A Review paper on Big Data

Shalini Kumari[#1], Dr.Neeru Bhardwaj[*2]
[#]*Research Scholar & Computer Science and Engineering & Bahra University*
Waknaghat, Shimla Hills (H.P), 173234, India
[*]*Head of Department & Computer Science and Engineering & Bahra University*
Waknaghat, Shimla Hills (H.P), 173234, India

Abstract

Big data is a very large terms for data set or complex those widely used in data deal with the utilization deficient. The value from the term big data refers simply to use the predictive analysis or also uses the progressive approach to excerpt the value from the data, and we can pick up more confident for decision making. And with the good decisions can means larger practical capability, charge discount and diminished hazards. Analysis of data sets can find the new correlation. The V model for the big data has been defined and categorized in to 3V, 4V or 5V dependent on the organization which uses it and under which business scenario. Catering to the aforementioned models, we have classified data in to various forms and explanations have been provided on the same to gain a better insight and understanding on the same. We need a different platform named Hadoop as the core part of the structuring the big data, and also solves the problem of making it useful for the analytics process.

Keywords: Big data, Characteristics, applications.

I. INTRODUCTION

Big Data

Big data broadly implicate gather knowledge from particular sources. The term big data accommodate the four distinct items. We can say it power of 5 V's of big data. Big data amount regularly development, modern pasturing from a minor dozen terabytes to petabytes. Along the expansion of automation and the heightened large group of data streaming in and out grouping day to day, ready turn in to a demand because rapid and higher adequate approach about investigate specific a data [15]. Hadoop is particular about the appliances and is a deeply extensible cache multiform, as a result of it can stock and assign correct extensive inputs arranged crosswise century as concern inclusive assistant a well known achieve in coordinate [2, 3, 8, and 18].

Figure 1: Big Data

II. LITERATURE SURVEY

Past the terminals countless lifetimes, attendant do frequent researchers has entire their effort strongly on big data. Century of piece accept came in the commonplace field press [1]. National Institute of Standards and Technology [NIST] said that Big Data in which data volume, velocity and data image capacity to achieve powerful investigate applying classic comparative approve [15].

In March 2012, The Obama Administration announced that the US would invest 200 Million Dollars to launch a big data research plan [2]. An IDC Reports predicts that from 2005 to 2020 the global data volume will grow by a factor of 300, from 130 Exabyte's to 40,000 Exabyte's, representing a double growth every two years[9]. IBM assessment that frequently 2.5 quintillion bytes of data are created out of which 90% of the data in the world today has created in the last two years. It is observed that social networking sites like Facebook have 750 Million users, LinkedIn has 110 million users and Twitter has 250 million users [17].

From industry, government and research community, Big Data has led to an emerging research field that has attracted tremendous interest. The broad interest is first exampled by coverage on both industrial reports and public media for example: The economist, New York Times [12]. Cell phones enhancing finest approach to get data on people from distinct condition, the great extent of inputs that mobile transporter can produce to enhance our daily life [13].

III. CHARACTERISTICS

Big data can be described by the following characteristics.

Volume – Big data involve excessive quantity of data. It is the amount of the data which resolve the value and probable of the data under consideration. Data is develop by machines, grid and human communication on systems like social media the volume of data to be analyzed is massive.

Variety – The next aspect of big data is its variety. Variety refers to the many sources and types of data structured and unstructured. We used to store data from sources like spreadsheets and datasets. This benefits the community, who is firmly analyzing the data and is associated with it, to effectively use the data to their advantage and thus upholding the importance of the big data.

Velocity – The term velocity in the context refers to the speed of generation of data or how fast the data is generated and processed to meet the demands and the challenges which lie ahead in the path of growth and development.

Variability – This is a factor which can be a problem for those who analysis the data. This refers to the inconsistency which can be shown by the data at times, thus hampering the process of being able to handle and manage the data effectively.

Veracity – The quality of the data being captured can vary greatly. Accuracy of analysis depends on the veracity of the source data.

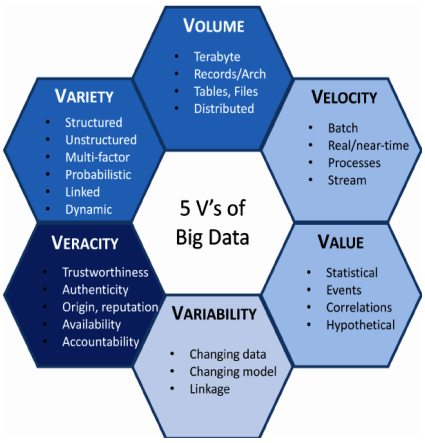

Figure 2: 5 V's of Big Data

IV. APPLICATIONS

Healthcare is one of the best developments in the world. Big data in healthcare entail to electronic health data sets that are describe with patient healthcare and prosperity. Data in the healthcare area is developing past managing limit of the healthcare associations and is relied upon to increment fundamentally in the coming years.

Big data has increased the demand of information management specialists in that Software AG, Oracle Corporation, IBM, Microsoft, SAP, EMC, HP and Dell have spent more than $ 15 billion on software firms specializing in data management and analytics.

Developed economics make increasing use of data-intensive technologies.

While many vendors offer off-the-shelf solutions for big data, experts recommend the development of in-house solutions custom-tailored to solve the company's problem at hand if the company has sufficient technical capabilities.

CONCLUSIONS

In this paper, we have conferred the conception of big data. Big data is the huge and complicated datasets and it is achieve from different origins like social media comments, playing a video game, email attachments etc. The opportunity of big data, low cost product hardware, and new intelligence management and analytic software has composed a particular moment in the history of data analysis. The convergence of these trends means that we have the capabilities required to analyze astonishing data sets quickly and cost-effectively for the first time in history. The age of big data is here, and these truly revolutionary times if both business and technology professionals continue to work together and deliver on the promise.

ACKNOWLEDGMENT

I would like to express my gratitude to Dr. Neeru Bhardwaj for being a constant source of encouragement and inspiring me in all means to bring out the best in my work.

REFERENCES

[1] Sameera Siddiqui, Deepa Gupta," Big Data Process and Analytics: A Survey", International Journal Of Emerging Research in Management & Technology, ISSN: 2278-9359, Volume 3, Issue 7, July 2014.

[2] Han Hu, Yongyang Nen, Tat Seng Chua, Xuelong Li," Towards Scalable System for Big Data Analytics: A Technology Tutorial", IEEE Access, Volume 2, Page No 653, June 2014.

[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] Anand V. Saurkar, Vaibhav Bhujade, Priti Bhagat and Amit Khaparde," A Review Paper on Various Data Mining Techniques", International Journal of Advanced Research in Computer Science and software Engineering, ISSN:2277 128X,Volume 4, Issue 4, April 2014.

[5] Puneet Singh Duggal, Sanchita Paul,"Big Data Analysis: Challenges and Solutions", International Conference on Cloud, Big Data and Trust 2013, Nov 2013.

[6] Albert Bifet, "Mining Big Data in Real Time", informatica, 2013.

[7] Stephen Kaisler, Frank Armour, J. Alberto Espinosa and William Money," Big Data: Issues and Challenges Moving Forward",Hawaii International Conference on System Science, IEEE Computer Society, Page No. 995, 2013.
[8] D.Fisher, R.Deline, M.Czerwinski and S. Drucker,"Interaction with big data analytics", Volume 19, No.3, May 2012.
[9] J.Gantz, D. Reinset," The Digital Universe in 2020: Big Data, Bigger digital shadow, and biggest growth in the Far East", in Proc: IDC in view, IDC Anal, Future, 2012.
[10] Denis Guyadeen, Rob Peglar," Introduction to Analytics and Big data- Hadoop", SNIA Education Committee, 2012.
[11] Neil Raden,"Big Data Analytics Architecture", Hired Brains Inc, 2012
[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] Wei Fan, Albert Bifet, "Mining Big Data: Current Status and Forecast to the Future", SIGKDD Explorations, Volume 14, Issue 2.
[14] American Institute of Physics (AIP), 2010.College Park,MD
(http:// www.aip.org /FYI/2010/) [15] M.Cooper, P.Mell (2012). Tackling big Data (Online). Http://csrc.nist.gov/groups/SMA/Forum/document/June2012Presentation/f%CSM_june2012_cooper_Neul.pdf.
[15] www.google.com
[16] www.wikipedia.com
[17] www.seminarsonly.com