IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

Big Data Analytics: Problems, Research Gaps, and Resources

Ajay Chakravarty, Assistant Professor,
College of Computing Sciences and Information Technology, Teerthanker Mahaveer University,
Moradabad, Uttar Pradesh, India
Email Id- ajay.chakravarty1@gmail.com

ABSTRACT: ***Big data refers to information or data sets that are so large or intricate that distributed databases rather than conventional data processing technologies are needed. Companies like Google, eBay, LinkedIn, and Facebook's parent companies were big data going back in time. It is a group of enormous and complicated data sets that contain enormous amounts of data, real-time data management, social media analytics, dates, etc. Issues with sensor design, capture, collection, sharing, storing, analysing, and visualizing data Privacy of information, etc. Big data is a term for datasets that are varied and pace, making it incredibly challenging to use conventional methods and instruments. The research process large amounts of data to find hidden relationships known as big analytics of data. Big Data is a complicated subset of data. Terabytes of data are produced every day by modern information systems and digital technologies like the Internet of Things and cloud computing. To examine these massive volumes of data and acquire knowledge for decision-making, a lot of labor is required on many different levels. Consequently, large data on a recent topic of study and development are analysed. The fundamental paper's goal is to observe the potential effects of large open research questions, data difficulties, and related tools of data analysis. Consequently, this paper offers a platform for the exploration of large data on several levels. This study provides an overview of the problems, research gaps, and resources of big data analytics. It also creates a new horizon for researchers to create the remedy, depending on the difficulties and active research questions.***

KEYWORDS: ***Big Data Analytics, Database, Management, Cloud Computing.***

1. INTRODUCTION

Data are created in the modern day from several sources, and big data has grown rapidly due to the speedy growth of digital technology. It provides evolutionary advances in a variety of domains using big dataset sets. It often refers to a group of large and complex datasets that are difficult to manage using standard database administration methods, tools, or data processing software. These are provided in a configuration that exceeds petabytes and is semi-structured, unstructured, and structured. Its formal definition ranges from 3Vs to 4Vs. 3Vs stands for volume, velocity, and variety. While velocity is used for the massive volumes of data created each day, volume denotes the speed at which the data is growing are collected for analysis. Every digital activity and social media engagement generates big data. Data is sent via systems, sensors, and portable gadgets. [1]–[5].

Big data is now available from a variety of sources at a terrifying rate of speed, volume, and variety. Demand appropriate computing capacity, analytical capabilities, and expertise to get useful value out of large data. More assured accurate big data can help in decision-making. Good Decisions increase operational effectiveness and cost lowering and lowered risk. Data sets can be analyzed to discover new relationships, to "identify and avoid business trends diseases, fight crime, and so forth". Researchers, business leaders, media professionals, and Governments and advertising frequently interact with issues with massive data collection in several contexts, including Internet search, money, and business information technology. Data is coming in extremely quickly, so it needs to be processed right away. RFID sensors and smart meters are driving the need to deal with fast-moving data in close to real-time. Figure 1 illustrate the Big Data analysis classification.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

Organizations are responding to data on time velocity. Velocity described how quickly data was processed. For urgent procedures like detecting fraud, huge must employ data. It enters your business to increase its worth. There are many different forms of data. Data that is structured and numerical in conventional databases information produced by line-of-business software. Email, video, audio, and unstructured text documents financial exchanges. Controlling, combining, and governing there are many distinct types of data Organizations continue to struggle with. Various sorts and Data sources are available. Various data bursts, including audio, video, unstructured or semi-structured data, organized and historical data stored in enterprise storage, etc. Data flows can be extremely erratic with periodic peaks due to the rising velocities and variety of data. In social media, it is popular. Daily seasonal high data demands caused by events cannot be accommodated and managed. There is even more unstructured data present. The data can occasionally display discrepancies, which can hinder the administration and management of the data process correctly [6]–[10].

Sometimes, the data's consistency can be seen might impair the handling and management of the data correctly. Various sources provide data today. Data transformation, comparison, and linking across systems continue to be difficult. However, it is essential to interactions, organizational structures, and various data links. Alternatively, your data might easily be out of control, and spiral. When massive amounts of data arrive the data handling is quite complex from numerous sources complicated. Data must be linked, connected, and especially associated with users' ability to recognize the information messages that the data should transfer. Parallelization in data analysis applications comes rather readily. Building applications using the parallel approach is of interest to many programmers. Requirement for an intermediary shared RAM or disc infrastructure that costs a lot.

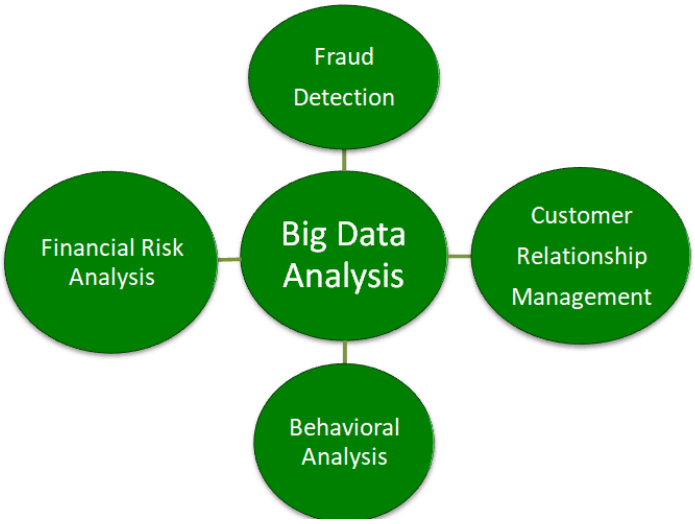

Figure 1: Illustrate the Big Data analysis classification [Geeks for geeks].

2. DISCUSSION

Utilizing several techniques, including radio frequency identification readers, mobile devices, aerial sensory technologies, remote sensing, etc. These records are kept at a great cost. However, they eventually destroyed or overlooked them because there was sufficient room to keep them. Consequently, the initial obstacle for big data analysis involves larger input/output ratios and storage mediums speed. Data accessibility in these circumstances must be of the maximum importance for knowledge representation and discovery. The fundamental justification is that it must be easily accessible and practical quickly to do further research.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

Data was formerly stored on hard disc drives by analysts, however, random input and output perform less well than sequential input and output. To avoid this phase, use the solid-state drive (SSD) concept and physical change memory (PCM). Finding and displaying knowledge in big data is a significant issue. It includes a variety of related fields, including representation, administration, information retrieval, archiving, and preservation. There are several instruments for information representation and discovery, like fuzzy sets "rough set," "soft set," "near set," and "formal," Principal component analysis and concept analysis are a few list a few.

In addition, numerous hybridization methods are built to handle challenges in real life. All of these strategies are contingent on the problem. Additionally, in several of these methods, huge datasets in a sequential computer wouldn't be appropriate. However, some of the strategies have positive qualities in terms of scaling on a parallel computer. Given that it is large data availability is growing exponentially. Analysis of large datasets requires more challenging computing operations. The key difficulty is handling the uncertainties and inconsistencies in the datasets. Typically, systematic is used to model the computational complexity. It is challenging to create an extensive mathematical framework that is often relevant to Big Data. However, specialized data Analyses can be carried out quickly by comprehending the specific complexities. A number of these developments potentially imitate large-scale data analytics for several topics. Several surveys and studies have been done utilizing machine learning in this direction. Methods that want the minimum amount of memory. The fundamental objective of this research is to reduce computational expense. Figure 2 illustrates the 5V'S of big data.

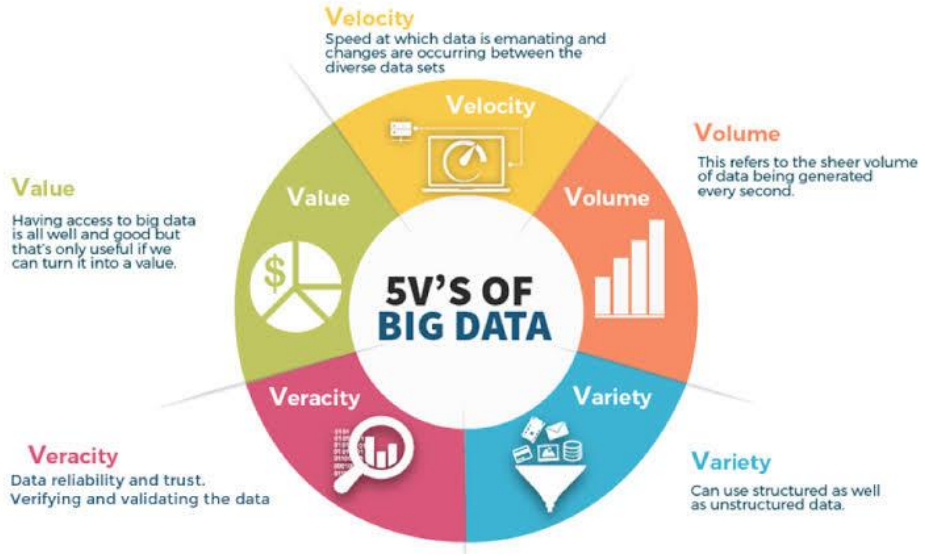

Figure 2: Illustrates the 5V'S of Big Data [Google].

Big data analytics and data science are dominating research in both academia and business. Data science aims to explore big data and extract insights from it. Applications of big data and data science, statistical learning, pattern recognition, and machine learning signal processing and storage include information science, uncertainty modeling, and uncertain data analysis. Future forecast movement of things will be made possible by an effective merging of technology and analysis. The main purpose of this part is to explore current issues with big data analytics research. Three broad categories of research issues involving large data analysis include cloud computing, the Internet of Things (IoT), quantum computing, and bio-inspired computing.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

Numerous personal characteristics, economic practices, cultural revolutions, and cross-border contacts have all been drastically altered by the Internet. The Internet of Things (IoT) is now being established by machines employing a variety of autonomous devices. As a result, appliances are utilizing the internet more and more, as people do with web browsers. Due to current research and challenges, the Internet of Things is becoming more and more well-known for its most fascinating possibilities. Future information, networks, and technology will be impacted in an essential social and economic way.

3. CONCLUSION

Data generation has substantially expanded during the last several years. It's difficult for a typical person to analyze this fact. To achieve this, we review the numerous research concerns technologies, and challenges involved in analyzing these huge data. According to this survey, every big data platform has its particular emphasis. Some of them are batch-oriented. Whereas some people excel in real-time analysis each big data platforms have particular features as well. Different statistical analysis is one of the methods utilized for the analysis. Data stream processing, cloud computing, quantum computing, data mining, and machine learning that in the future, researchers will focus more on innovative methods to efficiently address large data problems. Big data analytics have been used in a variety of settings and sectors. Thanks to big data analytics, healthcare costs might be reduced, treatments could be enhanced, and lives could be saved. It also helps businesses like financial institutions because it enables the study of client log files to aid in improving comprehension of consumer needs. The retail industry has a major impact. Big data analytics in this industry can again help managers comprehend the impact of society improved services are created in response to customer wants. Analyses of big data are also employed. In the telecommunications industry, they assist in keeping track of machine logs and resolving quality problems. Big data is a crucial area that offers several potential benefits and innovations.

REFERENCES:

[1] A. Oussous, F. Z. Benjelloun, A. Ait Lahcen, and S. Belfkih, “Big Data technologies: A survey,” *Journal of King Saud University - Computer and Information Sciences*. 2018. doi: 10.1016/j.jksuci.2017.06.001.

[2] C. W. Tsai, C. F. Lai, H. C. Chao, and A. V. Vasilakos, “Big data analytics: a survey,” *J. Big Data*, 2015, doi: 10.1186/s40537-015-0030-3.

[3] M. M. Hasan, J. Popp, and J. Oláh, “Current landscape and influence of big data on finance,” *J. Big Data*, 2020, doi: 10.1186/s40537-020-00291-z.

[4] W. A. Günther, M. H. Rezazade Mehrizi, M. Huysman, and F. Feldberg, “Debating big data: A literature review on realizing value from big data,” *J. Strateg. Inf. Syst.*, 2017, doi: 10.1016/j.jsis.2017.07.003.

[5] E. Al Nuaimi, H. Al Neyadi, N. Mohamed, and J. Al-Jaroodi, “Applications of big data to smart cities,” *J. Internet Serv. Appl.*, 2015, doi: 10.1186/s13174-015-0041-5.

[6] B. Ristevski and M. Chen, “Big Data Analytics in Medicine and Healthcare,” *J. Integr. Bioinform.*, 2018, doi: 10.1515/jib-2017-0030.

[7] D. Singh and C. K. Reddy, “A survey on platforms for big data analytics,” *J. Big Data*, 2015, doi: 10.1186/s40537-014-0008-6.

[8] J. Bughin, “Big data, Big bang?,” *J. Big Data*, 2016, doi: 10.1186/s40537-015-0014-3.

[9] Y. Wang, L. A. Kung, and T. A. Byrd, “Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations,” *Technol. Forecast. Soc. Change*, 2018, doi: 10.1016/j.techfore.2015.12.019.

[10] J. Fan, F. Han, and H. Liu, “Challenges of Big Data analysis,” *National Science Review*. 2014. doi: 10.1093/nsr/nwt032.