ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 10, March 2018

Big Data-Concepts, Techniques and Challenges

Hargurjeet Kaur

Assistant Professor in PG Department of Computer Science and Applications, SGGJGC, Raikot, Punjab, India

Email ID: hargurjeetmatharu@gmail.com

Abstract: Big data is the pristine term that refers to collection of huge volume of datasets. Data is as large and complex as there is a difficulty to store and process it by traditional data-handling techniques and tools. Data is not a different thing than any other term of data. Here big is a keyword that is used along with data. Big data identifies the grouped datasets due to its huge size and complexity i.e. data is either in Terabyte/Gigabyte/Petabyte/Exabyte or larger than it. Huge amount of data cannot be stored and managed with contemporary techniques or data mining tools. Big data mining is the capability of acquiring and organizing valuable information out of large volume of datasets due to its variability and velocity. There was no possibility before to do this. This study paper intends to explain the concepts of big data, applications technologies of big data, challenging issues and its related work.

Keywords: Big Data, 5 V's, Datasets, Hadoop, HDFS, MapReduce

I. INTRODUCTION

Now a days, different people and systems encumber the internet with urge amount of data. This amount of data is added everyday as a big space represented by the internet. Today is the era of internet. Anything, which we want to know about or anonymous to us, we Google it and in bit of time, we get a lot o links as a result. Google accommodates the enormous information. There are various vendors who have brought out different technologies.

Markets including Microsoft, Amazon, IBM, etc. focus on managing and handling big data. In big data, there are 3 types of data: structured data i.e. relational data, semi structured data i.e. XML data and unstructured data i.e. text, word, PDF.

II. CONCEPTS

Big data cannot be handled with the techniques that hold standard data management. This is because of irregularity and uncertainty of the manageable alliance. The concept of big data is simply defined by describing the 5 V's:

1. Volume

It concerns the large quantity of data that is collected by a company continuously. In the beginning, storage expenses were a big problem. However a better replacement needs to be evolved. It is mainly referred to the size of data that is either in terabytes or petabytes.

2. Velocity

Velocity refers to the speed of data in the way it is processed. Sometimes immediate response should be required in case of important data. Therefore, in big data for quick processing must be needed for efficiency.

3. Variety

Variety refers to the kind of data that may be structured or semi structured or unstructured. Because of heterogeneous nature, data comes in different formats and different types.

Structured data includes relational data while unstructured data may have text, words, audios, videos etc.

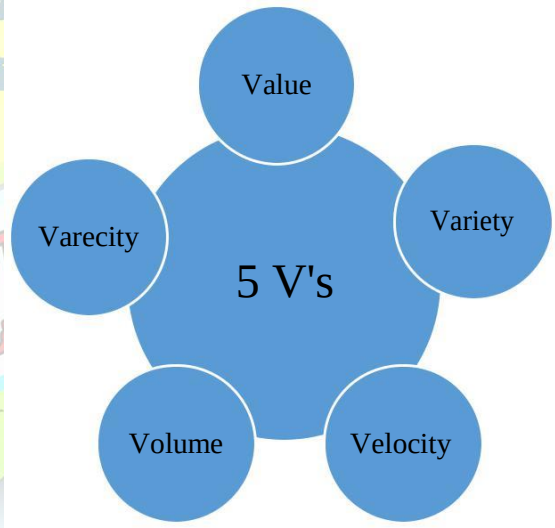

Fig. 1 5V's of big data

4. Value

Value is considered as the most powerful 'V' in big data because value is important in businesses, IT infrastructure systems for storing large quantities of values in datasets.

5. Veracity

Veracity refers to the trust of used information by the leader in order to take decision. It is the degree in which truth or fact, precision, accuracy are confirmed. While dealing with high volume, velocity and variety of data, not all data are 100% accurate, there will be chances of dirty data.

ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 10, March 2018

III. BIG DATA TECHNOLOGY

For the management of big data, there are various tools and techniques are used. These tools help in acquisition o data and analysis o that data. These tools are developed using Hadoop i.e. written in java language. Hadoop is constructed using two important projects:

1. Hadoop Distributed File System (HDFS)
2. MapReduce
3. Hive

1.HDFS:

This is based on distributed file system. Basically HDFS is designed for large clusters having commodity hardware. Here the word large used for 10 GB to 100 GGB or more than that. HDFS is totally based on batch processing instead of interactively used by users. File is creating once but used for many times as per requirements in Hadoop applications. IT detects the allure at application layer and then handles that detected failure.

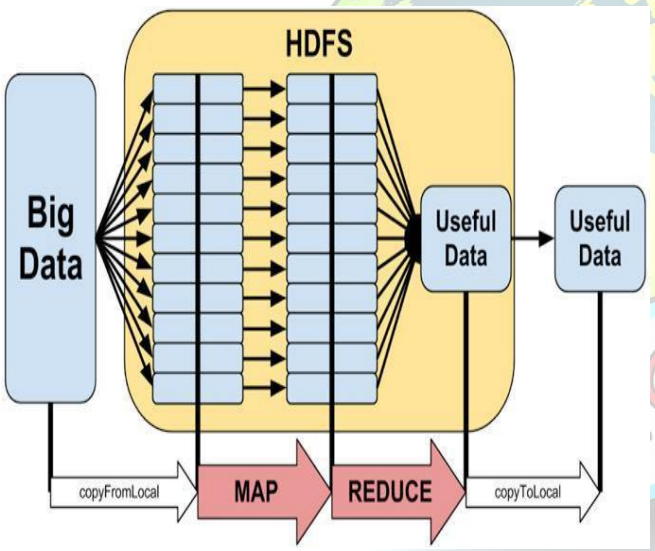

Fig. 2 Tools and techniques of big data

2. MapReduce:

Being an important programming construct, it generates huge datasets. MapReduce application's input data can be semi-structured or unstructured. Th output of MapReduce application is paired set of <key,value>. As name suggested, in MapReduce , the algorithm uses two main functions "Map" And "Reduce". Map function produces intermediate values of <key,value>, and the Reduce function combines all intermediate values.

3. Hive:

Hive is a decentralized system for developing applications by using local systems on networks. Apache hive as a component of data warehousing is an element of Hadoop ecosystem which uses cloud offers a language. This language is called HiveQL. HiveQL translates SQL-Like queries into MapReduce applications automatically. Apache Hive has various applications such as IBM, SQL, DB2 and oracle. Hive architecture is decomposed into Meta data for data storage, MapReduce.

IV. BIG DATA CHALLENGES

1. Proper Understanding of Big Data

The understanding of big data is very important. In order to discover the foremost scheme for a company, it is important that the data must be analysed properly. Time interval for this analysis is also important because sometimes it needs quick response in order to take decision for any in the business infrastructure.

2. Privacy and Security of Big Data

To maintain the privacy and security is one of the important challenges for Big Data. Because of its complex nature it is a difficult task for company to organise the data on various privacy levels and then registers the according security. Nowadays companies are dealing across the countries and continents and therefore differences in privacy principles have to be taken into consideration.

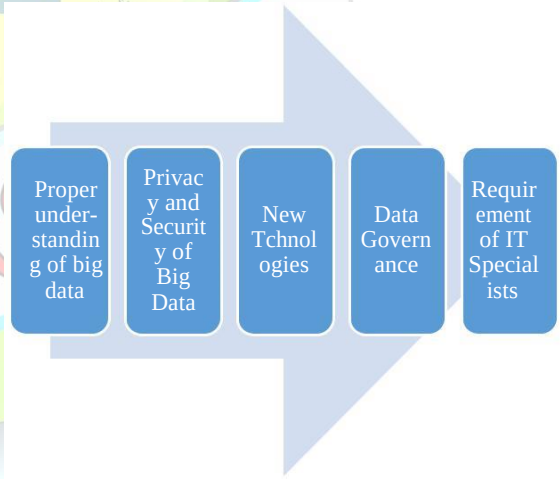

Fig. 3 Challenges of big data

3. New Technologies

New technologies are developing day by day which is also a big challenge for big data. Now a day, for organizations, big data is new technology. Therefore, it is mandatory for these organizations to make in use the new discovered technologies immediately. It is a going trend that brings competition among businesses.

4. Requirement of IT Specialists

It is also an obvious challenge associated with big data. According to study o big data, there is requirement for approximately 190,000 more employees along with analytical skills only in the United States. This statistics

ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 10, March 2018

proves that in order to achieve big data goal. For this, an organization has to train either existing employees or hire new employees having talent on the new field.

5. Data Governance

It is usually same type of data is getting by the organizations from different systems. Data from different systems may not agree. The process of getting those records to agree and for ensuring the accuracy is called data governance. For this, companies have to set up a team of employees to handle data governance and develop a number of policies and procedures.

V. IMPORTANCE OF BIG DATA

The Importance of big data cannot be defined by the size of data in a company has but it revolves around the efficient utilization of data. A company has more potential to grow when it utilizes its collected data in more efficient way. The data can e taken by any source. In this perspective, big data has following importance:

1. Cost Saving:

Tools of big data such as Hadoop and cloud based analytics can save cost in business while storing huge amounts of data. This helps in Progress of business.

2. Time Reducing:

Tools such as Hadoop and Cloud-Based Analytics can make quick decision. These decisions are based on learning.

3. Market-Conditions Understanding:

Analysis of big data can helps in better understanding of market‘s current conditions. It can be achieved by analyzing demand for a particular commodity. A company can manufacture the products according to demand of customers.

4. Development of New Product:

By analyzing the trends of customer’s requirements and level of their satisfaction, new product can be developed in the market.

5. Controlling Company’s Reputation:

Sentimental analysis can be done with the help of big data tools. Big data tools can help for monitoring the online presence of business

VI. CONCLUSION

In this paper, I have discussed the concepts of big data and various challenging issues of big data. This paper surveyed the techniques of big data i.e. HDFS and MapReduce that handles lager datasets from various sources and make significant improvement of systems. Support for fundamental research must be needed regarding technical challenges so that benefits of big data are achieved.

REFERENCES

[1] Rohit Pitre, Vijay Koolekar,” A Survey Paper on Data Mining With Big Data”, International Journal of Innoative Research in Advanced Engineering, Vol. 1, Issue 1, April 2014.

[2] Samiddha Mukherjee, Rani Shaw, “Big Data-Concepts, Applications, Challenges and Future Scope”, International Journal of Advanced Research in Computer nd Communication Engineering, Vol. 5, Issue 2, February 2016.

[3] Varsha B. Bobade, “Survey Paper on Big Data and Hadoop”, International Research Journal of Engineering and Technology, Vol 3, Issue 1, January 2016.

[4] C Lakshmi, V. V. Nagendra Kumar, “ Survey Paper on Big Data”, International Journalof Advanced Research in Computer Science and Software Engineering”, Vol. 6, Issue 8, August 2016.

[5] Ms. Vibhavari Chavan, Prof. Rajesh. N. Phursule, “Survey Paper On Big Data”, International Journal of Computer Science and Information Technologies, Vol. 5 (6), 2014.

[6]ApacheHadoop: http://developer.yahoo.com/hadoop/tutorial/module1.html.

[7] Kyuseok Shim, MapReduce,” Algorithms for Big Data Analysis”, DNIS 2013, LNCS 7813, pp. 44–48, 2013.

[8] Yuri Demchenko, “The Big Data Architecture Framework(BDAF)”,Outcomeofthe Brainstorming Session at the University of Amsterdam 17 July 2013.

[9] Elena Geanina ULARU, Florina Cameia PUICAN, Anca APOSTU, MAnole Velicanu,” Perspectives on Big Data and Big Data Analytics”, Database System Journal, Vol. III, no. 4/2012.

[10]https://www.google.co.in/search?q=hadoop&rlz=1C1CHBF_enIN768IN768&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjynJutt4raAhWHuI8KHYImBbwQ_AUICigB&biw=1366&bih=662#imgrc=8DHEIE6mv4rEPM:

[11]https://www.newgenapps.com/blog/importance-benefits-competitive-advantage-big-data