STATE-OF-THE-ART RECENT TECHNOLOGIES, BACKGROUND, TERMINOLOGIES, CHALLENGES: A PRELIMINARY SURVEY ON BIG DATA

C. Esther[1*], S.Keerthana[2], S.Pradhipa[3], C.A.Daphine Desona Clemency[4,] M.Joy Priyanka[5]

[1]Assistant Professor, Sri Sairam Engineering College, Chennai, Tamilnadu, India
[2]Assistant Professor, Sri Sairam Institute of Technology, Chennai, Tamilnadu, India.
[3]Assistant Professor, Sri Sairam Engineering College, Chennai, Tamilnadu, India
[4,5]Assistant Professor, Dhanalakshmi Srinivasan College of Engineering and Technology, Chennai, Tamilnadu, India

Abstract
Big data is a future research area acquiring substantial attention from academia and research communities. In advanced technologies, the extents of data generated and stored have dispersed within a short period of time. In context of big data, traditional data techniques and programs are less coherent. They show a slow responsiveness and lack of scalability, performance and accuracy. Subsequently, fast growing rate of data has created many demanding factors like speedy growth of data, transmission speed, data diverse, security and so on. At this point, distinct types of distributions, technologies, applications and latest trends have been developed. This paper explicitly presents the overall analysis on state-of-the-art big data technologies, background, terminologies, technical challenges and review of current approaches. In addition, future research suggestions in the field of big data are determined based on progress and definite open issues in big data analytics. Therefore research directions facilitate the study of domain with improvement of optimal techniques to expose big data.
Keywords: Big data, Data storage, Analytics, Security, Data processing

Introduction

Data Analytics is an experimental and analytical tool for discriminating raw data to retrieve information for obtaining knowledge. Further, data analytics conspire with data to organize multiple decisions from distinct perspectives for covering real world demanding [16]. The key part of analytics is to collect, store, process and analyze data to assign empirical methods in real world for decision making. Evidently, it is largely classified into descriptive, predictive and prescriptive analytics [3].

The study of data analytics extends as a process of identifying an enormous real time operating data thereby it varies in structure of data are termed as Big Data Analytics. Big data reveal a different range of applications such as confines for innovation, competition, productivity, business and trade forecasting accordingly the data is exponentially developed [2]. In contrast, the analytics on large data acknowledges hidden figure, unfound correlations, market tendency, customer requirements and future suggestions which in turn assist in captious decision-making action [1].

In general, data is stored in an exceedingly structured format to increase its informative details. However, actual data volumes are driven through both unstructured and semi-structured data [9]. Therefore, end to end processing can be blocked by the transcription between structured data in relational database management systems and unstructured data for analytical purpose [4]. On the other side, advancement rate of collected data expand diverse critical issues and challenges including expeditious data growth, transmission speed, differing data and some security issues.

The salient features of big data have been proposed by Doug Laney in the year of 2001. Big data defined various challenges and opportunities which can be increased overall data with 3Vs such as increase of Volume, Velocity and Variety [5]. In addition 3Vsexposes with the formation and selection of large data therefore scaling of data becomes increasingly big in volume. In relevance to velocity, the timeliness of big data especially collection data and analysis must be assisted promptly and well-timed which enhance maximum resources in the economic

value of big data. Variety implies the different categories of data like traditional structured data, semi-structured and unstructured data as audio, video, image, text, webpage and so on.

Further, the paper provides a survey of literature for various big data technologies investigated in recent times. The rest of the paper is organized as follows. Section 2 provides a general characterization about Vs model. Section 3 provides an analysis of diverse technologies in big data followed by the comparison table. The conclusion of the paper is inclined in section 4.

Characterization of Big Data

The investigative tools suitable on big data promote both technical and academic industries for processing conventional time span and for inheriting knowledge from the data [6].In recent, the standard features of big data are depicted as ten V's as demonstrated in figure 1 which represents the properties that are entitled below:

Volume: An enormous size of real time functioning data.

Variety: Data in disparate forms like text, images, music, videos, graphs, plots etc. · Velocity: Speed of the data outflow derivation from various origin.

Veracity: Reliable data appearance from large sites.

Validity: Data storage in terms of permissibility and certain period of time.

Value: Implication of data stored and computing the quality.

Variability: Alternating the data along with the data flow.

Venue: Repository and retrieval of data from a definite location in data entries. Vocabulary: Language expression, readability and understanding of grammatical characters.

Vague: Data is extensive and does not compose discern except it is determined.

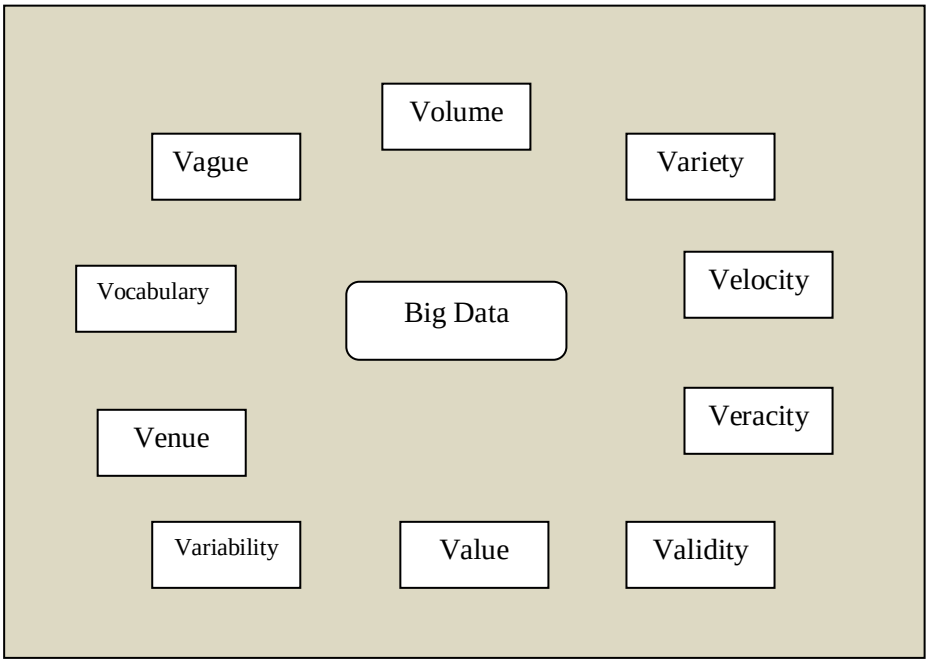

Figure 1. Characterization of Big Data

Survey on Big Data Technologies

With relevance to the big data analytics based driven agriculture has been analyzed [7] to examine by means of an immense and comprehensive datasets in plant breeding, bionomics, remote sensing and analytics has strongly attempted to maximize the overall benefit in agriculture. Consequently, to confer mechanisms for developing an acknowledged field phenotyping operation in addition to regulate protocols for assortment and evaluation of agricultural big data. In this phase, to examine the best employ the generous association of public and private plant cultivator and agronomists to resolve further challenges to form further use of data which provide application of coordinated technique to other datasets. Finally, to generate a report based on cross pointed, short and long term capitalization requirement for regular accomplishment in agriculture.

Hierarchical framework for the advancement of a structure business model established on big data has been introduced [8]. Here fuzzy set theory is used for covering out the unnecessary characteristic or aspect. In general, decision making and trial evaluation laboratory (DEMATEL) is highly recommended to maintain the complicated interrelationships among the respective attributes. Moreover, interpretive structural modeling (ISM) is used to segregate the position hierarchy which in turn finally compose a hierarchical framework. The final outcome reveal that (i)The fundamental links lies between the value proposition and public building in measurable management (ii) This technique ensures technical support in information technology as well as information management in value construction (iii) Additionally, increasing in customer with marketing value attain the expert power (iv) Last link is the value acquisition and that it is well-established by the support of determined value marketing and value organization.

The approval of Electronic Medical Records (EMRs) has been reviewed. In a contrast, emerging information technologies were proposed which capitalizes on large impact on the healthcare emergency. Primarily, EMR comprise health sensing for medical data collection, data investigation and implementation for definite detection and prediction [10]. On the other hand, cloud computing ensure scalable and cost-effective delivery of healthcare services. Ultimately, the actual case of scholastic exploration is dependable on convergent information technologies for new standard of healthcare maintenance.

Generation of Big Data

In this section, effective formations of big data need to be exploring which is the key role to generate the streaming of data.

Internet of Things (IoT)

The Internet of Things (IoT) point out the million of substantial devices around the world which are connected through the internet for collecting different surrounding environments and shared those information to the destination [11]. The concept of expanding sensor nodes and intelligence to fundamental things were determined throughout the 1980s and 1990sapart from various initial strategies including an internet connected vending machine evolution was gradually simple because the technology is not accessible. In a consequence, chips were too colossal and heavy and hence there is no way for things to communicate efficiently. Further, the IoT was originally highest interesting in business and manufacturing sectors where many applications are sometimes termed as machine-to-machine (M2M) but the significance is instantly filling distinct environment like homes and offices with smart devices which can transforming it into original information [12].

Biomedical and Medical Data

Bio and Medical relevant area such as bio informatics, clinical informatics, health related area fermenting the multidimensional dataset which are alternate source of big data generation. To ensure this clinical complementary activity, genome related program and real time health monitoring system are also computing huge data [13]. In addition to this medical imaging software including PET, CT scan, MRI, X-Ray and so on was generating the large amount of data even with more complex element. Moreover, the project called ProteomicsDB was started by the Swiss government authority to maintain the chromosome which is size of 5.17 TB.

Social Media

In recent time's social media like facebook, YouTube, twitter, instagram etc are generating huge amount of data generation. In instagram they are indulgence 7.1 million of active user in which 34.7 billion of user's share the photo with an average of 1650 million like per day [14]. Subsequently, in YouTube for every 300 hours of video subsistence upload per second further more than 3.25 billion hours video has been watched by users within a month. Likewise, the twitter holds 115 million active users every month in turn average of 58 million were tweet various information per day. In addition, more than 5 million people from worldwide can make call, text, tweet, and browse on mobile devices frequently. Various organisations could also communicate with their employees and alternate stakeholders such as students, customers and external advisors using diverse tools such as Yammer which is a private social network that assist collaboration across administration, region and other business standard applications [15].

Conclusion

In this paper, conferring of a review over recent big data research has been analyzed. Further research has extended into three main subdivisions as categorization of big data, latest technologies in big data and generation of big data. The primary objective of this paper is to provide a various survey of big data models

thereby handling techniques can maintain an enormous amount of data from different sources and enhance overall performance of systems and its standard resources. Therefore, the impact of big data has led to prime modifications within the business world. The advantages of big data are huge and can ready to employ individual like industries, medical, business, logistics and so on. For industrial sector holds trillions of dollars for business in every year hence big data is not depicted as a comfort but as an essential one. Researchers are continuously running with distinct algorithms to mine big data effectively and randomly. Furthermore, comparative analysis of selected big data technologies seems to be highly influential and also used for future research direction in real time big data applications.

References

1. Tripti M and Neha Mangle 2016, "A Survey Paper on Big Data Analytics using Map Reduce and Hive on Hadoop Framework" International Journal of Recent Advances in Engineering & Technology, Vol.4, No. 2, pp. 112- 118.
2. Mohammad Sultan Mahmud, Joshua Zhexue Huang, Salman Salloum, Tamer Z. Emara and Kuanishbay Sadatdiynov 2020, "A survey of data partitioning and sampling methods to support big data analysis"Big Data Mining and Analytics, Vol. 3, No. 2, pp. 1-10.
3. Burhan U.I.Khan, Rashidah F and Hunain A 2014, "Critical Insight for MapReduce Optimization in Hadoop" International Journal of Computer Science and Control Engineering, Vol.2, No.1, pp. 1-7.
4. Min Chen, Shiwen Mao and Yunhao Liu 2016, "Big Data: A Survey", Mobile Network Applications, Vol. 19, pp. 171-209.
5. Rabi Prasad P 2013, "Big Data Processing with Hadoop-MapReduce in Cloud Systems", International Journal of Cloud Computing and Services Science, Vol.2, No.1, pp. 16-27.
6. Surabhi Verma 2017, "Big Data and Advanced Analytics: Architectures, Techniques, Applications and Challenges", International Journal of Business Analytics, Vol.4, No.4, pp. 23-45.
7. Nadia Shakoor, Daniel Northrup, Seth Murray and Todd C. Mockler 2019, "Big Data Driven Agriculture: Big Data Analytics in Plant Breeding, Genomics, and the Use of Remote Sensing Technologies to Advance Crop Productivity", The Plant Phenome Journal, Vol.2, No. 1, pp. 1-8.
8. Xiaomin Du, Yang Gao, Linlin Chang, Xiangxiang Lang, Xingqun Xue, and Datian Bi 2020, "Assessing the application of big data technology in platform business model: A hierarchical framework", Plos One, Vol. 15, pp. 1-21.
9. Bernice Purcell 2019, "The emergence of "big data" technology and analytics", Journal of Technology Research, Vol. 3, pp. 1-6.

10. Ji-JiangYang, JianqiangLi, JacobMulder, YongcaiWang, ShiChen, HongWu, QingWang and HuiPan 2015, "Emerging information technologies for enhanced healthcare", Computers in industry, Vol. 69, pp. 3-11.

11. Somayya Madakam, R. Ramaswamy, Siddharth Tripathi 2015, "Internet of Things (IoT): A Literature Review",Journal of Computer and Communications, Vol. 3, pp. 164-173.

12. Chun Wang, Ming-Hui Chen, Elizabeth Schifano, Jing Wu and Jun Yan 2018, "Statistical Methods and Computing for Big Data", International Journal of Information Management, Vol.2, No.24, pp. 212-219.

13. Jake Luo, Min Wu, Deepika Gopukumar, and Yiqing Zhao 2016, "Big Data Application in Biomedical Research and Health Care: A Literature Review", Biomed Inform Insights pp. 13-24.

14.Vikas Dhawan and Nadir Zanini 2014, "Big data and social media analytics", RESEARCH matters, No.18, pp. 36-41.
15. P. Russom 2015, "Big data analytics," TDWI Best Practices Report, Fourth Quarter, Vol.2, pp. 32-45.
16. Demirkan H, C. Bess, J. Spohrer, A. Rayes, D. Allen and Y. Moghaddam 2015 "Innovations with Smart Service Systems: Analytics, Big Data, Cognitive Assistance, and the Internet of Everything." Communications of the AIS, pp.37- 35.