# Big Data Analytics and Challenges: A Survey

Trupti G.Ghongade [1]

Research Scholar, Department of Computer Science & Engineering, SGBAU, Amravati, India[1]

**ABSTRACT**: Now we are living in the age of on-demand world with vast majority data. In this world of information, the term Big Data has emerged with new opportunities, tools and challenges to deal with the huge amount of data. The world is producing huge amount of data which is not in same nature. Cloud computing, big data analytics, communication technologies, Internet of Things are flooded with huge amount of data. Managing such massive is one of the big issue. But traditional data analytics, tools may not be able to handle such large quantity of data. For managing data different tools are available. One of the best tool is Hadoop for managing data and solves the problem by making it useful for the big data analytics. This paper focuses on Introduction of Big data and Hadoop and challenges associated with big data.

**KEYWORDS**: Big Data, Hadoop, Heterogeneity, Scale, Analytics

## I. INTRODUCTION

In digital world, data are generated from various sources like social media, business,healthcare systems,banking sector etc.that leads to growth of big data. The arrival of big data is from multiple sources at a frightening velocity, volume and variety. We need optimal processing power, analytics capabilities and skills to extract meaningful value from big data [1].Scientists, business executives, practitioners of media, and advertising and governments alike regularly meet difficulties with large data sets in areas including Internet search, finance and business informatics.

Data sets grow in size in part because they are increasingly being gathered by cheap and numerous information-sensing mobile devices, aerial (remote sensing), software logs, cameras, microphones, radio-frequency identification (RFID) readers, and wireless sensor networks.These are available in structured, semi-structured, and unstructured format in petabytes and beyond [2]. Generally, Data warehouses have been used to manage the large dataset. In this case extracting the precise knowledge from the available big data is a foremost issue. Most of the presented approaches in data mining are not usually able to handle the large datasets successfully. The key problem in the analysis of big data is the lack of coordination between database systems as well as with analysis tools such as data mining and statistical analysis[2]. This paper This paper focuses on Introduction of Big data and Hadoop and challenges associated with big data

## II. CHARACTERISTICS OF BIG DATA

There are three main features that characterize big data:

### A. Volume

Volume means scale of data or amount of data generated per second in enterprise. Now a day's data is increasing from gigabytes to petabytes.

### B.Velocity

Velocity means speed at which data is generated and proceed on demand. In other words, velocity means speed of data.

### C.Variety

Variety includes the different formats and types of data such as text, numerical, images, video and audio type of data well as the different kinds of uses and ways of analysing the data.

### III. HADOOP

Hadoop is open source software used to process the Big Data. It is commonly used by organizations/researchers to analyse the Big Data. Hadoop is influenced by Google's architecture, Google File System and MapReduce. Hadoop processes the large data sets in a distributed computing environment. An Apache Hadoop ecosystem consists of the Hadoop Kernel, MapReduce, HDFS and other components like Apache Hive, Base and Zookeeper [3].
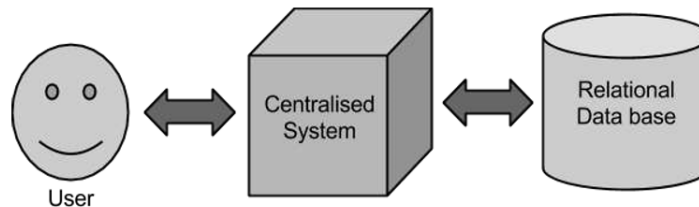


Fig.1 Actual Working of Hadoop

Traditional approaches worked fine with the application which processing the small amount of data. But it is not suitable for the application which is having the large amount of data to process. To solve this problem Google invented new technique called Hadoop. In this technique the hadoop uses the mapReduce algorithm, in which it divides the task into the small parts and assigns it to separate computer and collect the result in the dataset. Hadoop runs applications using the MapReduce algorithm, where the data is processed in parallel with others. In short, Hadoop is used to develop applications that could perform complete statistical analysis on huge amounts of data [4]

### IV. BIG DATA CHALLENGES

In recent years' huge amount of data is generated through various domains like social media, healthcare, banking sector, interdisciplinary scientific research etc. So, Heterogeneity, scale, timeliness, complexity, and privacy problems are coming with Big Data.

*A. Heterogeneity*

Generated data is always come up with two forms that is structure or unstructured from of data. When we deal with big data, data must be in structured form to analyse it. This is the biggest problem that analysers have to cope with

*B. Scale*

Big data contains huge amount of data and large size of data set. In past decades it is possible to deal with such data by increasing the power of processor. But data is continuing to increasing and power of processors is static. So the scalability is another biggest challenge in big data.

*C. Timeliness*

Another challenge with the speed is size. If the data sets are larger in size the longer it will take to proceed and analyse the data.

*D. Privacy of data*

Privacy of data is another big problem with big data. There is huge risk of security associated with big data. Security of big data can be enhanced by using techniques such as encryption, authentication and authorization.

### V. CONCLUSION

In recent years' data are generated at highest speed. Analysing such type of huge data is the biggest challenge. In this paper we survey characteristics of Big Data, Hadoop as best solution to the big data. Hadoop can handle the huge amount of data, it is very cost effective, and it can handle huge amount of data so processing speed is very fast, and also it can create a duplicate copy of data in case of system failure or to prevent the loss of data a well as and challenges associated with big data.

## REFERENCES

1. Mrs.Mereena Thomas,"A Review paper on Big Data", International Research Journal of Engineering and Technology (IRJET),Vol 2 Issue 9,Dec 2019

2. D. P. Acharjya, Kauser Ahmed P, "A Survey on Big Data Analytics: Challenges, OpenResearch Issues and Tools", (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 7, No. 2, 2016.

3.Rahul Beakta,'" Big Data And Hadoop: A Review Paper,Research Gate,Vol 2 Issue 2 ,2015

4.Rupali Jagadale, Pratibha Adkar,"A Review Paper on Big data & Hadoop", International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 6 Issue: 5

5. Chun-Wei Tsai1, Chin-Feng Lai2, Han-Chieh Chao1,3,4 and Athanasios V. Vasilakos5*," Big data analytics: a survey",Journal of Big Data Springer,2015.

6. http://blogs.mindsmapped.com/bigdatahadoop

7. https://www.knowledgehut.com/blog/bigdata-hadoop/top-pros-and-cons-of-hadoop