Big Data Analytics And Its Challenges

Jyotsna Tiwari[1]*, Dr. Monika Tripathi[2]

[1] Research Scholar, Shri Krishna University, Chhatarpur, M.P.

[2] Professor, Shri Krishna University, Chhatarpur, M.P.

Abstract - In order to make informed decisions and advance their strategies, companies are realizing they need access to massive amounts of data. The proliferation of the Internet and social media has led to an explosion in the volume of digital data being produced. "Big Data" refers to the massive amount of disparate digital information produced by businesses and individuals, and its characteristics necessitate the use of cutting-edge computer storage and processing methods. Examine the difficulties of big data analytics here.

Keywords - Big Data; Big Data Analytics; Hadoop; Internet; Security

- *X* -

INTRODUCTION

Big data often refers to datasets that have become too big and complicated for standard tools or database management systems to handle. It also indicates datasets with a lot of variation and velocity, necessitating the development of potential remedies to create profit and information from broad, dynamic datasets.[1]

Big data is described as "very massive data sets that may be computationally analysed to uncover patterns, trends, and relationships, particularly pertaining to human behaviour and relationships" in the Oxford English Dictionary. It has been suggested that this description does not adequately describe large data since big data must be distinguished from data that is challenging to manage using conventional data analysis. Big data's rapidly rising complexity necessitates more advanced approaches for processing it.[2]

Although the phrase "big data" has gained some popularity by 2011, the ProQuest Academic Library still provides a clearer picture of its frequency distribution.[3]
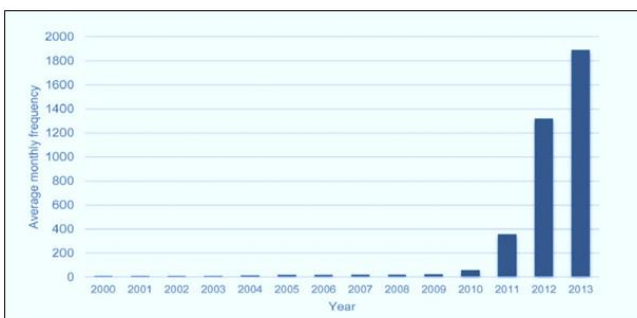

Figure 1: "Big data" frequency distribution in the ProQuest Research Library.

Various definitions of big data are employed in research and industry, according to a study. Big data definitions differ based on the user's knowledge, with some focusing on the properties of big data (volume, variety and velocity), others defining it based on their business's needs, and yet others focusing on what big data does. SAP commissioned Harris Interactive to conduct an online poll of 154 C-suite worldwide executives in April 2012, [4]

Early research focused on defining big data based on the 3Vs: volume, velocity, and variety. A overview of big data research was subsequently provided, which included an examination of security challenges and a definition of big data using the 5Vs, building on the work of Laney (2001), which had previously used the 3Vs to describe data's value and veracity.[5]

as a result, a set of current big data standards, as indicated in Table, was recently constructed.

Data volume forecasts from industry analyst firm IDC are seen in Figure 8. Managing and safeguarding the enormous amounts of big data has become more challenging due to the data's complicated structure and the sheer volume of it. Thus, big data has been a major emphasis in both technological and engineering fields since its conception.[6]

New ways of gathering data are needed in order for big data's full potential to be realised; this is evident in today's numerous connected gadgets, and the vast amounts of data that even people may access.

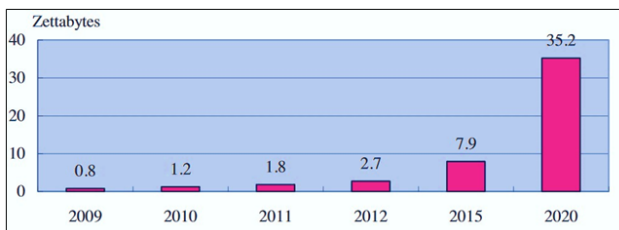The information in the database is expected to increase every 24 months by 2020, according to the forecast.[7]

Figure 2: IDC predicts a massive amount of data will be generated throughout the world.

At times, the definitions above are complementary to one other, such as the 5Vs definition of big data. Table 1 shows six sample definitions, however some of them are in conflict with those definitions. According to their definition of big data, they concentrated on the magnitude of the data, neglecting other aspects.[8]

These definitions, when seen through the eyes of the end user, reveal the many facets of how big data is being utilised in academia and industry today. Other aspects, like how data is kept or how it is used in a corporation, are as important, although they're not as prominent.[9]

However, the definition used in this study is one that includes all dimensions. This is because to its great density, timeliness, and variety of formats, structures, and sources, all of which need high-end processing..

CHARACTERISTICS OF THE BIG DATA.

Although other criteria matter, different definitions of big data show that size is the dominant factor. data management, as well as the three V's provide a shared paradigm for addressing such challenges. If one of these three dimensions changes, the likelihood of a change in the other two rises as well.[10]

The truthfulness and variability dimensions, , are often added to the list of big data characteristics. Big data is becoming more popular, and the five V's are a reflection of this. The quantity of data created is directly proportional to the first V, which is always volume. Because all data collecting and processing must be completed as quickly as possible, the second V stands for "velocity." As big data comes in a wide range of forms and structures, the third V alludes to variety, which is important because of the wide range of formats and structures available. Big data's "high value but very low density" (the "fourth V") causes significant issues when trying to extract value from large datasets. The authenticity of big data is called into doubt when information comes from external sources, which is the situation in the majority of instances. Veracity is connected to the data source's reputation, the correctness of the data, and how fit the data is for the planned purpose.[11]

With big data, the right technological architecture, analytics, or tools must be in place so that insights may be gleaned from previously undiscovered information, and they rely on the data's size, dispersion, variety, and velocity. However, the three most important characteristics of big data are its volume, velocity, and variety.[12]

In the early stages of big data, streaming data may be gathered in real time from a variety of sources, such as social media and blogs. Veracity has also been debated by scholars and organisations in this setting.. True veracity is based on the quality of a data, which might be excellent or poor owing to inconsistent data, incompleteness, ambiguity, delay, deceit, or approximations. The lack of control and uniformity that comes with massive data sources that are mostly external is a problem.[13]

Data management and value extraction are critical for today's businesses looking to gain an edge in the marketplace. When it comes to deriving commercial value from big data, there are both technological and business hurdles at play. As a result, demonstrating how big data contributes to organisational goals while also taking a technical viewpoint on the topic is now an essential part of study in this area. Organizations may get benefit from big data, according to Manyika et al. (2011).making knowledge more understandable and relevant;Incorporating advanced big data analytics to enhance the quality of decision-making; utilising big data to help design the future generation of goods and services by giving organisations the creation and storage of transactional digitally stored data.[14]

STORAGE AND ADMINISTRATION OF LARGE AMOUNTS OF DATA

In order to properly manage enormous amounts of data, the most challenging difficulty is storage; dealing with vast numbers and kinds of data is not always straightforward.

It is possible to store and analyse large amounts of data in a variety of ways. A data centre may be required to store and handle massive amounts of data due to the sheer number of users and devices. In order to collect this quickly created data, a network architecture must be established and transferred to the data centre so consumers can access it.

An initial data network, bridges used to link and send data to data centres, as well as at least one centre are identified by Yi et al. (2014) in their research.[15]

According to another research, users are unable to pick and choose data from the network while accessing huge data from particular areas. Ultra-scalable solutions may cause inefficiency by

blocking the processing of particular data sources, which is a major concern for storage models. Big data technology is difficult to build, and each new system must be able to collect and distribute information across nodes located all over the globe.

"Relational databases, data marts, or data warehouses" are all examples of structured data storage and retrieval. The database is loaded with data that has been retrieved from external sources, altered to meet operational requirements, and then reloaded. There are tools for extracting, transforming, and loading the data from the enterprise data warehouse into a long-term storage location. There are many steps involved in preparing the data for usage.[16]

In contrast to the typical EDW context, a big data context need analytical abilities. The big data ecosystem welcomes and expects any and all data sources conceivable. There are certain caveats to the EDW approach to data sources since it is more focused on structured data. Big data storage necessitates agile databases because of the ever-increasing number of sources of information and analysis that may be performed. Large datasets can only be fully analysed by using sophisticated statistical approaches.[17]

PROCESSING OF BIG DATA ANALYTICS

a. Analytic processing, the next step following massive data storage, involves four important criteria.In order to accelerate query execution, disc and network interference is minimised.It is essential that real-time requests be handled as rapidly as possible to meet the needs of the end user. There should be a way to handle numerous inquiries as the amount of requests grows.

b. In order to keep up with the increasing growth of user activity, they need scalable storage and computational power. It is essential to control data storage throughout processing to deal with space concerns adaptively since disc space is limited.

c. Because data processing has variable workload patterns and huge dataset analysis includes many distinct applications and users with various intents and techniques, the underlying system must be very adaptable.[18]

Big data frameworks have transformed the environment of knowledge discovery from data, particularly in terms of data mining processes or pre-processing, with a particular emphasis on the increase of data which was before in cloud computing, according to the study given here. A wide range of data preprocessing techniques, including big data and data preprocessing, were addressed in the solution. Factors including such maximum supported data size and preprocessing of data were also evaluated. There was also discussion of other big data frameworks, including Hadoop, Spark, and Flink.[19-21]

CHALLENGES IN THE USE OF BIG DATA ANALYTICS

Data mining, visualisation, statistical analysis, or machine learning have been the subject of many research; nevertheless, new analytic methodologies must be developed to deal with the problems of big data, such as the time necessary for processing when the amount of data is very vast. Deep learning, deep learning, gradual techniques, and granular computing have all been shown to be challenging to implement in practise.

Similarly, we looked at several methods for dealing with big data concerns, including such cloud computing or quantum computing, in order to determine their usefulness. Concepts, data features, and processing paradigms; state-of-the-art methodologies for making decision in big data; applications in social science; and big data's current issues and future prospects were all covered comprehensive review of big data.[22]

- **Issues Related To Data Security**

Managers and officials that work in public affairs should focus on resolving privacy and internet access inequities, as well as legal and security problems. The restricted budgets, diverse stakeholders, and short time limits for information extraction constrain public managers and policymakers. the colossal amount of data

- **The Problem Of Data Privacy**

Collection data from users may lead to privacy issues since the gathering process may alter the data context or semantics, resulting in inaccurate and inefficient rules.

An issue with big data is that it frequently contains sensitive information like medical records and financial transactions, which isn't suited for standard data transfer methods. Considering safety and privacy issues before establishing a mechanism for exchanging data is thus essential. Secure certification procedures remain difficult to deploy, and anonymization schemes reduce data trust, despite the fact that these issues are widely

Jyotsna Tiwari[1]*, Dr. Monika Tripathi[2]

understood.[23]

• **Quality Of Data In The Form Of Storage, Capture, And Storage**

Data collection and storage is a difficult task, made more difficult by the rising quantity and complexity of data collections. Many industries, such as finance and medicine, are obliged to destroy data since there isn't enough storage capacity. A hefty price is attached to acquiring and generating important data.

The properties of big data, include the fact that it is processed by a variety of analytical tools and visualisations. The components and technologies of the big data platform layer were discussed. Technology was compared and big data systems were categorised as according their characteristics and the services they provide to consumers.. It was shown that there are still a slew of technical challenges with big data utilisation that need to be worked out. They also addressed large data computer systems' obstacles, analysing difficulties on a variety of levels "including data acquisition, storage, finding, sharing, analysis, management, and visualisation," among others. Concerns about safety and privacy were part of the investigation. Big data is becoming more and bigger all the time, and present technology can't keep up.[24]

CONCLUSION

Big data refers to the massive amounts of numerical information generated by the widespread use of new technology, whether for private or commercial purposes. The process of analyzing large amounts of data for useful information, such as patterns, market trends, and customer preferences, is known as Big Data analytics. Technology-wise, Big Data analytics is on the rise. It has been adopted by unexpected areas and grown into its own market. The difficulties of analyzing this data in the context of Big Data, however, might seem intrusive at times. Analysis is what data scientists do. Big data is the means by which questions are posed, whereas business intelligence is concerned with answering those questions. Analytics tools are used when a company has to know or wants to know what will happen in the near future, and BI tools help translate such forecasts into intelligible language. Many people consider Big Data to be the next step in the development of business intelligence.

REFERENCES

1. Anastasia, February 2015. Big data and new product development. Entrepreneurial Insights http://www.entrepreneurial-insights.com/big-data-new-product-development/, Accessed on June 15, 2015.
2. Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M.A.S. ,Buyya, R., 2015. Big Data computing and clouds: Trends and future directions, Journal of Parallel and Distributed Computing, Volumes 79–80, May, Pages 3-15.
3. Bakshi, K., 2012. Considerations for big data: Architecture and approach. In: 2012 IEEE Aerospace Conference, Big Sky Montana. pp. 1-7.
4. Beck, J., Mostow, J., 2008. How who should practice: Using learning decomposition to evaluate the efficacy of different types of practice for different types of students. In: Proceedings of the 9th International Conference on Intelligent Tutoring Systems. pp. 353-362.
5. Benjamins, V. R., 2014. Big data: From hype to reality? In: Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS14). ACM, New York, NY, USA, pp. 2:1-2:2.
6. Bhadani, A., 2011. Cloud Computing and Virtualization. 1,Saarbrucken: VDM Verlag Dr. Muller Aktiengesellschaft& Co. KG. 116 s. ISBN 9783639347777
7. Bhadani, A., and Chaudhary, S., 2010. Performance evaluation of web servers using central load balancing policy over virtual machines on cloud, Proceedings of the Third Annual ACM Conference, Bangalore, ACM Blount, M., Ebling, M., Eklund, J., James, A., McGregor, C., Percival, N., Smith, K., Sow, D., 2010. Real-time analysis for intensive care: Development and deployment of the Artemis analytic system. IEEE Engineering in Medicine and Biology Magazine 29 (2), 110-118.
8. Boyd, D., Crawford, K., 2011. Six provocations for big data. In: A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society.
9. Boyd, D., Crawford, K., 2012. Critical questions for big data. Information, Communication & Society 15:5, 662-679.
10. Brewer, E. A., 2000. Towards robust distributed systems (abstract). In: Proceedings of the Nineteenth Annual ACM Symposium on Principles of Distributed Computing (PODC '00). ACM, New York, NY, USA, p. 7.
11. Brown, B., Chui, M., Manyika, J., October 2011. Are you ready for the era of big data? McKinsey Quarterly, http://www.mckinsey.com/insights/strategy/are you ready for the era of big data, Accessed on: Jan 20, 2015.
12. Chen, M., Mao, S., Liu, Y., 2014. Big data: A survey. Mobile Networks and Applications 19 (2), 171-209.
13. Coumaros, J., de Roys, S., Chretien, L., Buvat, J., KVJ, S., Clerk, V., Auliard, O., 2014. Big data alchemy: How can banks maximize the value of their customer data? Capgemini Consulting White Paper https://www.capgemini.com/resources/big-data-customer-analytics-in-banks, Accessed on: March 20, 2015.
14. Dhar V., 2013. Data science and prediction. Communications of ACM 56(12), 64–73.

Facebook, 2015. http://www.statisticbrain.com/facebook-statistics/, accessed on February 20, 2016.

15. Fisher, D., DeLine, R., Czerwinski, M., Drucker, S., 2012. Interactions with big data analytics. Interactions 19 (3), 50-59.

16. Forsyth, C., January 2012. For big data analytics there"s no such thing as too big. Cisco White paper http://www.cisco.com/en/US/solutions/ns340/ns517/ns224/big data wp.pdf, Accessed on Feb 20, 2015.

17. Foster, I., Zhao, Y., Raicu, I., Lu, S., 2008. Cloud computing and grid computing 360-degree compared. In: Grid Computing Environments Workshop, 2008 (GCE '08). pp. 1-10.

18. Gandomi, A., Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 35 (2), 137-144.

19. Gantz, J., Reinsel, D., 2011. Extracting value from chaos. Tech. rep., IDC.

20. Garlasu, D., Sandulescu, V., Halcu, I., Neculoiu, G., Grigoriu, O., Marinescu, M., Marinescu, V., 2013. A big data implementation based on grid computing. In: 11th Roedunet International Conference (RoEduNet), Sinaia. pp. 1-4.

21. Gilbert, S., Lynch, N., 2002. Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services. SIGACT News 33 (2), 51-59.

22. Jin, X., Wah, B. W., Cheng, X., Wang, Y., 2015. Significance and challenges of big data research. Big Data Research 2 (2), 59-64.

23. Jothimani, D., Bhadani, A.K., Shankar, R., 2015. Towards Understanding the Cynicism of Social Networking Sites: An Operations Management Perspective, Procedia - Social and Behavioral Sciences, 189, 117–132

24. Kaisler, S., Armour, F., Espinosa, J., Money, W., 2013. Big data: Issues and challenges moving forward. In: 46th Hawaii International Conference on System Sciences (HICSS), Hawaii. pp. 995-1004.

Corresponding Author

Jyotsna Tiwari*

Research Scholar, Shri Krishna University, Chhatarpur M.P.