International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 12 | December - 2022 Impact Factor: 7.185 ISSN: 2582-3930

A Survey on Big Data Analytics: Challenges, OpenResearch Issues and Tools

Dr C K Gomathy-Assistant Professor, Department of CSE, SCSVMV Deemed to be University, India

Mr.S.Sai Krishna Chaitanya, Mr. R.Lakshmi Kartheek Kumar Reddy, Mr. R.Rahul, Mr.T.Sarath Kumar
UG Scholars- SCSVMV Deemed to be University, India

ABSTRACT

Internet of Things and cloud computing, a huge repository of terabytes of data is generated each day from modern information systems and digital technologies. For decision making, analysis of these massive data requires a lot of efforts at multiple levels to extract knowledge. Therefore, big data analysis is a current area of research and development. The basic objective of this paper is to explore the potential impact of big data challenges, open research issues, and various tools associated with it. As a result, this article provides a platform to explore big data at numerous stages. Based on the challenges and open research issues, it opens a new horizon for researchers to develop the solution additionally.

Keywords: Big data analytics; Hadoop; Massive data; Structured data; Unstructured Data

I. INTRODUCTION

Data are produced from a variety of sources, and the quick development of digital technology has contributed to the expansion of big data in the digital world. With the compilation of massive datasets, it offers evolutionary advancements in a variety of domains. It often refers to a collection of very vast and complicated datasets that are challenging to handle with conventional database administration software or data processing tools. These are accessible in petabytes and beyond in organised, semi-structured, and unstructured formats. From 3Vs to 4Vs is how it is defined formally. Volume, Velocity, and Variety are referred to as 3Vs. The vast quantity of data that are produced every day is referred to as volume, and the pace of increase and the speed at which the data are gathered for analysis is referred to as velocity. Structured, unstructured, semi-structured, and other forms of data are described by variety. Veracity, which combines availability and accountability, is the subject of the fourth V. Processing data with a large volume, velocity, diversity, and veracity while utilising a variety of conventional and computationally clever methodologies is the main goal of big data analysis. Gandomi and Haider have talked about a few of these information extraction techniques

II. CHALLENGES IN BIG DATA ANALYTICS

Big data has grown in recent years in a number of industries, including healthcare, public administration, retail, biochemistry, and other multidisciplinary scientific projects. Big data is regularly encountered by web-based applications, including social computing, online text and documents, and internet search indexing. While internet search indexing includes ISI, IEEE Xplorer, Scopus, and Thomson, social computing encompasses social network analysis, online communities, recommender systems, reputation systems, and prediction markets.

We need to understand different computational difficulties, information security, and computational methods to evaluate large data in order to meet the problems. For instance, many statistical techniques that work well with little amounts of data do not scale to large amounts of data. Similar issues arise when evaluating large amounts of data using various computer algorithms that excel at handling little data. Many scholars studied the many difficulties that the health industry faces. The difficulties associated with big data analytics are categorised here into four main groups: data analysis and storage; knowledge discovery and computational complexity; scalability and data visualisation; and information security. In the subsections that follow, we briefly address these concerns.

A. Data Storage and Analysis

Through a variety of tools, including mobile devices, aerial sensory technologies, remote sensing, radio frequency identification readers, etc., data size has drastically increased in recent years. These data are kept at a high cost, but in the end they are ignored or erased since there is not enough room to keep them. Therefore, faster input/output speeds and storage media are the first obstacle to large data analysis. In these situations, the data accessibility must be given top importance for knowledge representation and discovery. The main reason is that material has to be quickly and easily accessible for additional examination. Hard disc drives have been used by analysts to store data for decades, but their poorer random input/output performance than sequential input/output is a drawback.

The variety of data is an additional difficulty for big data analysis. The number of data mining tasks has substantially expanded with the continuous growth of datasets. Additionally, while working with enormous datasets, data reduction, data selection, and feature selection are crucial tasks.

B. Knowledge Discovery and Computational Complexities

Large dataset analysis necessitates increasingly complicated computing tasks. Dealing with the uncertainties and discrepancies in the datasets is the main challenge. In most cases, computational complexity is modelled systematically. Establishing a complete mathematical framework that is universally applicable to Big Data may be challenging. However, if the relevant complexity are understood, domain-specific data analytics may be completed quickly. Such developments might be used to emulate big data analytics in several fields. The least memory-intensive machine learning algorithms have been used in a lot of research and surveys in this area. The primary goal of this study is to reduce computing complexity and expense.However, current big data analysis tools have poor performance in handling computational complexities, uncertainty and inconsistencies. It leads to a great challenge to develop techniques and technologies that can deal computational complexity, uncertainty, and inconsistencies in an effective manner.

C. Scalability and Visualization of Data

Scalability and security are the biggest problems facing large data analysis approaches. In recent years, academics have focused on speeding up data analysis and the computers that handle it in line with Moore's Law. For the former, sampling, online, and multiresolution analytic approaches must be developed. In the area of large data research, incremental approaches provide high scalability properties. There is a natural dramatic change in processor technology as embedded with an increasing number of cores since data capacity is rising considerably faster than CPU speeds. Parallel computing emerges as a result of this change in processors. Parallel computing is necessary for real-time applications such as navigation, social networks, banking, internet search, timeliness, etc.

D. Information Security

Massive amounts of data are correlated, examined, and mined for useful patterns in big data analysis. Various organisations have different rules in place to protect their private data. In large data analysis, information preservation is a crucial problem. Big data comes with a significant security risk. Information security is therefore becoming into a big data analytics issue. The use of mechanisms for authentication, authorisation, and encryption can improve the security of massive data. Big data applications must contend with a number of security measures, including the size of the network, the variety of devices, real-time security monitoring, and

the absence of an intrusion detection system. Information security has focused on the security issue brought on by large data. A multi-level security policy model and preventative mechanism must be developed as a result.

III. OPEN RESEARCH ISSUES IN BIG DATA ANALYTICS

Research in companies and academia is increasingly focusing on big data analytics and data science. The goal of data science is to study massive data and knowledge extraction from it. Information science, uncertainty modelling, uncertain data analysis, machine learning, statistical learning, pattern recognition, data warehousing, and signal processing are a few examples of applications for big data and data science. In order to accurately predict the future course of events, technologies and analyses must be integrated effectively. This section's primary goal is to address open-research concerns related to big data analytics. The internet of things (IoT), cloud computing, biologically inspired computing, and quantum computing are the three broad categories into which the research questions relevant to large data analysis are divided. But it's not only about these things. Husing Kuoetal has additional studies on big data in healthcare-related concerns.

IV. CONCLUSION

In recent years, data production has increased dramatically. For a regular individual, analyzing this facts presents a challenge. In order to do this, we examine the many research problems, difficulties, and analytical techniques in this study. It is clear from this poll that each big data platform has a distinct focus. While some of them excel in real-time analytics, others are better suited for batch processing. Additionally, each big data platform offers particular features. Statistical analysis, machine learning, data mining, intelligent analysis, cloud computing, quantum computing, and data stream processing are some of the several approaches utilised for the analysis. We think that in the future, academics will focus more on these methods to effectively and efficiently address big data issues.

V.REFERENCES

[1] DR.C.K.Gomathy , V.Geetha , S.Madhumitha , S.Sangeetha , R.Vishnupriya Article: A Secure With Efficient Data Transaction In Cloud Service, Published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 4, March 2016, ISSN: 2278 – 1323.

[2] Dr.C.K.Gomathy,C K Hemalatha, Article: A Study On Employee Safety And Health Management International Research Journal Of Engineering And Technology (Irjet)- Volume: 08 Issue: 04 | Apr 2021

[3] Dr.C K Gomathy, Article: A Study on the Effect of Digital Literacy and information Management, IAETSD Journal For Advanced Research In Applied Sciences, Volume 7 Issue 3, P.No-51-57, ISSN NO: 2279-543X,Mar/2018

[4] Dr.C K Gomathy, Article: An Effective Innovation Technology In Enhancing Teaching And Learning Of Knowledge Using Ict Methods, International Journal Of Contemporary Research In Computer Science And Technology (Ijcrcst) E-Issn: 2395-5325 Volume3, Issue 4,P.No-10-13, April ’2017

[5] Dr.C K Gomathy, Article: Supply chain-Impact of importance and Technology in Software Release Management, International Journal of Scientific Research in Computer Science Engineering and Information Technology (IJSRCSEIT) Volume 3 | Issue 6 | ISSN : 2456-3307, P.No:1-4, July-2018.

[6] C K Gomathy and V Geetha. Article: A Real Time Analysis of Service based using Mobile Phone Controlled Vehicle using DTMF for Accident Prevention. International Journal of Computer Applications 138(2):11-13, March 2016. Published by Foundation of Computer Science (FCS), NY, USA,ISSN No: 0975-8887

[7] C K Gomathy and V Geetha. Article: Evaluation on Ethernet based Passive Optical Network Service Enhancement through Splitting of Architecture. International Journal of Computer Applications 138(2):14-17, March 2016. Published by Foundation of Computer Science (FCS), NY, USA, ISSN No: 0975-8887

[8] C.K.Gomathy and Dr.S.Rajalakshmi.(2014), "A Software Design Pattern for Bank Service Oriented Architecture", International Journal of Advanced Research in Computer Engineering and Technology(IJARCET), Volume 3,Issue IV, April 2014,P.No:1302-1306, ,ISSN:2278-1323.

[9] C. K. Gomathy and S. Rajalakshmi, "A software quality metric performance of professional management in service oriented architecture," Second International Conference on Current Trends in Engineering and Technology - ICCTET 2014, 2014, pp. 41-47, doi: 10.1109/ICCTET.2014.6966260.

[10] Dr.C K Gomathy, V Geetha ,T N V Siddartha, M Sandeep , B Srinivasa Srujay Article: Web Service Composition In A Digitalized Health Care Environment For Effective Communications, Published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 4, April 2016, ISSN: 2278 – 1323.

[11] C.K.Gomathy.(2010),"Cloud Computing: Business Management for Effective Service Oriented Architecture" International Journal of Power Control Signal and Computation (IJPCSC), Volume 1, Issue IV, Oct - Dec 2010, P.No:22-27, ISSN: 0976-268X .

[12] Dr.C K Gomathy, Article: A Study on the recent Advancements in Online Surveying , International Journal of Emerging technologies and Innovative Research (JETIR) Volume 5 | Issue 11 | ISSN : 2349-5162, P.No:327-331, Nov-2018

[13] Dr.C.K.Gomathy,C K Hemalatha, Article: A Study On Employee Safety And Health Management International Research Journal Of Engineering And Technology (Irjet)- Volume: 08 Issue: 04 | Apr 2021

[14] Dr.C K Gomathy, V Geetha , T.Jayanthi, M.Bhargavi, P.Sai Haritha Article: A Medical Information Security Using Cryptosystem For Wireless Sensor Networks, International Journal Of Contemporary Research In Computer Science And Technology (Ijcrcst) E-Issn: 2395-5325 Volume3, Issue 4, P.No-1-5,April ’2017

[15] C.K.Gomathy and Dr.S.Rajalakshmi.(2014), "Service Oriented Architecture to improve Quality of Software System in Public Sector Organization with Improved Progress Ability", Proceedings of ERCICA-2014, organized by Nitte Meenakshi Institute of Technology, Bangalore. Archived in Elsevier Xplore Digital Library, August 2014, ISBN:978-9-3510-7216-4.

[16] Parameshwari, R. & Gomathy, C K. (2015). A Novel Approach to Identify Sullied Terms in Service Level Agreement. International Journal of Computer Applications. 115. 16-20. 10.5120/20163-2253.

[17] C.K.Gomathy and Dr.S.Rajalakshmi.(2014),"A Software Quality Metric Performance of Professional Management in Service Oriented Architecture", Proceedings of ICCTET’14, organized by Akshaya

College of Engineering, Coimbatore. Archived in IEEE Xplore Digital Library, July 2014,ISBN:978-1-4799-7986-8.

[18] C.K.Gomathy and Dr.S.Rajalakshmi.(2011), "Business Process Development In Service Oriented Architecture", International Journal of Research in Computer Application and Management (IJRCM) ,Volume 1,Issue IV, August 2011,P.No:50-53,ISSN : 2231-1009

AURTHOR'S PROFILE

1. S.Sai krishna chaitanya Student, B.E. Computer Science and Engineering, Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. His Area of Interest in Big Data Analytics.
2. R. Lakshmi Kartheek kumar reddy Student, B.E. Computer Science and Engineering, Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. His Area of Interest in Big Data Analytics.
3. R.Rahul Student, B.E. Computer Science and Engineering, Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. His Area of Interest in Big Data Analytics.
4. T.Sarath kumar Student, B.E. Computer Science and Engineering, Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. His Area of Interest in Big Data Analytics.
5. Dr.C.K.Gomathy is Assistant Professor in Computer Science and Engineering at Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India.