IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified

Vol. 6, Issue 6, June 2017

Review Paper on Big Data: Applications and Different Tools

Simranjot Kaur[1], Er. Sikander Singh Cheema[2]

Research Student, M. Tech. (CE), Department of Computer Engineering, Punjabi University Patiala[1]

Assistant Professor, Department of Computer Engineering, Punjabi University Patiala[2]

Abstract: The term Big Data play very important role in various fields. It is the collection of data in the large amount which has to store, capture, manage and process. For this process we have introduced various tool and techniques in this paper. Big data is a term which is being used almost in every field. Big data is basically used in various fields just like Medical department, Business, Agriculture, Education, BDA, Finance etc. We have also introduced the applications of Big Data in different fields. This paper will give the detailed explanation of different tools and different applications in which Big Data is used in very vast amount.

Keywords: Big data, Big data Analytics, Hadoop, MapReduce, HDFS, Data Mining.

1. INTRODUCTION

Big data is the term which can be described in the structured, semi-structured and unstructured form of data. The data which is in the proper format or tabulation form is referred as the structured form of the data [1]. The data which contains the images in addition to the text data is come under the semi-structured for of the data. The data which contains the videos, images and text etc. and not in the proper format is comes under the unstructured form of the data. This type of data contains the records in billion forms which are not so easy to process, manage and store with the help of relational databases. So Big Data Analysts need some other tools and techniques for this purpose [2]. So this is very difficult task for Big Data Analysts to deal with tools and techniques.

A. Big data analytics: It indicates the process of managing, collecting and storing of large amount of data sets in order to discover useful information [3]. The main aim of the big data analytics is to make businessman, scientists and other professionals to make more accurate decisions to their large amount of data.

B. Characteristics of Big Data: On the basis of large amount of data, it can be characterized by 3 V’s.

1. Volume: Volume of data refers to the size of data which lies between the Petabytes and Exabytes.
2. Velocity: Velocity defines the speed of data flow. The flow of data is continuous. It is used to analyze the increase in profit of business before the information lost.
3. Variety: variety refers to the sources of data or we can say that different types of data such as structured and unstructured data [1].

2. ROLE OF BIG DATA

In Data Mining: Decision trees are used in the data mining concept. The process of extracting the useful data from the Big data is called data mining technique. This technique is used by the big data analysts for the processing and managing of data.

1. In BDA (Big Data Analytics Applications): It is the new category of the software application. Hadoop technique is used in this application to analyze the data. In this technique first of all, the analysis is done by the developers in a small scale by talking some small samples. After that they use the whole data in the analysis process [6].
2. In clustering: Clustering is also a technique which is used in managing and processing of Big data. K-means algorithm is used in this technique. It helps to identify the similar groups of data and collaborate in one cluster.
3. In banking: In the banking sector, Big data is used at large scale. As we know that banking sector includes important information regarding customer’s earning, savings, insurance policies etc. So the banking sector uses the Big data for the sake of security purposes [13].
4. In Agriculture: In this field, big data is used at very rapid rate. As we all know that crops are totally dependent upon the weather conditions. So, Big data helps to identify the weather conditions by applying different algorithms and tools [7].

DOI10.17148/IJARCCE.2017.6694

5. In Medical field: Now a day's medical department is also using big data in term of storing the large amount of patient information in a particular way. Record of patients, workers, staff members, medicines etc. needs a specific pattern to store and process [9].
6. In Education: Every college or university have vast amount of data which is very difficult to manage and process. Even storage of date is a difficult task. Every educational department is using big data techniques to store and process the information of student's record and record of teachers and many more [7].
7. In weather forecasting: Big data is the term which is widely used in the field of weather forecasting. With the help of K-means clustering and R tool, we can easily predict the weather for upcoming days. This method is also applicable in various fields in which we have to predict the upcoming data [2].
8. In smart phones: Big data is the term which is widely used in every field. In today's era, every person have smart phone in their pockets. As in the I phones, there are some applications in which we have to store the person to person data in order to satisfy the android application's requirements. For example, in I phones, there is an app in which we have to store the images for security locks as facial recognition. So the data which is being stored in that application might be in millions. So Big data is used in that type of applications also in order to simplify the uses of modern applications.
9. In Conservation: Big data helps to keep the data in an isolated, merged and proper format which gives the benefit in business department.
10. In Finance: Big data is used to manage the wide range of data including credit cards, checking, savings, mortgages, and investment data [4].

3. TOOLS AND TECHNIQUES

1. Hadoop: Hadoop is one of the important techniques which is a programming framework and developed by Google's MapReduce. Hadoop is used to handle the large amount of data with the help of divide and conquer method. Hadoop functionality includes two steps:
Map(): The main function of Map is to divide the data into number of sub parts.
Reduce(): The function of Reduce is to collect all the answers generated by the sub parts and combine them to make an appropriate output [12]. There are some Hadoop technologies which are used:
a.) Apache PIG
b.) Apache HBase
c.) Apache Hive
d.) Apache Sqoop
e.) Apache Flume
f.) Apache Zookeeper

2. HDFC: Hadoop distributed File System is a client-server architecture which is used to process the large amount of data. It contains one Name Node and multiple Data Nodes. HBase, Pig, Hive, Sqoop, Chukwa, Flume are some other components of Hadoop [8].

3. HPCC: It is an open source platform defined by the users. This system is used to manage complex problems. It is not only single platform system but also a single architecture and single programming language which are used to process the data. There are some components of HPCC:
a.) HPCC data refinery
b.) HPCC data delivery
c.) Enterprise Control Language

4. Grid Computing: Grid computing is the technique in which computers are interconnected and share resources to one another. In Big data, Grid Computing is used with the help of Hadoop. In some cases, hadoop is unable to store the huge amount of data. But with the help of Grid Computing, It can store the data by parallelizing the data in different data nodes [14].

5. Data mining: Data mining is the technique which is used to extract the useful information from the large set of data. Now a day's there are various data mining techniques like R tool, KNIME, WEKA, KEEL, RAPIDMINER, ORANGE etc.

6. R tool: R is the free software programming language which is used for the statistical computing and graphics. We can say that R is the platform for statistical computing. R was first released in 1997. All the statisticians can do the complicated analysis with the help of R without having knowledge of functions of computing system. C, FORTRAN and R languages are used in this tool [10].

DOI10.17148/IJARCCE.2017.6694

7. KEEL: KEEL is Knowledge Extraction based on Evolutionary Learning. It is application software of machine learning tools. It helps to solve the data mining problems with the use of evolutionary algorithms. All the preprocessing, post processing techniques for data manipulation are inbuilt in this software [14].

8. WEKA: WEKA is Waikato Environment for Knowledge Analysis. In order to solve the data mining problems WEKA works on the machine learning algorithms. In this tool, we use the Java codes in order to apply these algorithms on the data sets or we can use these algorithms directly to the data sets. It is the platform independent software.

9. KNIME: Konstanz Information Minor is an open source data analytics and integration platform. It is used in pharmaceutical research, CRM customer data analysis, business intelligence and financial data analysis. KNIME was released on 2004. It is written in Java language. There are some limitations of KNIME like, no error measurement method and no wrapper method.

10. RAPIDMINER: RAPIDMINOR is the software platform which is used for industrial and business applications. Apart from that, it is also used in the field of training, research, rapid prototyping, education, application development etc. In order to support the data mining processes, it uses the client/ server model. RAPIDMINOR is a machine learning environment which is used to solve huge number of learning problems.

11. ORANGE: ORANGE was developed in 2009. C++ and Python are used to implement this application. As it contains the set of components for data preprocessing, it refers as the component-based data mining software. ORANGE is very easy to learn for the programmers as it is written in python [11].

4. CONCLUSION AND FUTURE SCOPE

In recent years, data is produced at very rapid amount from almost every field. In this paper, we had surveyed on different types of data produced from different fields and how this data is applicable for different departments like agriculture, education, medical, finance, business etc. in order to process and store this huge amount of data, we have studied various tools and techniques. These tools and techniques play an important role to manage such a huge amount of data in a proper way. If we focus on the problems related to big data then storage capacity and processing of data are the main problems. Further we will discuss about the tools that are used to store and process the large amount of data.

REFERENCES

[1] Arti Chandani, M. M. (2015). BANKING ON BIG DATA: A CASE STUDY. ARPN Journal of Engineering and Applied Sciences , 4.

[2] Basvanth Reddy, P. B. (2016). Weather Prediction Based on Big Data Using Hadoop Map Reduce Technique. International Journal of Advanced Research in Computer and Communication Engineering , 5.

[3] Dr. Siddaraju, S. C. (2014). Efficient Analysis of Big Data Using Map Reduce Framework. International Journal of Recent Development in Engineering and Technology , 5.

[4] Harashawardhan S. Bhosle, P. D. (2014). A Review paper on Big Data and Hadoop. International Journal of Scientific and Research Publications , 7.

[5] Harikesh S. Nair, S. M. (2014). Big Data and Marketing Analytics in Gaming: Combining Empirical Models and Field Experimentation. Big Data Marketing Analytics (p. 43). Chicago-Booth: Vineet Kumar and Puneet Manchanda.

[6] Kalpana Rangra, D. K. (2014). Comparative Study of Data Mining Tools. International Journal of Advanced Research in Computer Science and Software Engineering , 8.

[7] Kuchipudi Sravanthi, T. S. (2015). Applications of Big data in Various Fields. International Journal of Computer Science and Information Technologies , 4.

[8] M. R. Bendre, M. R. (2015). Big Data in Precision Agriculture : Weather Forecasting for Future Farming. 1st International Conference on Next Generation Computing Technologies, (p. 7). Dehradun.

[9] P.Surya, D. A. (2016). The role of big data analytics in agriculture sector : a survey. International Journal of Advanced Research in Biology Engineering Science and Technology , 9.

[10] Patil, S. (2016). Big Data Analytics Using R. International Research Journal of Engineering and Technology , 7.

[11] Pradeepa. A, D. A. (2013). Significant Trends of Big Data Analytics in Social Network. International Journal of Advanced Research in Computer Science and Software Engineering , 5.

[12] Raghu Garg, H. A. (2016). Big Data Analytics Recommendation Solutions for Crop Disease using Hive and Hadoop Platform. Indian Journal of Science and Technology , 6.

[13] Utkarsh Srivastavaa, S. G. (2015). Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks. ELSEVIER , 11.

[14] Yuvraj S. Sase, P. A. (2014). Big Data Implementation Using Hadoop and Grid Computing. International Journal of Innovative Research in Science, Engineering and Technology , 6.

DOI10.17148/IJARCCE.2017.6694