ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 7, October 2015

TADOOP: Traffic Analysis and Traffic Solution Using Hadoop

Ashwini. E.M, Sunil. M.E

Department of CSE, PESITM, Shimoga, Karnataka, India

ABSTRACT: Traffic is the one of the problem in daily life, so proper measure has to be taken to control it. In this paper traffic density is generated using random number generator (instead of sensors) and analyze it using Hadoop Framework, an open source software implementation of the map reduce is used to big data analytics. Here we provide the live feed of traffic density in various nodes. In the live feed, the nodes with higher density are indicated by RED marker and nodes with lower density are indicated by GREEN marker on the map. Data generated is providing a platform to analyze various nodes statistically and graphically. It produces the useful information that can be transferable to the knowledge. Knowledge of traffic data helps end user to analyze the traffic density and used to make useful decision of a particular node over a period of time.

I. INTRODUCTION

Data that cannot be computed using traditional computational technique is referred as Big data. There are many tools available to analyze and compute Big data. One of such tool is HADOOP. HADOOP is a framework which runs on a special file system known as HDFS (Hadoop Distributed File System). Here any data stored is converted into blocks and distributed across the cluster nodes.

Hadoop provides a distributed file system and a framework for the analysis and transformation of very large data sets using the Map-Reduce paradigm. An important characteristic of Hadoop is the partitioning of data and computation across many (thousands) of hosts, and executing application computations in parallel close to their data. HDFS, the Hadoop Distributed File System, is a distributed file system designed to hold very large amounts of data (hundreds of terabytes or even petabytes), and provide high throughput access to this information. Files are stored in a redundant fashion across multiple machines to ensure their durability to failure and high availability to much parallelized applications.

Hadoop Map-Reduce is a software framework that process large amount of data, in parallel, in a fault tolerant manner. It is highly efficient and scalable in nature, and is used to process huge datasets. HDFS has master/slave architecture. As the data produced in traffic nodes are unstructured and are of large size, HADOOP framework is used to analyze and perform computations on these data easily. By analyzing the traffic density, we are going to provide statistical analysis and live feed of the traffic.

1.1 Objective

The main objective of our project is to analyze the traffic density by using HADOOP framework. Here we provide the live feed of traffic density in various nodes. In the live feed, the nodes with higher density are indicated by RED marker and nodes with lower density are indicated by GREEN marker on the map. In this project we are also providing a platform to analyze the data of various nodes statistically and graphically. It helps the end user to analyze the traffic density and other features of a particular node over a period of time. As it provides the live feed of the traffic, it also specifies the level of congestion at a particular traffic node.

1.2 Overview

This paper traffic data analysis using Hadoop was developed aiming at reducing the traffic density at various nodes. The idea is simple, instead of collecting traffic data using sensors, we generate traffic data using range of random numbers and these generated data are analyzed using Hadoop. Authenticated user can observe these variations of traffic density through live feed. The whole concept is divided into 3 modules.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 7, October 2015

First module is mainly concentrated on generation of traffic data using random numbers. These generated data are transmitted through mobiles. For these data transmission we have created mobile applications for respective nodes.
Second module is mainly concentrated on providing authentication and also user interfaces. Here we provide suitable features which are helpful for analyzing the data.
Third module is mainly concentrated on storage of traffic data in Hbase of Hadoop. These data in database are used to provide results for user queries.
Here data generated in first module will be sent to second module and third module through socket connections and also from second module to third module Php-Hive communication. Processing of these data is done in third module and is sent to second phase as results for user queries.

II. LITERATURE SURVEY

Itdescribes the background of existing traffic management system, its drawbacks and our proposed system as a better solution towards the problem

We compare how better our proposed system is compared to existing system.

2.1 Existing System

At present, Traffic information is commonly shared among traffic agencies by means of voice and/or data communications. To perform data communication between management centers, a common language and a frame of reference is required. There are many ways to analyze traffic data like manually counting each vehicle or through image processing and so on.
Manual counting requires a team to count each vehicle and update the statistics and a separate team is required to analyze these statistics. The disadvantage of these systems is

- Requires huge man power.
- Time consuming.
- Create Heavy Traffic Jams.
- If there is no traffic, the team still needs to wait at junctions.
- Lack of Traffic Information to users: Present traffic systems fail to provide traffic information including congested roads and alternate routes available in case of congestion.
- The cost involved in maintenance of field components is also high.
- It is also difficult to setup the system initially.

2.2 Proposed System

In our proposed system, traffic data from traffic nodes are assumed to be generated through random numbers (instead of sensors) and software used for traffic management system are open source software such as HADOOP, Code Igniter, Phone Gap, Leaflet etc.

User interface will be having an admin panel, where user can login with appropriate username and password. Thereby he will be provided with options such as statistical analysis, live feed, graphical analysis etc. Data from traffic nodes is stored onto HBASE of the HADOOP, from where they are taken into another data node of the HDFS cluster. In the data node they are analyzed using Map-Reduce technique. Finally statistical reports are generated based on user requirement.

III. SYSTEM ANALYSIS

It describes about the requirements of the system like functional, non-functional requirements, software requirements, use case diagram and activity diagram of the systems.

3.1 Functional Requirements

These are the requirements that are directly concerned with specific functions delivered by the software. These directly speak about the functionality that our software should provide after it is built. It should response exactly in the way we have mentioned.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 7, October 2015

- Implementation of big data characteristics (variety, volume, velocity)
- Visual representation of status of traffic node
- Graphical representation of traffic data
- Periodical updating of node information table
- Optimization of traffic node
- Providing suitable authentication to admin
- Statistical analysis of traffic node

Providing suitable user interface the software requirement specification is produced at the culmination of the analysis task. The function and performance allocated to software as part of system engineering are refined by establishing a complete information description as functional representation, a representation of system behavior, an indication of performance requirements and design constraints, appropriate validation criteria.

3.1.1 Sequence diagram for Interaction between Admin and Hadoop Server

In software and systems engineering, a use case is a list of steps, typically defining interactions between a role (known in UML as an "actor") and a system, to achieve a goal. The actor can be a human or an external system.

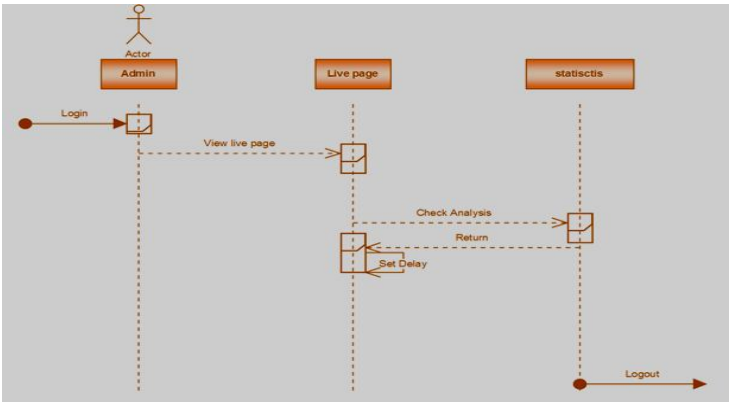

Fig 1 - Interaction between Admin and Hadoop Server

3.1.2 Use Case Diagram for Interaction between Admin and Analytical Server

A Use Case diagram at its simplest is a representation of a user's interaction with the system that shows the relationship between the user and the different use cases in which the user is involved. A use case diagram can identify the different types of users of a system and the different use cases and will often be accomplished by other types of diagram as well.

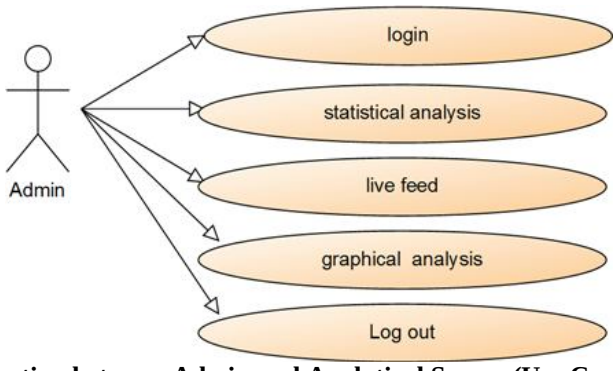

Fig.2 Interaction between Admin and Analytical Server (Use Case ID 1)

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 7, October 2015

3.1.3 Block Diagram for Entire System

Block diagram is a diagram of a system in which the principle parts or functions are represented by blocks connected by lines that show the relationship of the blocks.

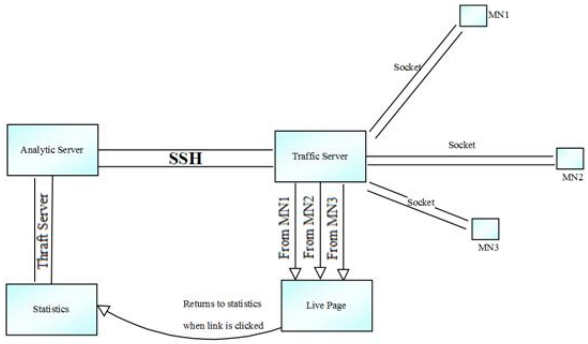

Fig3 Block Diagram for Entire System

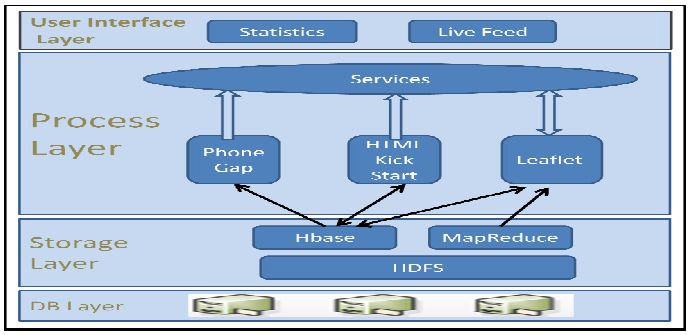

Fig4: Overall System Design

Fig 4 Shows architecture traffic management system. Architecture is composed of 4 layers. User interface layer is the entry point of the whole system, which provides services such as live feed, statistics and previous records. Process layer process the given input. It act as intermediate between user interface layer and storage layer. It consists of PhoneGap, HTML Kick Start and Leaflet. Storage layer retrieves data from database layer. It includes Hbase, MapReduce and HDFS. Database (DB) layer is used for storage purpose.

IV. RESULTS AND DISCUSSION

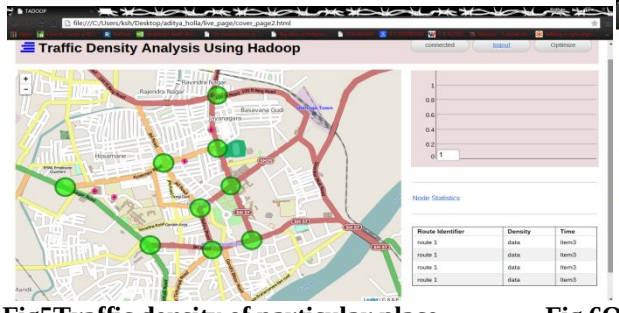

Fig5Traffic density of particular place

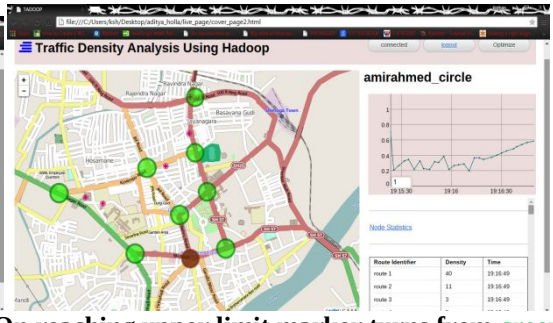

Fig 6On reaching upper limit marker turns from *green* to *red*

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 7, October 2015

Fig5 shows the live page provided to user. Live page contains live map, live graph and live tables corresponding to the values from the mobile application.
Fig 6 shows the density value of the particular place which is reaching the maximum limit set.

V. CONCLUSION

Huge amount of data can be efficiently accessed by Hadoop framework. In our project, as density data from traffic nodes is of huge amount; it is analyzed using Hadoop framework. Live feed of traffic density of various nodes are provided. In the live feed, the nodes with higher density are indicated by RED marker and nodes with lower density are indicated by GREEN marker on the map. We have also provided a platform to analyse the data of various nodes both statistically and graphically. Data is sent through mobile nodes. This project considerably reduces the time and effort for traffic data management when compared to existing system.

REFERENCES

1. https://www.youtube.com/watch?v=w6JlMY-MWBA – Hadoop Setup
2. http://tecadmin.net/install-oracle-java-8-jdk-8-ubuntu-via-ppa/ - JDK on Ubuntu
3. https://archanaschangale.wordpress.com/2013/08/31/installing-pseudo-distributed-hbase-on-ubuntu/ - Hbase on Ubuntu installation guide
4. http://www.tutorialspoint.com/hadoop/ - Hadoop installation guide
5. http://hortonworks.com/hadoop/ - Hadoop concepts in a Nutshell
6. https://en.wikipedia.org/ratchet - Ratchet Plug-in information
7. http://www.pacestar.com/uml/ -UML Diagram
8. http://www.99lime.com/ -HTML Kick Start main page
9. http://socketo.me/ - Web Socket home page
10. http://dygraphs.com/ - Dygraphs home page
11. http://leafletjs.com/ - Leftlet Mapping library
12. http://www.wampserver.com/en/ - Wamp Server Home page
13. https://jquery.com/ - Documentation on Jquery
14. https://developer.mozilla.org/en/docs/WebSockets/Writing_WebSocket_client_applications#Sending_data_to_the_server - components of Web Socket
15. http://cjihrig.com/blog/how-to-use-websockets/ - Web Socket guide
16. http://daneden.github.io/animate.css/ - Animation class
17. http://www.aptana.com/products/studio3/download.html - Aptana Studios
18. http://www.w3schools.com/ - HTML guide
19. https://www.youtube.com/watch?v=U12pqIxwuYE – Phone Gap Setup
20. http://goratchet.com/ - Ratchet guide
21. http://docs.phonegap.com/en/4.0.0/guide_cli_index.md.html#The%20Command-Line%20Interface – Phone gap documentation
22. http://socketo.me/docs/install - Socket installation guide
23. http://phonegap.com/blog/2012/08/10/introducing-phonegap-emulation/ - Introduction to Phone gap
24. http://speedutilities.com/su/tracker.php?process_name=php.exe&item=078e0214576dcc81a6ce&name=php.exe-is-not-recognized-as-an-external-or-internal-command – PHP execution details
25. https://build.phonegap.com/apps/1389181 - Phone gap app build