



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirce.com

Vol. 5, Issue 8, August 2017

A Survey on Big Data

Prachi Jain, Kanchan Jha

Assistant Professor, Computer Science Department, Sagar Institute of Research and Technology, Bhopal, India

Assistant Professor, Computer Science Department, Sagar Institute of Research and Technology, Bhopal, India

ABSTRACT: We live in on-demand world with vast majority of data. People and devices are constantly generating data, while streaming a video, active in social media, playing games, search any location using GPS. This data increase day by day from many resources, various types of techniques and technologies. The data is categories as “Big Data”. The term, ‘Big Data’ has been coined to refer to the gargantuan bulk of data that cannot be dealt with by traditional data-handling techniques. Big Data is still a novel concept, and in the following literature we intend to elaborate it in a palpable fashion. Big Data term appeared for First time in 1998 in Silicon Graphics (SGI) Slide Deck By john Massey with the title of Big Data. Big Data is very vast in majority and Complex data. Heterogeneity, scale, timeliness, complexity, and privacy problems with big data hamper the progress at all phases of the process that can create value from data.

KEYWORDS: Big Data, Networking, Structured Data, Unstructured Data.

I. INTRODUCTION

The term, ‘Big Data’ has been coined to refer to the gargantuan bulk of data that cannot be dealt with by traditional data-handling techniques. Big Data is still a novel concept, and in the following literature we intend to elaborate it in a palpable fashion. It commences with the concept of the subject in itself along with its properties and the two general approaches of dealing with it.

Big Data term appeared for First time in 1998 in Silicon Graphics (SGI) Slide Deck By john Massey with the title of Big Data. Big Data is very vast in majority and Complex data. Heterogeneity, scale, timeliness, complexity, and privacy problems with big data hamper the progress at all phases of the process that can create value from data. There are various resources of Big Data For Example: Audio, Videos, and Post in Social Media, Various Database Tables, and Email Attachment etc.

Big Data describes innovative techniques and technologies to capture, store, distribute, manage and analyze pet byte or larger sized datasets with high-velocity and different structures. Big data can be structured, unstructured or semi-structured, resulting in incapability of conventional data management methods. Data is generated from various different sources and can arrive in the system at various rates.

Firms like Google, eBay, LinkedIn, and Face book were built around big data from the beginning. It is a collection of massive and complex data sets that include the huge quantities of data, social media analytics, data management capabilities, real time data etc.

II. BIG DATA

Big data is the new term that contains large and complex datasets. It is difficult to manage these datasets without new technology. The McKinsey Global Institute (MGI) published a report on big data that describes the various business opportunities that big data opens [8]. Paulo Boldi, One of the authors says “Big Data does not need big machines, it needs big intelligence” [9]. There are two types of Big Data are as follows:

3.1 Structured Data These data can be easily analyzed. It is in numerical form, figures, and transaction data etc.

3.2 Unstructured Data These data contain complex information such as Email attachments, Images comments on social networking sites. These data cannot be easily analyzed. Doug Lancy was the first one talking about 3V's in big data management [3]: Volume - It describes the amount of data. It refers to mass quantities of data. Variety - It

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirce.com

Vol. 5, Issue 8, August 2017

describes different types of data and sources including structured, semi-structured and unstructured data. Velocity - It defines the motion of data. Data created rapidly, processed and analyzed.

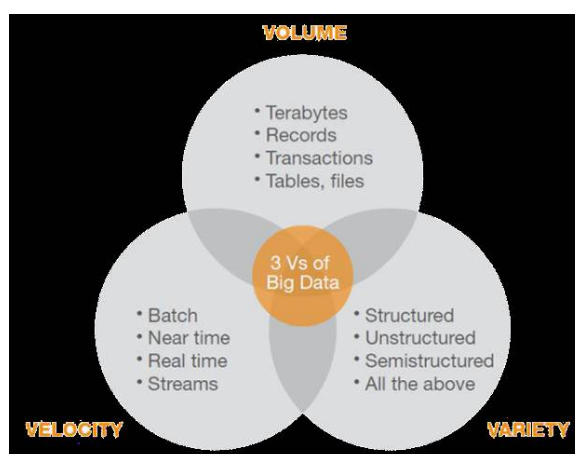


Fig 2.1 Three V's of Big Data

III. LITERATURE REVIEW

Over the last many years, there are many researchers has completed their work successfully on big data. Hundreds of articles have appeared in the general business press (For example Forbes, Fortune, Bloomberg, Business week, The Wall street journal, The Economist). National Institute of Standards and Technology [NIST] said that Big Data in which data volume, velocity and data representation ability to perform effective analysis using traditional relational approaches. In March 2012, The Obama Administration announced that the US would invest 200 Million Dollars to launch a big data research plan.

In this paper author explores the potential impact of big Data challenges, open research issues, and various tools associated with it. This article provides a platform to explore big data at numerous stages. Additionally, it opens a new horizon for researchers to develop the solution, based on the challenges and open research issues. Author survey the various research issues, challenges, and tools used to analyze these big data. From this survey, it is understood that every big data platform has its individual focus. Some of them are designed for batch processing whereas some are good at real-time analytic. Each big data platform also has specific functionality. Different techniques used for the analysis include statistical analysis, machine learning, data mining, intelligent analysis, cloud computing, quantum computing, and data stream processing. Author believes that in future researchers will pay more attention to these techniques to solve problems of big data effectively and efficiently. [1]

This survey paper author discusses Big Data from its infancy until its current state. It elaborates on the concepts of big data followed by the applications and the challenges faced by it. Finally we have discussed the future opportunities that could be harnessed in this field. Big Data is an evolving field, where much of the research is yet to be done. Big data at present is handled by the software named Hadoop. However, the proliferating amounts of data are making Hadoop insufficient. To harness the potential of Big Data completely in the future, extensive research needs to be carried out and revolutionary technologies need to be developed.[2]

In this paper, author presented the concept of big data. Big data is the large and complex datasets and it is generate from various sources like social media comments, playing a video game, email attachments etc. There is complexity in big data such as velocity, variety and volume. These three terms are more challenging for big data analytics. Author provided literature survey shows exponential growth of data in industries from 2005 year. There are variations possible while generating and storing data whether data is in audio, video, images and text. In big data analytics, Researchers divided generated data into various big data application such as structured data analytics, text analytics, web analytics,



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirce.com

Vol. 5, Issue 8, August 2017

multimedia analytics and mobile analytics. Many challenges in the big data system need further research attention. Research on typical big data application can generate profit for businesses, improve efficiency of government sectors.[3]

In this paper author describes about the Big Data with its 3Vs Volume, Velocity and Variety, evolution, problem definition and its techniques and technology with describes the Hadoop, HDFS and HPCC system in brief for efficient and fast processing. Big-data analysis fundamentally transforms operational, financial and commercial problems in aviation that were previously unsolvable within economic and human capital constraints using discrete data sets and on-premises hardware. Hadoop is open source software for data storage and fast processing. [4]

In this paper author talk about the problems of big data challenges, many researchers proposed a different system models, techniques for big data. The high performance computing paradigm is required for data mining to solve the problem of big data. There are still chances to improve the algorithms and techniques for data mining .In this paper, big data are facing lots of challenges, issues and provide a solutions to handle the big data. [5]

IV. CONCLUSION

Today, all the IT professionals, engineers and researchers are working on big data. Big data is term of concerning about large volume of complex data sets. In order to solve problems of big data challenges, many researchers proposed a different system models, techniques for big data. The high performance computing paradigm is required for data mining to solve the problem of big data. We conclude that there are still chances to improve the algorithms and techniques for data mining .In this paper, big data are facing lots of challenges, issues and provide a solutions to handle the big data.

REFERENCES

1. D. P. Acharjya Kauser Ahmed P A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools” (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 2, 2016.
2. Samiddha Mukherjee1, Ravi Shaw “Big Data – Concepts, Applications, Challenges and Future Scope” *International Journal of Advanced Research in Computer and Communication Engineering* Vol. 5, Issue 2, February 2016 Copyright to IJARCCCE DOI 10.17148/IJARCCCE.2016.5215 66
3. Tiwarkhedel, Prof. Vinit Kakde “A Review Paper on Big Data Analytics” *International Journal of Science and Research (IJSR)* ISSN (Online): 2319-7064.
4. Miss. Chetna Kashyap, Mrs. Ranjana Sharma” Review paper of Big Data and Methodology” 4th International Conference on System Modeling & Advancement in Research Trends (SMART) College of Computing Sciences and Information echnology (CCSIT) ,Teerthanker Mahaveer University , Moradabad.
5. Nirmal Kaur, Gurbinder Singh” A Review Paper On Data Mining And Big Data” Volume 8, No. 4, May 2017 (Special Issue) *International Journal of Advanced Research in Computer Science*.
6. Samiddha Mukherjee , Ravi Shaw “Big Data – Concepts, Applications, Challenges and Future Scope” *International Journal of Advanced Research in Computer and Communication Engineering* Vol. 5, Issue 2, February 2016 Copyright to IJARCCCE DOI 10.17148/IJARCCCE.2016.5215 66.
7. D. P. Acharjya, 632014 Kauser Ahmed” A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools” (IJACSA) *International Journal of Advanced Computer Science and Applications*, Vol. 7, No. 2, 2016.
8. James Manyika, Michael Chui, Brad Brown, Jacques Buhin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers, “Big Data: The next frontier for innovation, competition and productivity”, June 2011.
9. Albert Bifet, “Mining Big Data in Real Time”, *informatica*, 2013.