
Vmpk= (((a1^mux mod n)*(a2^mus mod n)).mod(n))*(ACOM^~mueprime mod n)).mod(n))
Vrev= ((w^mut mod l))*(BCOM^~mueprime mod l)))mod lE = E0+E1+E2
V = V0+V1+V2
reste = ACOM+BCOM+V+Vmpk+Vrev
Set c = f(E + reste + message)
Construct the following numbers
taux = c * (x+mux)
taus = c * (s+mus)
taut = c *(t+mut)
tauePrime = c *(Eprime+mueprime)
tauE = (c*(roE+muE)) mod(o)

Return
σ=(E0,E1,E2,ACOM,BCOM,c,taux,taus,tauePrime,taut,tauE)
End

3.5 Signature Verification Algorithm

Input: System parameters and signature σ = (E0,E1,E2,ACOM,BCOM,c,taux,taus,tauePrime,taut,tauE)
Output: True or False

Compute the following values
taue = (c*(expKe+tauePrime);
tauEG = g *tauE
a0a1 = (a0^c mod n)*(a1 ^ taux mod n))mod n
a2A= (a2^taus modn)*(ACOM^ ~taue mod n)mod n
Vmpk = (a0a1 * a2A)mod n
Bw= ((b^c mod l)*(w^taut mod l))mod l
Vrev = (bw*(BCOM^~tauePrime mod l)) mod l
E = E0+ E1+E2
V = V0+V1+V2
reste = ACOM+ BCOM+ V + Vmpk+ Vrev
if c = f(E + reste + message)
Return True
else
Return False
End

3.6 Algorithm used for Busy/Idle Traffic Prediction

Input: Traffic details from MapReduce
Output: Time when the traffic is low, average, high
Step 1: Filter traffic between 6 to 10
Step 2: Find highest number of vehicles
For all traffic details
{
Find: number of vehicle / highest number of vehicles
}
Step 3: Sort average
Step 4: Display time when traffic is less
if (val<0.5)
{
Display as best case
}
else if (val<0.8)
{
Display as average case
}
else
{
Display as bad case
}
Step 5: Stop

4. RESULTS AND IMPLEMENTATION

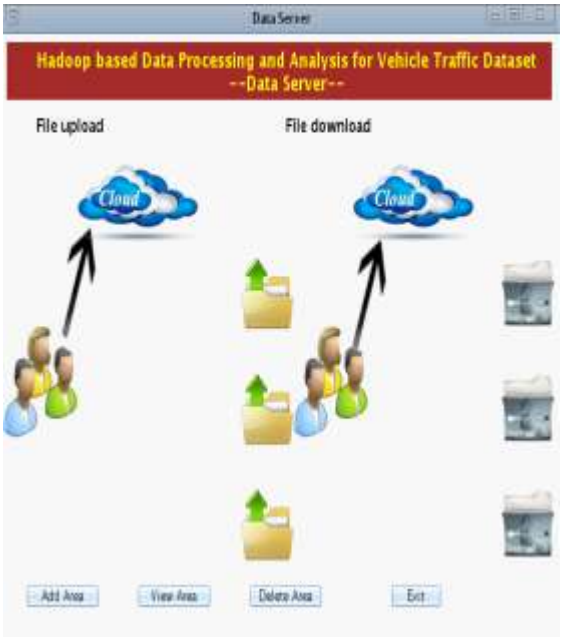


Fig 4.1 Server Page

The figure 4.1 shows the server screen. For user's request to be processed this page should be running always. The admin has control over this page and can add areas, view list of areas and delete areas any time he/she wishes to.

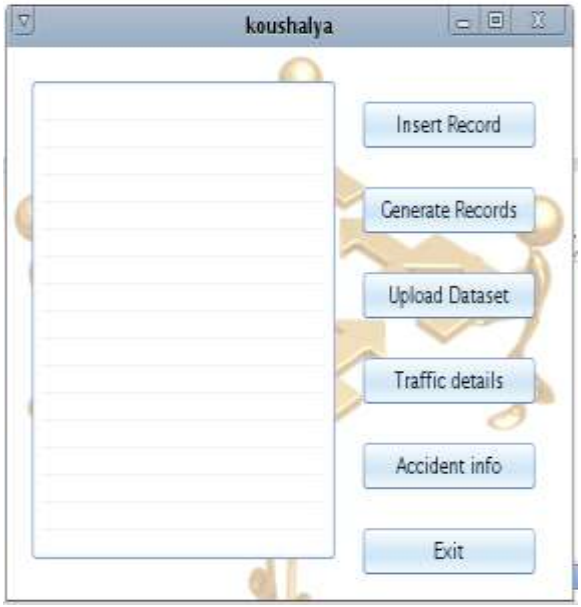


Fig 4.2 User's Main Menu

The figure 4.2 shows the page which appears as the user logs in successfully, where in the logged in user's name will be displayed on the top. The user has full access to all the buttons as shown in the snapshot.

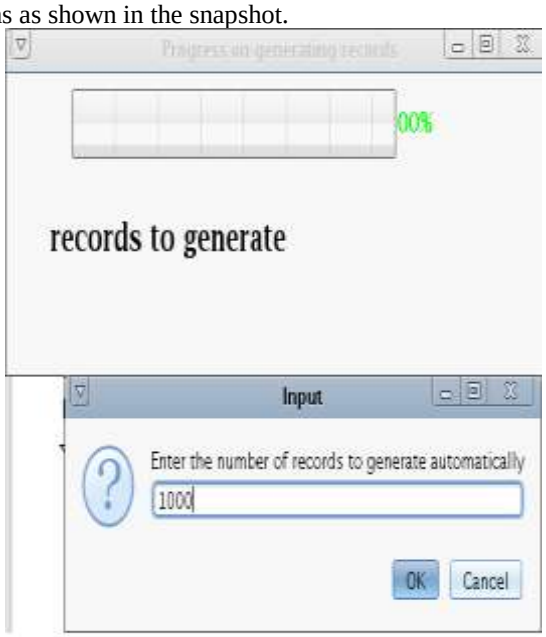


Fig 4.3 Generation of Dataset

As shown in figure 4.3, the user can generate the dataset by specifying the numbers of records to be generated by the application automatically.

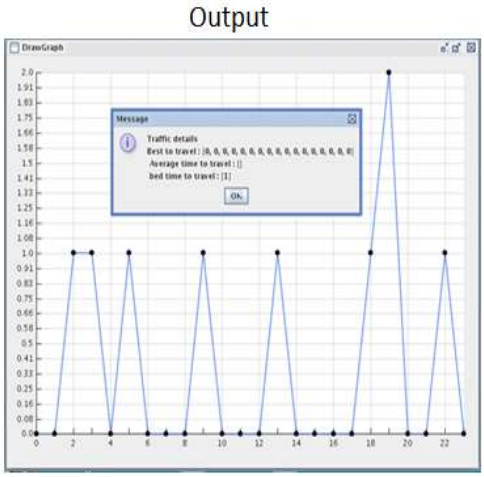


Fig 4.4 Traffic Flow Analysis and Prediction

This snapshot gives the prediction of traffic flow for a particular date. It indicates which is the best, average and bad time to travel based on hourly traffic density.

CONCLUSION AND FUTURE SCOPE

The project presented a secure approach to collect the user requests for traffic data details. Each user is authenticated and only upon signature verification the requested details are provided to the users. The proposed system is designed for secure analysis and giving predictions about the traffic flow details in a day and number of accidents happening per month. Overall, that data which the user wishes to retrieve can be efficiently fetched using the Big Data approaches such as Hadoop and MapReduce technology, by overcoming all the limitations that were encountered by the existing system.

At present the application operates offline, in the future changes that make it work online should be incorporated. More useful predictions can be made if the data to be operated on is the real-time data which can be fetched easily from open source platforms.

References

[1] "*Traffic Prediction and Analysis using a Big Data and Visualisation Approach*" - Declan McHugh, Department of Computer Science, Institute of Technology Blanchardstown, March 10, 2015

[2] "*Traffic Accident Analysis Using Machine Learning Paradigms*" - Miao Chong, Ajith Abraham and Marcin Paprzycki, Computer Science Department, Oklahoma State University, USA, December 20, 2004

[3] "*A Perspective Analysis of Traffic Accident using Data Mining Techniques*" - S.Krishnaveni, Dr.M.Hemalatha, International Journal of Computer Applications, Volume 23– No.7, June 2011

[4] "*Application of Big Data in Intelligent Traffic System*" - Gang Zeng, IOSR Journal of Computer Engineering, Volume 17, Issue 1, Jan – Feb. 2015

[5] "*A Comprehensive Analysis of The Vehicle Traffic Prediction*" - Ravindra M, Lokesha V, Prasanna Kumara, International Journal of Engineering Research & Technology, Vol. 1, Issue 8, October 2012

[6] Abellan J, Lopez G, Ona J. "*Analyis of traffic accident severity using decision rules via decision trees*", vol. 40. Expert System with Applications: Elsevier; 2013.

[7] Tom White, *Hadoop: The Definitive Guide*, 3d. Ed. CA, 2012.

[8] Improving Decision Making in the World of Big Data http://www.forbes.com/sites/christopherfrank/2012/03/25/improvingdecision- making-in-the-world-of-big-data/

[9] Pakize S, Gandomi A (2014), "*Comparative Study of Classification Algorithms Based On MapReduce Model*", International Journal of Innovative Research in Advanced Engineering (IJIRAE), 1(7), 251-254.

[10] Bhagattjee B (2014), "*Emergence and Taxonomy of Big Data as a Service*".

[11] Ashish Vijay, Priyanka Trikha, Kapil Madhur, "*A New Variant of RSA Digital Signature*", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 10,October 2012, ISSN: 2277 128X.

[12] Gang Zeng, "*Application of Big Data in Intelligent Traffic System*", IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN:2278-0661, p-ISSN: 2278-8727, Volume 17, Issue 1, Ver. VI(Jan –Feb. 2015), PP 01-04.

[13] Ashwini E.M, Sunil M.E, "*Tadoop:Traffic Analysis and Traffic Solution Using Hadoop*",International Journal of Innovative Research in Computer and Communication Engineering, Vol.3, Special Issue 7, October 2015.