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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

Weather Forecasting using Map-Reduce

B.Anurag[1], Manoj Prakash[2], Vakeesh Kanna[3], Pelash Choudhary[4]

[1]Bachelor of Technology, Department of Computer Science and Engineering, SRM University, Chennai, India

[2]Bachelor of Technology, Department of Computer Science and Engineering, SRM University, Chennai, India

[3]Bachelor of Technology, Department of Computer Science and Engineering, SRM University, Chennai, India

[4]Bachelor of Technology, Department of Computer Science and Engineering, SRM University, Chennai, India

ABSTRACT: Big data is defined as the collection of large sets of data for computation purposes. The assets of big data stores large volume of data. The challenge of Map-Reduce framework in Hadoop is to extract the useful information from the stored sets of data. The efficiency of the extraction also plays a key role. Keeping these thing in mind there is a need to design a system that is at least close to moderate accurate and efficient system that uses the Mapreduce framework of Hadoop to predict and forecast the weather. Therefore the proposed system consists of four major modules: Data Attainment, Data storage, Query Resolution, Representation unit. The input of the data points is taken from Kaggle and National Climatic Data Centre, which is a website that hosts various datasets and kernels. The data sets are stored in HDFS (Hadoop Distributed File system). The main objective of our project is create an application that would predict climatic and environmental changes. Our project mainly focuses on citizens who live in areas which are prone environments disasters such as flood and also to assist farmers in adapting to climate-smart weather forecasting system using big data Approach which increase the income and productivity of farmers. Apart from predicting the day to day weather this project would create awareness and caution among people.This paper presents an enhanced automated prediction technique based on hadoop framework for efficient and scalable weather data analyzing and forecasting system.

KEYWORDS: Big data, Map-Reduce, Decision trees, prediction, Hadoop framework, HDFS

I. INTRODUCTION

Big data is basically defined as large sets of unstructured data which is further used for processing. In traditional data analysis only structured data can be processes unlike here where we can process unstructured data as well. Big data normally three defining properties of big data: Volume, Velocity, Variety.Volume refers to the amount of data, variety refers to the number of types of data and velocity refers to the speed of data processing.

Weather prediction is mainly important for business class, farmers etc. where they have to plan their work according to the weather. The prediction of weather is seeking a lot of attention in the recent times the best example of this can be the floods caused in the cities of Tamil Nadu in the year 2015. The drawback of the existing system are basically the techniques used in data mining in turn reflecting the efficiency of the system. In this project we use the map reduce framework from hadoop to process the offline data collected from various sources to predict the minimum and the maximum temperatures of the particular area based on the previous year's data.

1.1 **Aim and Objective:**

- To provide the maximum and the minimum temperature.
- To predict the climatic changes in a defined area.
- To be able to predict the rainfall and caution the people in case of emergency.
- To provide statistical analysis of the climatic change of an area over a period of time.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

1.2 Problem Statement

Due to the increase in global warming and over exploitation of natural resources resulted in drastic change the overall climate, therefore having adequate accuracy and efficiency is a common issue.

II. PROPOSED SYSTEM

The proposed model is to find the solution to find solution to the above mentioned problems. The wide variety of the given input data gives us a broader coverage of climatic change. Our proposed system can be explained in detail by using figure 1.

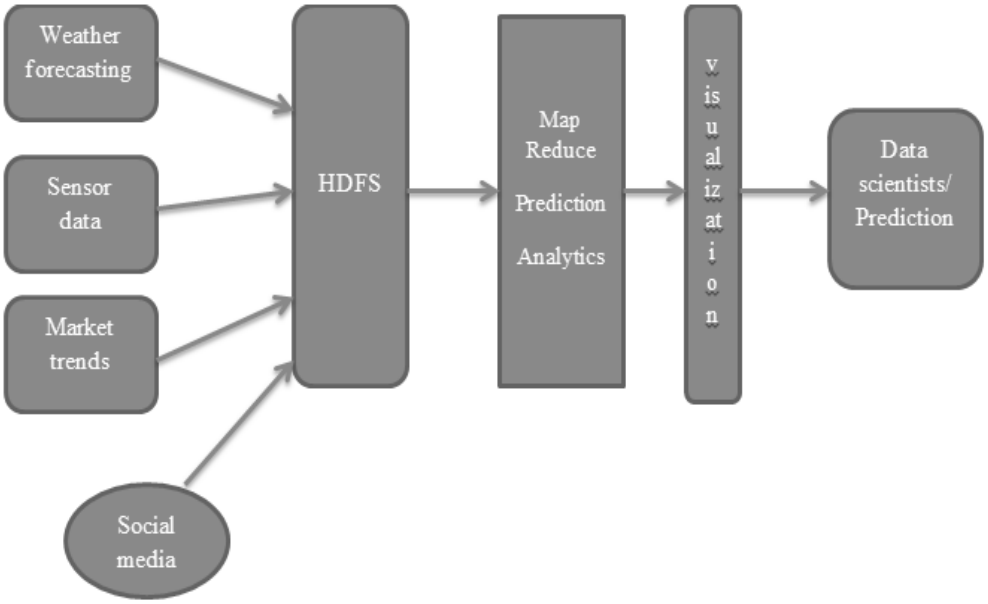


Figure 1: System Architecture

As shown in figure 1 the architecture of the proposed system using the components of Hadoop i.e. HDFS, Map reduce.

III. IMPLEMENTATION

The implementation basically consists of these following modules. The module flow diagram is given based upon the reference of figure: 1.

- Data Attainment
- Data Storage
- Query Resolution
- Data Representation

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

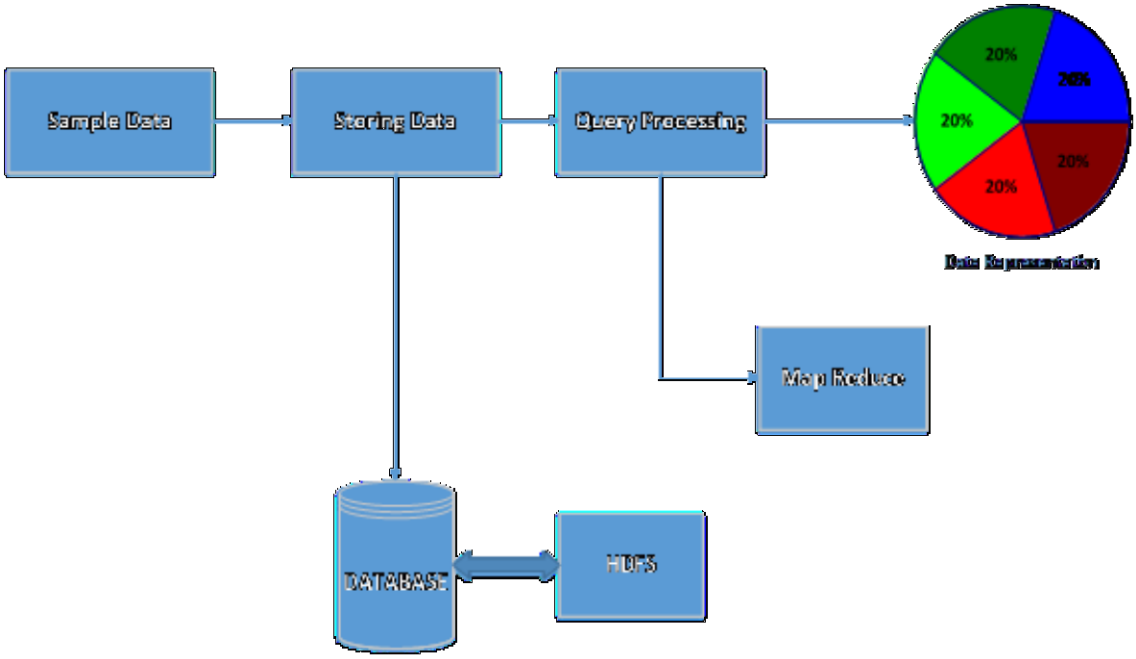


Figure 2: Module Flow of the proposed system

a) ***Data Attainment***

This module collects the data from the source. This collected data is first stored in a regular oracle or any other data base. In our case the date we are using is collected from an online form called Kaggle or NCDC as implementation of sensors is beyond our scope and budget. This collected data is later transferred to the storage unit and deleted from the database automatically.

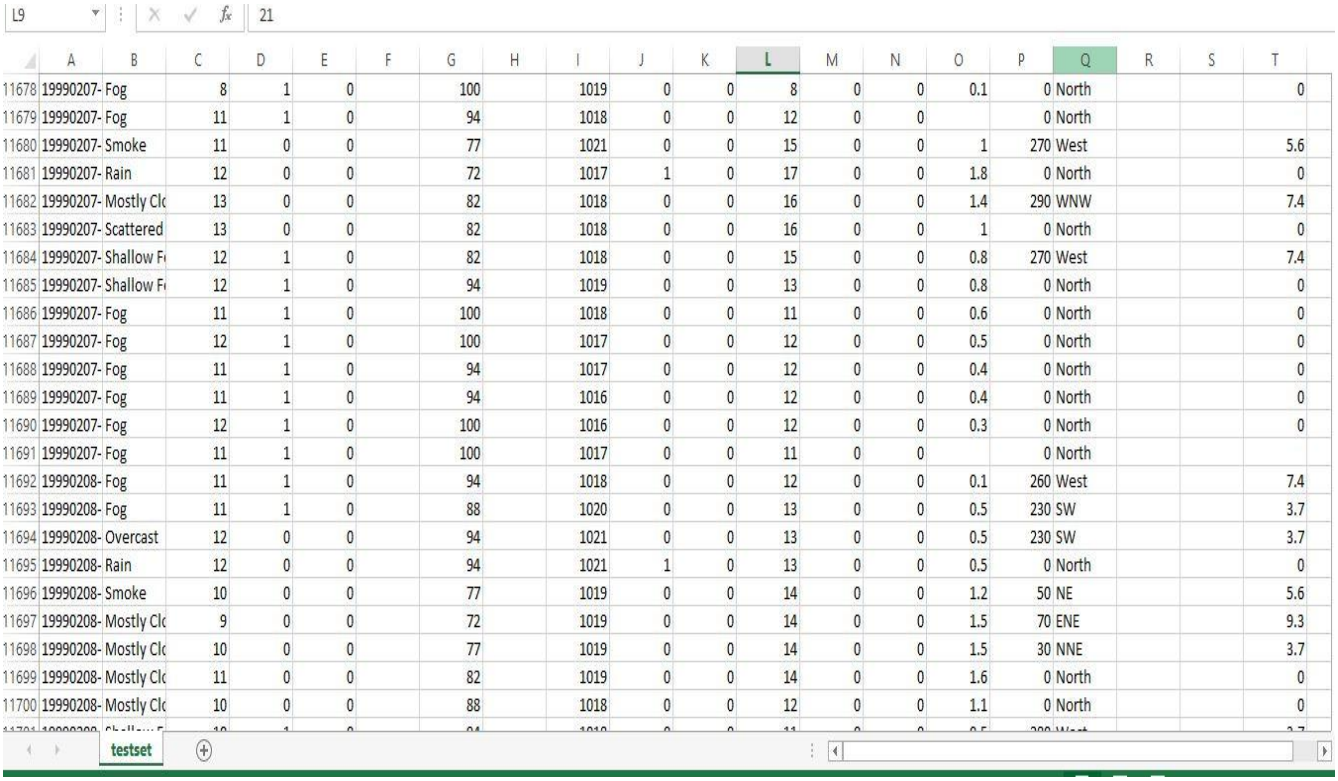
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
11678	19990207-	Fog	8	1	0		100		1019	0	0	8	0	0	0.1	0	North			0
11679	19990207-	Fog	11	1	0		94		1018	0	0	12	0	0		0	North			
11680	19990207-	Smoke	11	0	0		77		1021	0	0	15	0	0	1	270	West			5.6
11681	19990207-	Rain	12	0	0		72		1017	1	0	17	0	0	1.8	0	North			0
11682	19990207-	Mostly Clc	13	0	0		82		1018	0	0	16	0	0	1.4	290	WNW			7.4
11683	19990207-	Scattered	13	0	0		82		1018	0	0	16	0	0	1	0	North			0
11684	19990207-	Shallow F	12	1	0		82		1018	0	0	15	0	0	0.8	270	West			7.4
11685	19990207-	Shallow F	12	1	0		94		1019	0	0	13	0	0	0.8	0	North			0
11686	19990207-	Fog	11	1	0		100		1018	0	0	11	0	0	0.6	0	North			0
11687	19990207-	Fog	12	1	0		100		1017	0	0	12	0	0	0.5	0	North			0
11688	19990207-	Fog	11	1	0		94		1017	0	0	12	0	0	0.4	0	North			0
11689	19990207-	Fog	11	1	0		94		1016	0	0	12	0	0	0.4	0	North			0
11690	19990207-	Fog	12	1	0		100		1016	0	0	12	0	0	0.3	0	North			0
11691	19990207-	Fog	11	1	0		100		1017	0	0	11	0	0		0	North			
11692	19990208-	Fog	11	1	0		94		1018	0	0	12	0	0	0.1	260	West			7.4
11693	19990208-	Fog	11	1	0		88		1020	0	0	13	0	0	0.5	230	SW			3.7
11694	19990208-	Overcast	12	0	0		94		1021	0	0	13	0	0	0.5	230	SW			3.7
11695	19990208-	Rain	12	0	0		94		1021	1	0	13	0	0	0.5	0	North			0
11696	19990208-	Smoke	10	0	0		77		1019	0	0	14	0	0	1.2	50	NE			5.6
11697	19990208-	Mostly Clc	9	0	0		72		1019	0	0	14	0	0	1.5	70	ENE			9.3
11698	19990208-	Mostly Clc	10	0	0		77		1019	0	0	14	0	0	1.5	30	NNE			3.7
11699	19990208-	Mostly Clc	11	0	0		82		1019	0	0	14	0	0	1.6	0	North			0
11700	19990208-	Mostly Clc	10	0	0		88		1018	0	0	12	0	0	1.1	0	North			0

Figure 3: Dataset

b) ***Data Storage***

This module as the name suggests is responsible for the storage. The data is stored in HDFS format. This distributed file system can handle large files with sequential Read/write operations. The architecture of HDFS is shown in the below figure 4. In HDFS large data is broken down into small sets and then stored in various locations called Data nodes. These data notes are further connected to a Name node. HDFS provides parallel processing since the data I stored in a reluctant fashion also avoiding failure. The Name node acts as the server and controls the data nodes using hadoop framework.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

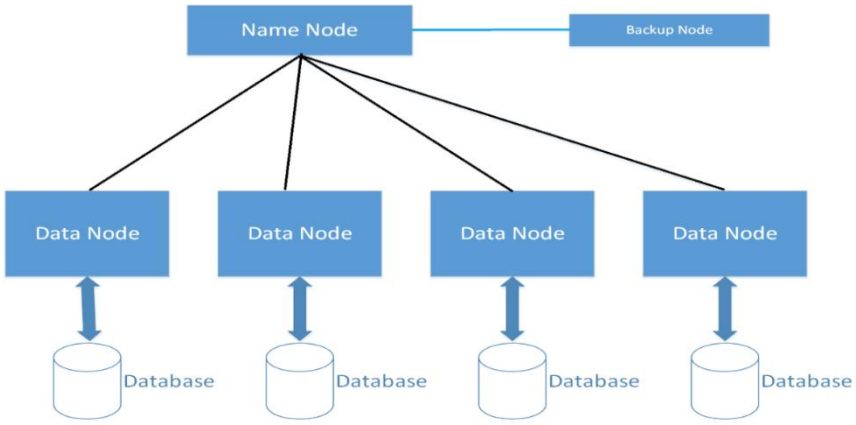


Figure 4: Architecture of HDFS

c) ***Query Resolution***

This phase includes the resolution of data stored in HDFS. This is done using map reduce framework. The basic architecture of map reduce is shown in figure 5. MapReduce framework works on a **split-apply-reduce-combine** strategy. The final output is the value which has more key-value occurrences.

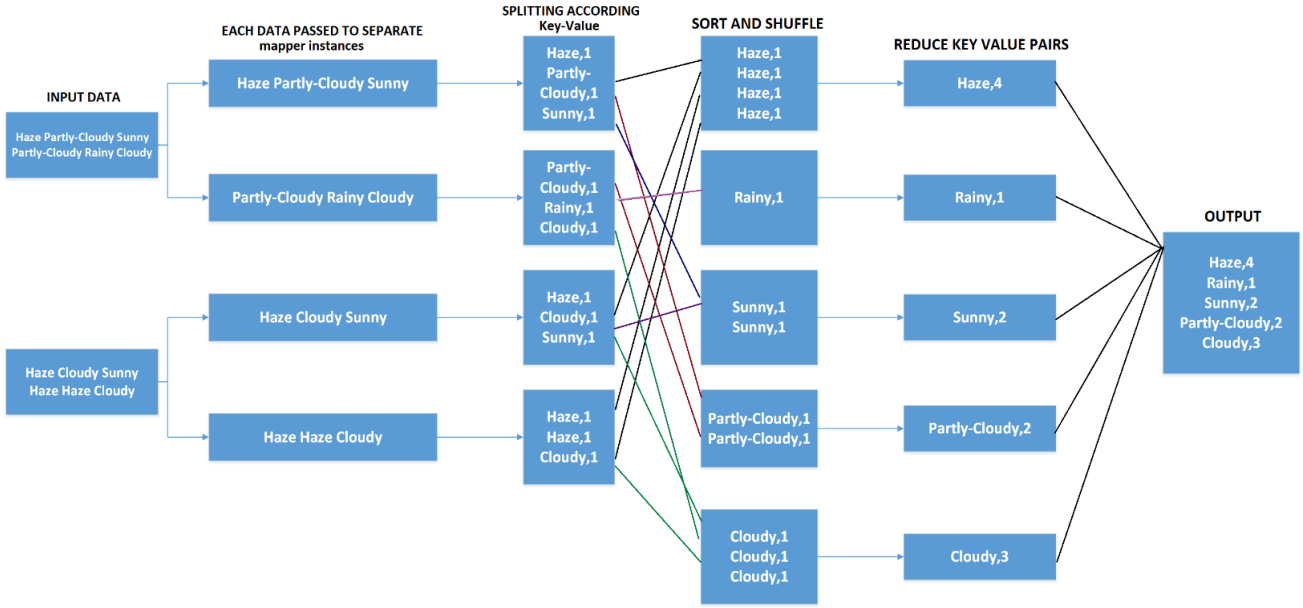


Figure 5: Architecture of Map reduce

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

From the above **figure 5,** the final output will be HAZE as it has more key occurrences. Map reduce is an efficient way of data resolution of large sets of data in a cluster. The map reduce program is composed of a map() procedure that performs percolating and sorting the required data into the data node. This module follows a split-apply-combine strategy. This module also uses HIVE which generally runs on the workstation, converting SQL query into series of map reduce jobs.

d) ***Data representation***

This module is used to represent the filtered data. The representation can be in done any preferred format like graphs, pie charts etc. This module is used for analysis of the result that is obtained. As per the datasets given above the sample presentation can be given as:

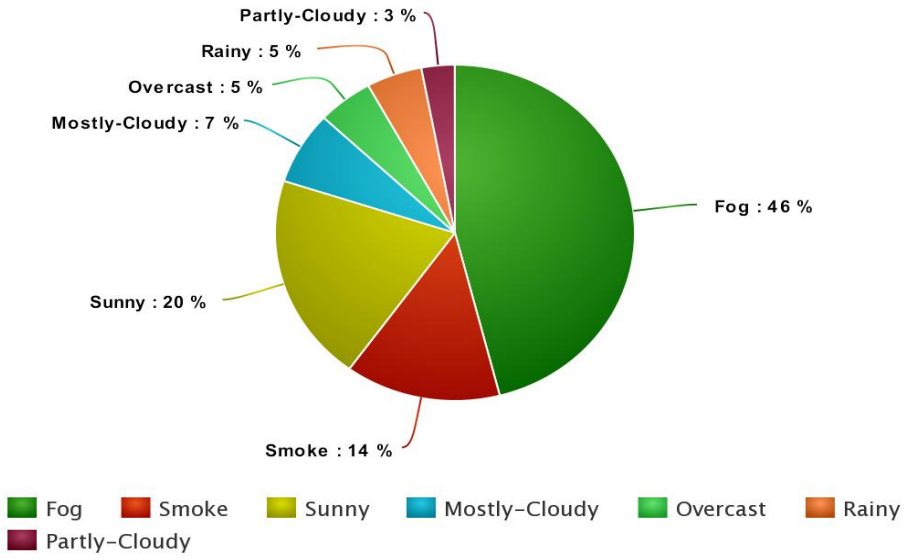


Figure 6: Sample result from figure 3

As show in figure 6 we can see that the forecast of the day would be foggy. In the similar way minimum and maximum temperature would also be predicted.

IV. RESULTS

In the proposed system uses the data sets from 1996 to 2017 of the area New Delhi, originally collected from Kaggle. This data is stored in the form of HDFS and map reduce is performed on the same as mentioned above to produce the desirable results i.e. the min and max temperature and predict the forecast of the day.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

V. CONCLUSION

We would like to conclude by saying that, in this paper we have proposed a system that is used to predict the forecast of the weather using map reduce which is a framework of Hadoop. The input of data can be taken from any trustable sources like sensors, online databases, social media etc.

Hence we predict the forecast of the day, Minimum and maximum temperature, rainy days etc. This might help the common people to plan their day accordingly.

VI. ACKNOWLEDGEMENT

This research was supported by our project guide Mr. Jayakumar S (Assistant professor(SL.G)), We thank him for providing us an opportunity to do the project research on Big-data Hadoop field at SRM University, Ramapuram, Chennai and by giving us the support and guidance for completing the project research duly.

REFERENCES

1) "Hadoop-based ARIMA Algorithm and its Application in Weather Forecasting", Authors: Leixiao Li, Zhiqiang Ma, Limin Liu, Yuhong Fan, International Journal of Database Theory and Application Vol.6, No.5 (2013), pp.119-132.
2) Dr. Doreswamy and Ibrahim Gad " BIG-DATA TECHNIQUES: HADOOP AND MAP REDUCE FOR WEATHER FORECASTING" , International Journal of Latest Trends in Engineering and Technology Special Issue SACAIM 2016, pp. 194-199.
3) Ramya M G, Chetan Balaji, Girish L, "Environment Change Prediction to adapt Climate-Smart Agriculture using Big-data Analytics" , International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 5, May 2015.
4) Dhanashri V.Sahasrabuddhe, PallaviP.Jamsandekar, "Data Structure for Representation of Big Data of Weather Forecasting", International Journal of Computer Science Trends and Technology (IJCST) – Volume 3 Issue 6, Nov-Dec 2015.
5) Riyaz P.A , Surekha Mariam Varghese, " Leveraging Map Reduce With Hadoop for Weather Data Analytics ", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 3, Ver. II (May – Jun. 2015), PP 06-12.
6) Rabi Prasad Padhy, "Big Data Processing with Hadoop-MapReduce in Cloud Systems", International Journal of Cloud Computing and Services Science (IJ-CLOSER) Vol.2, No.1, February 2013, pp. 16~27 ISSN: 2089-3337.
7) Tianyi Yang and Anne Hee Hiong Ngu, "Implementation of Decision Tree Using Hadoop Map Reduce", International Journal of nI ISSN: 2090-4924 Biomedical Data Mining.
8) Vidyullatha Pellakuri, Dr.D. Rajeswara Rao, "Hadoop Mapreduce Framework in Big Data Analytics", International Journal of Computer Trends and Technology (IJCTT) – volume 8 number 3– Feb 2014.
9) L . Greeshma, G. Pradeepini, "Big Data Analytics with Apache Hadoop MapReduce Framework", Indian Journal of Science and Technology, Vol 9(26), DOI: 10.17485/ijst/2016/v9i26/93418, July 2016.
10) Madhavi Vaidya, "Parallel Processing of cluster by Map Reduce", International Journal of Distributed and Parallel Systems (IJDPS) Vol.3, No.1, January 2012.
11) Ch. Shobha Rani, Dr. B. Rama, "MapReduce with Hadoop for Simplified Analysis of Big Data", Volume 8, No. 5, May-June 2017 International Journal of Advanced Research in Computer Science, ISSN No. 0976-5697.
12) Shvachko, Konstantin, et al. "The hadoop distributed file system". Mass storage systems and technologies (MSST), 2010 IEEE 26th symposium on. IEEE, 2010.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

BIOGRAPHY

B. ANURAG
BACHELOR OF TECHNOLOGY, COMPUTER SCIENCE AND ENGINEERING AT SRM UNIVERSITY, RAMAPURAM, CHENNAI (TN).

MANOJ PRAKASH. P
BACHELOR OF TECHNOLOGY, COMPUTER SCIENCE AND ENGINEERING AT SRM UNIVERSITY, RAMAPURAM, CHENNAI(TN).

VAKEESH KANNA T.I
BACHELOR OF TECHNOLOGY, COMPUTER SCIENCE AND ENGINEERING AT SRM UNIVERSITY, RAMAPURAM, CHENNAI(TN).

PELASH KUMAR CHOUDHARY
BACHELOR OF TECHNOLOGY, COMPUTER SCIENCE AND ENGINEERING AT SRM UNIVERSITY, RAMAPURAM, CHENNAI(TN).