International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 7, 2017 ISSN: 2348-6600
Reference ID: IJCS-190 PAGE NO: 1158-1166
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

A REVIEW ON BIG DATA ANALYTICS

Dr.J.Suguna,
Associate professor,
Department of Computer Science,
Vellalar College for Womens,
Tamilnadu, India
Email:sugunajravi@yahoo.co.in

P.Kalaiyarasi,
PhD (Research scholar),
Department of Computer Science,
Vellalar College for Womens,
Tamilnadu, India.
Email:kalaipavi6@gmail.com

ABSTRACT- We live in on-demand world with vast majority of data. The name "big data" refer to as the massive amount of gigantic dataset in zeta byte or bigger sized dataset generated from various sources in daily life such as medical science or healthcare data, social media sites, satellites data, various sensors etc., with very high velocity. This vast collection of data, called Big Data, has caused the traditional tools incompetent for managing it from either of storage, computing or analytical perspective. There is an immense need of architectures, platforms, tools, techniques and algorithms to handle Big Data. The available technologies deal with two broad aspects related to Big Data that are Big Data Storage Management and Big Data Computing, focused toovercome various challenges such as scalability, faster processing speed, multiple format data processing, availability, faster response time and analytics etc.

Keywords: Big data Computing, BDSM, Big data.

1. INTRODUCTION

Big Data term appeared first time in 1998 in Silicon Graphics (SGI) Slide Deck by john Massey[3].The term big data is now widely used, particularly n the IT industry, where it has generated the various job opportunities. It consists of large datasets that cannot be managed efficiency by the common database management system. Data is everywhere in every industry in the form of numbers, images, videos, and text. As data continues to grow, so does the need to organize it. Collecting a huge amount of data would just be a waste of time, effort, and storage space if it cannot be put to any logical use. The needs to sort, organize, analyze, and offer this critical data in a systematic manner leads to the rise of the much discussed

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 7, 2017 ISSN: 2348-6600
Reference ID: IJCS-190 PAGE NO: 1158-1166
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

term, big data. The process of capturing or collecting big data is known as "datafication". Big data is „datafied" so that it can be used productively big data has many opportunities like financial services, Healthcare, Retail, Web/social, Manufacturing and Government [10]. Big data has now reached every sector in the global economy. It is estimated that by 2005, nearly all sectors in the US economy had an average of 200 terabytes of stored data per company with more than 1,000 employees [12]. Big data is continued to evolve rapidly, driven by innovation in underlying technologies. In August 2010, The White house, OMB, Proclaimed that big data is national challenge and priority along with healthcare and national security [14].The challenges include analysis, capture, duration, search, sharing, storage, transfer, visualization, and privacy violations [11].Scientists regularly encounter limitations due to large data sets in many areas, including meteorology, genomics, connectomics, complex physics simulations, and biological and environmental research. The limitations also affect Internet search, finance and business informatics. Data sets grow in size in part because they are increasingly being gathered by ubiquitous information-sensing mobile devices, aerial sensory technologies (remote sensing), software logs, cameras, microphones, radio- frequency identification (RFID) readers, and wireless sensor networks [1].

2. BIG DATA

Big data refers to a process that is used when traditional data mining and handling techniques cannot uncover the insights and meaning of the underlying data. There are three types of big data that is structure data, unstructured data and semi structure data [20].

Structured data: The term structured data generally refers to data that has a defined length and format for big data. These data can be easily analyzed and its numerical forms, figures and transaction data etc.

Unstructured data: These data are cannot be access easily. These data contain some complex information such as Email attachments, Images comments on social networking sites.

Semi structure data: These data it contain XML formatted data, CSV file and RDF.

3. CHARACTERISTICS OF BIG DATA

Big Data is vital because it enables organizations to gather, store, manage, and manipulate vast amount of data at the right speed, at the right time, to gain the right insights. As we all know the big data referred to as a large volume of huge dataset which is categorized into five main characteristics or 5V"s such as velocity, volume, variety, veracity and value. Each aspects put challenges in processing and handling massive volume of huge dataset to extract some meaningful information [19].

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 7, 2017 ISSN: 2348-6600
Reference ID: IJCS-190 PAGE NO: 1158-1166
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

Volume: This is amount of data generated by organizations or individuals. The volume of data in most organizations is approaching exabytes. Some experts predict the volume of data to reach zettabytes. **Velocity:** Enterprises can capitalize on data only if it is captured and shared in real time. Information processing systems such as CRM and ERP face problems associated with data, which keeps adding up but cannot be processed quickly.

Variety: Since, data is being generated at a very fast pace , and from different types of sources, such as internal, external, social, and behavioral, and comes in different formats, such as images, text, videos etc. It can be in varied formats, for example, GPS and social networking sites, such as Facebook, produce data of all types, including text, images, videos, etc.

Veracity: Its generally refers to the uncertainty of data, i.e., whether they obtained data is corrected or consistent. Out of the huge amount of data that is generated in almost every process, only the data that is correct and consistent can be used for further analysis..

Volume: It refers to the important feature of the data which is defined by the added-value that the collected data can bring to the intended process, activity or predictive analysis hypothesis [4].

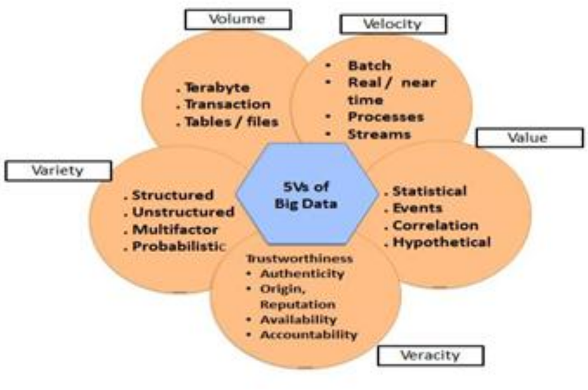

Fig 3 Characteristics of big data

4. PROCESS OF BIG DATA ANALYTICS

Data, which is available in abundance, can be streamlined for growth and expansion in technology as well as business. When data is analyzed successfully, it can become the answer to one of the most important question how can business acquire more customers and gain business insight? The key to this problem lies in being able to source, link, understand, and analyze data. It enables the organizations to analyze a mix of structured, semi structured and unstructured data in search of valuable business information.

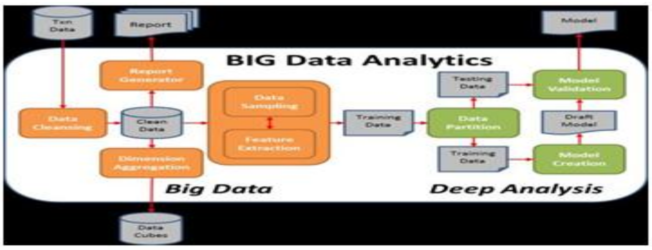

Fig: 4 Process of big data analytics

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
SKRGC Since 2012
http://www.ijcsjournal.com Volume 5, Issue 1, No 7, 2017 ISSN: 2348-6600
Reference ID: IJCS-190 PAGE NO: 1158-1166
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

Makinsey"s internal Think- Tank, the Mckinsey Global Institute, published a major study in June 2011 on Big Data [13]. Its overloading conclusion: Big Data is "a key basis of competition and growth". The term Analytics is often used broadly to cover any data-driven decision making [8]. The term analytics divided into two groups: Corporate analytics and Academic Research Analytics.

In Corporate Analytics, Team uses their expertise in statistics and Data mining. In Academic Analytics, Researchers analyze data to test Hypotheses and form theories [8].

Table: 4.1 Analytics approaches associated with bigdata.

Approach	Possible evaluations
Predictive analysis	This approach is an unstructured data. Real time analysis across its different domains that are sentiment data, social media, clickstream and multimedia.
Behavioral Analysis	It is a business leverage complex data in order to create new methods for driving business outcomes, decreasing business cost, strategy and improving overall customer satisfaction.
Data interpretation	It can be estimated from the available data and should be analyzed for new product innovation.

Prescriptive – This type of analysis discloses what actions should be taken. This is the most valuable kind of analysis and usually results in rules and blessings for next steps.

Predictive – An analysis of likely scenarios of what might happen. The deliverables are usually a predictive forecast.

Diagnostic – A look at past performance to determine what happened and why. The result of the analysis is often an analytic dashboard.

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 7, 2017 ISSN: 2348-6600
Reference ID: IJCS-190 PAGE NO: 1158-1166
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

Descriptive – What is happening now based on incoming data. To mine the analytics, use a real-time dashboard and/or email reports.

5. TECHNIQUES OF BIG DATA ANALYTICS

There are several techniques available in big data. This paper provides a list of some techniques applicable across a range of industries. This list is by no means exhaustive. Indeed, researchers continue to develop new techniques and improve on existing ones, particularly in response to the need to analyze new combinations of data. Some of them can be applied effectively to smaller datasets (e.g., A/B testing, regression analysis). However, all of the techniques listed here can be applied to big data and, in general, larger and more diverse datasets can be used to generate more numerous and insightful results than smaller, less diverse ones[16]. A/B testing: A technique in which a control group is compared with a variety of test groups in order to determine what treatments (i.e., changes) will improve a given objective variable, e.g., marketing response rate. This technique is also known as split testing or bucket testing. Association rule learning: A set of techniques for discovering interesting relationships, i.e., "association rules," among variables in large databases. This technique consists of a variety of algorithms to generate and test possible rules. One application is market basket analysis, in which a retailer can determine which products are frequently bought together and use this information for marketing used for data mining [16].

Classification tree analysis: Classification tree analysis is the best way in which different text data can be analyzed. Text analytics can also manifest itself in the form of classification tree analysis.

Genetic algorithms: These techniques that are used to identify the most possibly viewed videos, TV shows and other forms of media. There is an evolutionary pattern that can be identified by genetic algorithms. Video and media analytics can be done by the use of genetic algorithms.

Machine learning: Machine learning is another technique that can be used to categories and determine the probable outcome of a specific set of data. Machine learning defines software that can be able to determine the possible outcomes of a certain set of event. An example of predictive analytics is probability of winning legal cases or the success of certain productions. Generally, the field of machine learning is divided into three subdomains: supervised learning, unsupervised learning, and reinforcement learning [9].

Regression analysis: Regression analysis is used to estimate the strength and direction of the

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 7, 2017 ISSN: 2348-6600
Reference ID: IJCS-190 PAGE NO: 1158-1166
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

relationship between variables that are linearly related to each other. This is a technique that takes the use of independent variables and how they affect dependent variables.

Sentiment analysis: This is the ultimate technique that is used in text analytics. It looks at the actual sentiments of different people and then cross references them with the experience that is described in the text or audio response. Sentiment analysis is a categorization technique that is text based but can have applications in audio analytics [13].

Social network analysis: Social network data refers to the data generated from people socializing on social media. On a social networking site, different people will constantly adding and updating comments, statuses, preferences, etc. and all these activities generate large amounts of data [7].

6. TOOLS OF BIGDATA

Pentaho Business Analytics: Pentaho is a software platform that began as a report generating engine; it is, like JasperSoft, branching into big data by making it easier to absorb information from the new sources.. Karmasphere Studio and Analyst: Many of the big data tools did not begin life as reporting tools. Karmasphere Studio, for instance, is a set of plug-ins built on top of Eclipse. It's a specialized IDE that makes it easier to create and run Hadoop jobs. Talend Open Studio: Talend also offers an Eclipse- based IDE for stringing together data processing jobs with Hadoop. Its tools are designed to help with data integration, data quality, and data management, all with subroutines tuned to these jobs.

Hadoop: Hadoop is an Apache-managed software framework created using Map Reduce and Big Table. Hadoop allows applications based on Map Reduce to run on large clusters of commodity hardware. Hadoop is designed to parallelize data processing across computing nodes to speed computations and diminish latency [4].

Tableau: The Tableau is an American Software Company with its headquarters located in Seattle [5].Caching assists work with few latency of a Hadoop cluster. The ability to reslice data across various graphs gives an artistic effect to the software [6].

CHALLENGES IN BIG DATA ANALYTICS

Data integration

The ability to combine data that is not similar in structure or source and to do so quickly and at reasonable cost. With such variety, a related challenge is how to manage and control data quality so that you can meaningfully connect well understood data from your data warehouse with data that is less well understood.

Data volume

The ability to process the volume at an acceptable speed so that the information is available to decision makers when they need it. Skills availability

Big Data is being harnessed with new tools and is being looked at in different ways. There a shortage of people with the skills to bring together the data, analyze it and publish the results or conclusions. Solution cost Since Big Data has opened up a world of possible business improvements, there is a great deal of experimentation and discovery taking place to determine the patterns that matter and the insights that turn to value. To ensure a positive ROI on a Big Data project, therefore, it is crucial to reduce the cost of the solutions used to find that value [17].

Privacy

A lot of big data contains personal information about customers, clients, patients, and other types of users.People are concerned about how information relating to them is used, particularly how it is used to affect [2].

7. CONCLUSION

This study it aims to find out the big data analytics and its techniques are available and characteristics of big data. We have discussed about various tools used in big data analytics and what are the challenges in big data. And also this paper to review the fundamentals of big data analytics. Big Data is new and requires investigation and understanding of both technical and business Requirements. 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.

REFERENCES

[1] S. Vidhya, S. Sarumathi, N. Shanthi "Comparative Analysis of Diverse Collection of Big Data Analytics Tools " International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:8, No:9, 2014.

[2] Dylan Maltby "Big data analytics "New Orleans, LA, USA ASIST 2011, October 9-13, 2011.

[3] Bharti Thakur, Manish Mann," Data mining for big data: A Review", International journal of advanced Research in Computer Science and
Software Engineering, ISSN: 2277 128x, Volume 4, Issue 5, May 2014.

[4] Hiba Jasim Hadi, Ammar Hameed Shnain, "Big Data And Five V‟s Characteristics" Proceedings of IRF International Conference, 01st November 2014,
Tirupati, India, ISBN: 978-93-84209-61-2.

[5] http://casci.umd.edu/wp-content/uploads/2013/12/Tableau-Tutorial.pdf.

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 7, 2017 ISSN: 2348-6600
Reference ID: IJCS-190 PAGE NO: 1158-1166
Sri Vasavi College, Erode Self-Finance Wing *3rd February 2017*
National Conference on Computer and Communication *NCCC'17*
http://www.srivasavi.ac.in/ nccc2017@gmail.com

[6] http://www.infoworld.com/d/business-Big-Data-Analytics.pdf intelligence/7-top-tools-tamingbig- data-191131.

[7] Vivekananth.P, Leo John Baptist.A ," An Analysis of Big Data Analytics Techniques", nternational Journal of Engineering and Management Research Volume-5, Issue-5, pno: 17-19 October-2015.

[8] D.Fisher, R.Deline, M.Czerwinski and S. Drucker, "Interaction with big data analytics", Volume 19, No.3, May 2012.

[9] B Adam, IFC Smith, F Asce, Reinforcement learning for structural control. J compute Civil Eng 22(2), 133–139 (2008.

[10] Denis Guyadeen , Rob Peglar," Introduction to Analytics and Big data- Hadoop", SNIA Education Committee, 2012.

[11] Parmeshwari P. Sabnis, Chaitali A.Laulkar, "SURVEY OF MAPREDUCE OPTIMIZATION METHODS", ISSN (Print): 2319- 2526, Volume -3, Issue -1, 2014

[12] James Manyika, Michael Chui, Brad Brown, Jacques Bhuhin, Richard Dobbs, Charles Roxburgh, Angela Hungh Byers, "Big Data: The next frontier for innovation, competition and productivity", June 2011.

[13] Vivekananth.P, Leo John Baptist.A " An Analysis of Big Data Analytics Techniques " International Journal of Engineering and Management Research Volume-5, Issue-5, pp.no: 17-19 October-2015.

[14]American Institute Of Physics(AIP), 2010. College Park, MD(http:// www.aip.org /fyi/2010/)

[15]www.ingrammicroadvisor.com/data-center/four- types-of-big-data-analytics-and-examples-f-their-use.

[16]www.bigdata-madesimple.com/26-popular-techniques-for-analysing-big-data.

[17]www.gsma.com/membership/wp-content/uploads/2013/07/The-Top-Challenges-of-

[18] Bakshi Rohit Prasad and Sonali Agarwal" Comparative Study of Big Data Computing and Storage Tools: A Review " International Journal of Database Theory and Application Vol.9, No.1 (2016), pp.45-66

[19] Ms. Vibhavari Chavan, Prof. Rajesh. N. Phursule "Survey Paper On Big Data"International Journal of Computer

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 7, 2017 ISSN: 2348-6600
Reference ID: IJCS-190 PAGE NO: 1158-1166
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

Science and Information Technologies, Vol. 5 (6) , 2014, pp.7932-7939.

[20] Ankita S. Tiwarkhede1, Prof. Vinit Kakde" A Review Paper on Big Data Analytics International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2013):4.438.