Prevention Of Road Accidents Using Bigdata Analysis

PANGA.VIJAYA KUMAR

Department Of Computer Science and Enginnering

Hindustan Institute Of Technology and Sciences

CHENNAI,INDIA

NALLABOLU. RAMESH

Department Of Computer Science and Enginnering

Hindustan Institute Of Technology and Sciences

CHENNAI,INDIA

VELAGANTI.BHARGAVA RAJU

Department Of Computer Science and Enginnering

Hindustan Institute Of Technology and Sciences

CHENNAI,INDIA

Dr.S.SATHYALAKSHMI

Department Of Computer Science and Enginnering

Hindustan Institute Of Technology and Sciences

CHENNAI,INDIA

*Abstract— **The aim of this project is to analyze the road accidents data in India and to generate the graph which include conditions to follow to prevent happening of Accidents. Result scrutiny displays that road accidents and fatalities in India differs in accord with gender, age, time, weather conditions . Age group 20-49 ages is the most endangered population group though males faces complex level of fatalities and hurts than the female complements. Moreover accidents on roads are comparatively high and more in risky climate conditions and during employed times. Analysis of accidents data situation at state level and city level shows that there is a lot of difference in accident risks across states and cities. Demise menace in 16 out of 36 states and union territories are complex than average all in India. Even though, risk of happening accidents on roads in India is slightly low in its urban cities, nearly 55% of the towns face complex accident risks than their uncountable counterparts. In total, whereas in many advanced countries including China, Brazil, United States road safety situations are generally improving but India faces a weakening condition. Without increasing actions and new visions, number of accidents happening on roads in India is probable to cross the number 2,60,000 by the year 2025. There is a crucial requirement to find out the worst condition in deaths on roads and injuries and to take suitable actions.***

I. INTRODUCTION

In this paper we are analyzing Road Safety data by using Hadoop tool along with some Hadoop ecosystems like HDFS, MapReduce, sqoop, hive and pig. By using these tools we can process no limitation of data, no data lost problem, one can get high throughput, maintenance cost also very less and it is an open source software, it is compatible on all the platforms since it is Java based. In Road safety, what precautions we need to take in order to decrease the happening of accidents in cities. By observe the road traffic routine wise what's happening in our real world, meanwhile crossing each stage of traffic issues which can be analyzed into several years and implemented in Transport research wing(TWR).

II. RELATED WORKS

Liling Li, Gongzhu HU IEEE(2017) examined the traffic accidents using datamining techniques that maybe reduce the accident rate. By using database saved in the name of Road safety that allows to reduce the decease by applying road safety programs at local and national levels. Those database scheme which describes the accidents on roads including conditions like roadway condition, number of persons involved and other data which could be useful for case evaluation, collecting added evidences.

(IRTAD)International Road Traffic and Accident Database, GLOBESAFE website for ARC networks are the finest sources for collecting data related to accidents. Using collected web data a map which is self organized for analyzing patterns is generated. Map classifies the data and deliver warnings as an audio. It was recognized that accidental rates are higher in crossings then other parts of roads.

III. DEFINITIONS

A.BIGDATA

The word Big Data has been functional to sets of data that has been growing so large in size and it becomes difficult working with using outdated database systems. There datasets whose size is outside the capabilities of normally using software tools and storage systems to store, capturing, managing and processing the data within a acceptable stipulated time. Sizes of datasets in Big Data are constantly increasing ranging from dozens of terabytes (TB) to petabytes (PB). Therefore, the difficulties related with big data include capturing, storage, searching, sharing, analysis, and visualizing. Today, initiatives are discovering with large volumes of detailed data so as to discover out the facts they didn't know early.

Therefore analysis of big data is where new and progressive techniques are applied on big data sets. Analysis founded on larger data sets reveals and leads to change. So, the larger the dataset size the more it becomes problematic to manage .In this segment, we are going to start by discussing the features of Big Data, as well as its importance. Normally, business benefits can be derived commonly by analyzing more complex data sets that require real time or non-real time capabilities. However, this leads to a need for new data designs, analytical methods, and tools. Therefore the following section will give detailed information on Big Data analytics tools and methods, in particular, starting with the storage of Big data and management of Big data, then moving to the processing of Big Data. It results with some of the several big data studies which have grown in usage with big data.

B.HADOOP

It is a open source software and processing framework that does work on data processing and storage of data for big data sized applications which are running in collected systems. It is one of the growing systems of bigdata technologies which are mainly used to provision advanced analysis initiatives including predictive analysis, data mining and machine learning applications. Hadoop can deal with structured data unstructured forms of data as well by giving more flexibility to user for processing, collecting and analyzing data compared to relational databases. Hadoop is primarily used for analysis uses, and it's additional features like process and storage of various forms of data which makes Hadoop as a decent fit for big data analysis applications. Big data analysis involves not only analyzing huge volume of data, but also structured data and unstructured forms of data as well such as records of internet clickstream and web server data and mobile application logs etc.

C . BIGDATA TOOLS INVOLVED IN HADOOP

The tools or schemes which are built up Hadoop in includes lot of other open source technologies that can match and extends its basic abilities of data analysis. Following are the list of tools in Hadoop:

Apache HIVE: HIVE tools are used to process the structured data present in Hadoop.

Apache PIG: pig is high level scripting language which is used in Hadoop.

Apache SQOOP: sqoop tool is used for transfer of bulk of data in among Hadoop and other structured and relational databases.

D.WHY HADOOP

The Apache Hadoop software is an opensource framework which allows large sized data sets processing and allow dispersed processing of large data sets across clusters in network. It was designed to operate like thousands of computers connected to single server, each local computer of the network providing local computing and storing of the data from data sets. Rather than depending on hardware to provide high availability, the library in the Hadoop designed in such a way that Hadoop highly available facility on top of group of computers, each of which may be results in failures. Hadoop provides applications with cost-effective storage for business applications and help businesses applications to easily have access to new data sources and different types of data to construct valued outcome from that data. Hadoop storage method is based on a distributed file systems that it works like basically ''mapping'' data to anywhere it is situated on any groups of computers. The tools which are available for processing data in Hadoop are frequently on the same servers where the information is situated, results in faster data processing. When data is directed to any individual node in the network, that data is then simulated to other nodes in cluster, which means that in case of failure, there will be copy available in another node for use.

E..HADOOP MAPREDUCE

It is a technique and called as core of Hadoop for tremendous scalability across thousands of Hadoop collection of computers on a product hardware. This perfect uses distributed algorithm to process large unstructured, structured data sets. MapReduce is a programming technique for performing distributed computations and this computations are based on java. Algorithm contains two important tasks, Map task and Reduce task. The process in Map step includes taking data sets as input and then mapping to another data sets, somewhere independent essentials are fragmented down into number of key/value pairs. Secondly, the process in reduce task includes which takes the map stage output and associate those data key/value pairs into a small set of keys, As in the name the algorithm follows the sequence which means reduce task is

achieved only after map task. The special benefit of Map Reduce algorithm is that it can easily scale the processing of data over multiple computing nodes. There are two primitives in Map Reduce model one is mappers and other is called as reducers. But, once after writing an application in the Map Reduce form, making those applications to run on thousands in a cluster is definitely a outline change. This scalability measure of MapReduce model what that has fascinated many programmers to use this model.

IV. METHODOLOGY

A. Preprocessing Road Accidents Database:

In this module, analyzing the data with different kinds of fields in Microsoft Excel then it converted into comma delimited format which is said to be csv(comma separator value) file and moved to mysql backup through Database.


B.Data Storage in HDFS:

In this module we are getting all those backup data which we have stored in MYSQL and importing all those data by use of sqoop commands to HDFS(Hadoop Distributed file System).Now the information are stored in HDFS were it is ready to get processed by use of hive.
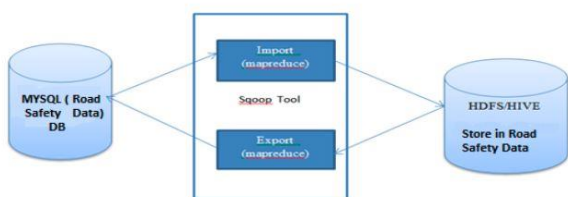

C.Analyse Query(HIVE):

In this module we are getting all those data from HDFS to HIVE by use of sqoop import command .were hive is ready to analyze . here in HIVE we can process only structured data to analyze. by extracting only the meaningful data and neglecting uncleansed data we can analyze the data in more effective manner by use of hive.
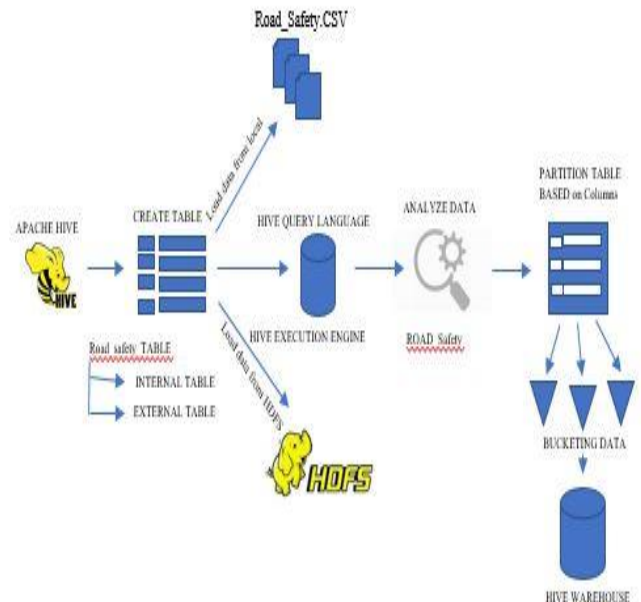

D.Scripting Process(PIG):

To analyze Road safety using Pig, programmers need to write scripts using Pig Latin language and execute them in interactive mode using the Grunt shell. All these scripting processes are within changed to Map tasks and Reduce tasks. After making a call to Grunt shell, one can run their Pig scripts in the shell excluding loading and storage tasks, but can perform all other operations. Pig statements take one relation schema as in data and yield extra relational schema as out result. After entering load statements in Grunt shell it will checks out syntactic and semantic errors. To have a view on the insides of schema use operator named Dump. Loading of data will be accepted out only afterwards executing Dump operation. For operations like grouping, filters etc, pig provides lot of built in operators.
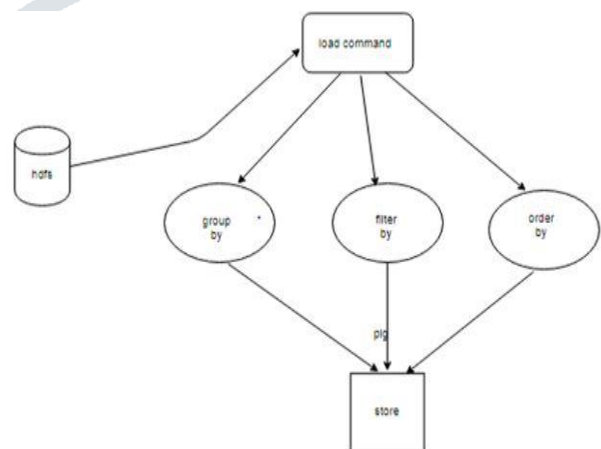

E.Parallel Process(MAPREDUCE)

MapReduce is a open source framework by using which one can write applications for processing enormous volumes of Road traffic Data in parallel. The MapReduce procedure comprises two significant responsibilities, namely Map task and Reduce task.

MapReduce algorithm takes three steps to execute which include map stage, shuffle stage, and then reduce stage. The mapper's work is to practice the input data. Normally the input statistics is in the process of file or directory which is warehoused in the HDFS. The information in the input file is then accepted to the mapper function one line after another. The map phase deals the data and generates numerous minor portions of data. Next step is the grouping of the Reduce phase and the Shuffle phase. The reduce task processes the information that derives from mapper. After completing processing it produces a new result set which is stored in HDFS.
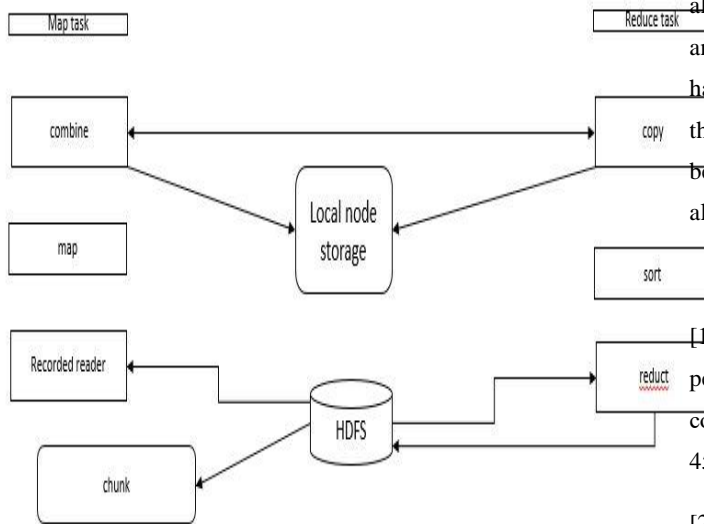
V.SYESTEM ARCHITECTURE
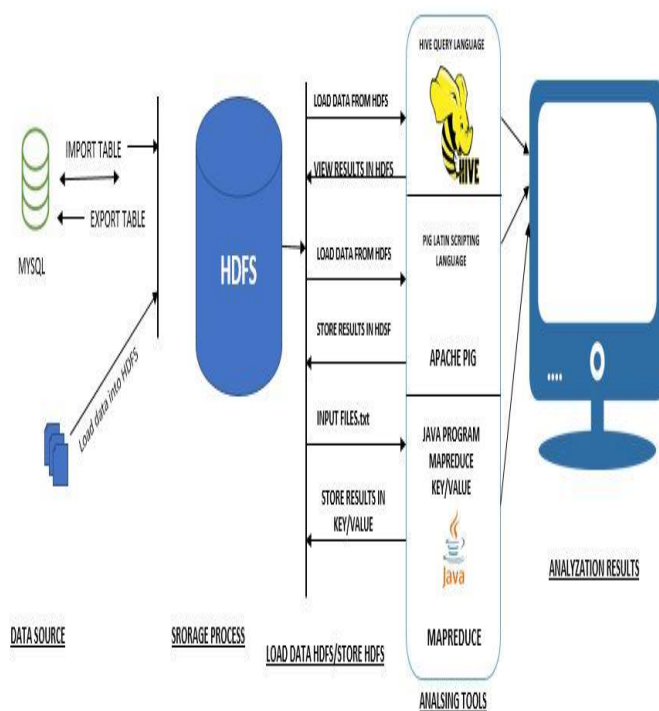
VI.CONCLUSION

We presented a Road accidents report by generating real time data backup hardly taken two years of data. Due to advancement of bigdata analytics its processing speed is fast and useful for prediction purposes which indicates to perform a proper maintenance.

VII.FUTURE ENHANCEMENT

Hadoop is opensource distributed processing framework which also overseen by apache software basis and it is used for storage and processing of enormous datasets with a collection of product hardware. The transport research wing able to find issues behind the road traffic and also it will be helpful to find the reasons beforehand by predict with the help of bigdata analytic report. We also use spark implementation for 100 times faster further.

REFERENCE

[1] M. Bagheri, M. Siekkinen, and J. K. Nurminen, "Cloud-based pedestrian road safety with situation adaptive energy efficient communication,"IEEE Intell. Transp. Syst. Mag., vol. 8, no. 3, pp. 45–62, Jul. 2016.

[2] S. Tao, V. Manolopoulos, S. R. Duenas, and A. Rusu, "Real-time urban traffic state estimation with A-GPS mobile phones as probes," J. Transp.Technol., vol. 2, no. 1, pp. 22–31, 2012.

[3] M. Khabazian, M. Mehmet-Ali, and S. Aissa, "Analysis of continuous communication availability in vehicular ad hoc networks," IEEE Syst. J.,vol. 7, no. 1, pp. 137–150, Mar. 2013.

[4] L.-W. Chen and P.-C. Chou, "BIG-CCA: Beacon less, infrastructure less, and GPS less supportive collision escaping based on vehicular sensor networks," IEEE Trans. Syst., Man, Cybern., Syst., vol. 46, no. 11,pp. 1518–1528, Nov. 2018. VanetMobiSim.   Accessed: Apr.   1,   2017.   [Online]. Available:http://vanet.eurecom.fr

[5] M. A. Javed and E. B. Hamida, "On the Interrelation of Security, QoS,and Safety in Supportive ITS," IEEE Trans. Intell. Transp. Syst., vol. 18,no. 7, pp. 1943–1957, Jul. 2017.

[6] S. Ucar, S. C. Ergen, and O. Ozkasap, "Multihop-cluster-based IEEE802.11p and LTE hybrid architecture for VANET safety message dis-semination," IEEE Trans. Veh. Technol., vol. 65, no. 4, pp. 2621–2636,Apr. 2016.

[7] BIGDATA Analytics. Applications and Theoretical Aspects: 14th Industrial Conference, ICDM 2014, St. Petersburg, Russia, July 16-20, 2014

[8] International Traffic Safety Data And Analysis Group(IRTAD) https://www.itf-oecd.org/IRTAD