IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115125

Divergent Big Data Tools and Its applications in Different Domains

Vaishali B. Bhagat[1], Dr. V. M. Thakare[2]

[1]Research Scholar- Computer Science & Engineering, Sant Gadge Baba Amravati University, Amravati Maharashtra 444606 India

[1]Asst. Prof, P. R. Pote College of Engineering & Management Amravati

[2]Professor & Head, P G Department of Computer Science & Engineering, Sant Gadge Baba Amravati University, Amravati 444606 India

Abstract: The proposed paper surveys the big data usage in multiple sectors and calls attention to big data tools and big data analytics in numerous fields. In the proposed paper, the usage of big data and its tools have been discussed and numerous examples have been shown. Regarding the efficient tools in big data, more than 10 papers have been reviewed and their summaries and important takeaways are mentioned in the literature survey of the paper. This paper will provide an overview to the researchers and practitioners about the extensive usage of big data and its tools. Use of big data in the healthcare sector has been explained along with the names of the sources where data is collected from. This paper explains the importance of big data in our life and details the numerous ways in which it solves real-life issues. Big data tools such as Apache, Hadoop and HPCC has been highlighted in this paper with their attributes and advantages.

Keywords: Big data, Hadoop, usage, medical-sector, dataset, big data tools

1. INTRODUCTION

With the exponential growth and development in technology, the rise of data which is present online has been in immense amounts and big data is the data present online and is growing tremendously with time. There are multiple tools introduced in order to use big data in organizations and firms. Extensive usage of big data calls for the methods which can be efficient, and the traditional methods available to process the big data are not sufficient in providing the efficiency which is required by the firms. With the failure of traditional methods, numerous new methodologies and big data tools have been introduced and they are widely used for processing and analyzing the big data. To process the huge set of data present online i.e., Big Data Hadoop is one of the online tools which is designed in a way that it is used for scaling up single servers to numerous machines. Hadoop is known for providing flexibility and speed in processing the data. One of the big data tools that provides all in one access to the global range of platforms is Atlas. It is utilized in the projects containing numerous coded data segments and more than thousands of documents. Tool which is noticeably the one of the most efficient big data tools is HPCC and it finishes large tasks with fewer code. Another advantageous attribute of HPCC is that it offers availability and high redundancy. Other famous big data tools are Storm, Qubole, Cassandra, Stats iQ, CouchDB, Pentaho, Flink, Cloudera, Openrefine, Rapidminer, DataCleaner, Kaggle,Hive etc.

Figure 1: Big Data tools

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115125

Big data is playing a vital role in industries and so many other organizations. Production of data in bulk from numerous different sectors such as electronic records, smart devices, wearables, handheld devices, healthcare information systems, and so on. Development in the computational techniques in the area of information technology and the rapid growth in medical big data has allowed the researchers and practitioners to obtain the big data after extraction and visualize it in a whole new rainbow of use. To manage the data and handle it carefully, different tools which are mentioned in the above paragraph are used.

Figure 2: Usage of big data in different sectors

1.1 Organization of the paper

Section 2 of the proposed paper is the literature survey which details the usage of big data in numerous different fields and it also explains about the tools and their performances. Furthermore, section 3 presents the conclusion and the future scope and marks the end of the paper.

1.2 Contribution of the paper

- The paper provides details about big data and its various tools and compares them on the basis of their performance.
- Furthermore, the paper explains the application of big data in different sectors.
- The paper provides an overview of the datasets and how the analysis is carried for huge data using various big data tools.
- The literature survey provided in the paper comprises useful information which has been extracted from different research articles.
- Additionally, the paper also gives an insight of future works which could be done and the applications of big data can be enhanced so that it can be more beneficial.

2 . LITERATURE SURVEY

In the proposed literature survey, more than 10 papers have been analyzed and reviewed and afterwards the essential and key takeaways have been explained. Various Big data tools and the usage of big data in numerous domains have been detailed in the survey. Additionally, the usage of big data in the medical sector has been explained and it also details about the numerous huge datasets and their extraction and analysis using multiple big data tools available.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115125

According to [1], Earlier industries used manual methods to conduct the audits but now with the advancement of technology, computerized systems like big data analytics are used. To overcome the redundancy of identical data, a big frame is used. The data comes from multiple sources and to process large amounts of data which is coming from various sources would be difficult with traditional methods available. The advantages provided by Bigdata to companies make big data more preferable to use in their business processes, [1]. According to a survey, the percentage of companies who planned to use big data by the end of 2018 was 75%. To produce different kinds of diverse information, big data analytics can be carried out which will reveal the connections and patterns which exist in big data. The failure of traditional methods to audit comes due to the rapid growth of data, [1]. Bigdata makes the process more efficient and it provides a clearer picture of things which are done by the company which positively affect the relevant decision making. Frequent updating and training in the use of technology and dataset will make it easier to obtain big data faster and accurately,[1].

Big data is the data generated at an exponential rate [2]. It has different specifications and characteristics such as structured, semi-structured and unstructured data formats. According to [2], their business contains various valuable information from the numerous different types of stakeholders depending upon their needs and it is not possible to process and meet them by the use of traditional tools and methods. Here Big data technology comes into picture and plays a vital role to handle, process, store the huge amount of data. By providing analytics and techniques which are predictive, big data analytics assists entrepreneurs and companies to make more subtle and informed decisions related to business. According to [2], Apache spark can process different types of data, and not only process the high-volume data but it also concentrates on accelerating the workloads of batch processing. The prime advantage of using Apache spark is its ability to get deployed as a separate cluster or it can be combined with an already present Hadoop cluster,[2].

According to [3], size plays a crucial role in big data and it should be considered one of the defining characteristics. Usually, the users of computers are habituated to think of data in terms of megabytes and gigabytes, but the requirement of big data is much larger units, [3]. It was predicted in [3] that Bigdata will emerge as a key basis of competition and growth for individual firms and that if the current situation is taken into consideration the prediction stood out to be 100% true. The role of big data analytics in science research is ever increasing. Due to the cheap rate and small size of sensors, carrying out scientific experiments depends majorly on collecting so much data in order to get analyzed only by big data techniques, [3].

Usage of big data in large or medium sized organizations enables them to generate content which is user-obtained from the users and it also enables them to collaborate with customers, patterns and suppliers effectively on social media channels, [4]. According to [4], insights from the content which is user generated and collaboration with the customers is critical for the victory in the era of social media. It is mentioned in [4] that US healthcare can create more than $300 billion every year if it plans to use Bigdata effectively and creatively to drive quality and efficiency.

For companies to gain deeper and richer insights and gain an advantage in the competition, big data needs to be implemented, analyzed and executed as accurately as it can get, [5]. According to [5], McKinsey Global Institute presented the potential of big data in five major areas, they are Health Care, public sector, retail manufacturing and personal location data. In healthcare, clinical decision support systems, personalized medicine, applying individual analytic for patient profile analyzing disease patterns etc. are majorly using big data and its applications,[5].

Big data analytics is a process where the advanced analytical techniques are applied on the big data sets and to reveal and leverage the business change analytics based on large data becomes crucial,[6]. To use big data in healthcare, data is generated from various heterogeneous places such as clinical data, hospital operations, pharmaceutical data, patients symptoms etc[,7]. Collection of data in this way enables the health services to become more personalized, [8]. Usage of big data helps in optimizing the hospital operations and it lowers down the overall health cost expenditure,[9]. Big analysis which is provided by big data opens the doors for big opportunities which advances the quality of life and also it enables the world to solve many mysteries problems,[10].

One of the best concerns of the world is healthcare and big data has become an essential part of healthcare in terms of implying data-sets that are electronic and are recognized with healthcare and prosperity of the patient, [11]. In healthcare, the data has been growing excessively. The main reason for expansion and development of data in healthcare is the development of cell phones and sensors with web association,[11]. According to the paper [11], production of data by this method is the perfect resource for enterprises in interpreting the business procedures policies. It is mentioned in the paper that, to prepare and divide humongous amounts of data, cloud facilities were used. As a next step, the cloud services transformed the big data model in a way that it handles the on-demand controlling,[11].

DOI: 10.17148/IJARCCE.2022.115125

In the research carried out in [11], the major six challenges faced in big data were applied ontology, Security, Storage and transport, storage and transport, accessibility, inconsistencies and mobility. Apart from these challenges, another issue which is highlighted in the paper,[11] is the data management issue. According to [11], it is the most complicated issue while dealing with the enormous amount of data and information. In management, the issues which are declared are crucial are data privacy, governance, and ethical issues [12].

Big Data is one of the latest and recent technique used for analysis the data. The big data is available with low-cost and new management software tools have created a data analysis history. Various tools such as Apache Hadoop, Apache Spark, Apache Drill, Dryad, Knime, Apache Storm and MongoDB are used for big data analysis. [13].

The data is generated over the recent years for analysis data and try to overcome the Challenges and predict the accurate which can improve the business growth . They have reviewed about the challenges of tools while analyzing the data [14].

The MongoDB is very interesting NoSQL application that uses C, C++ and JavaScript and other indexing replication and load balancing which also involves the file storageand aggregation such as server side and java script application and script collection [15].

CONCLUSION AND FUTURE SCOPE:

The proposed paper begins with explaining the definition of big data and then it explains how traditional methods are not sufficient to process and analyze the huge data sets and then the new technologies and methods have been detailed in the paper with their advantages and working tendencies. Furthermore, use of big data in various different sectors has been detailed in the paper. Hadoop, Apache spark and HPCC were also highlighted in the paper. Various sources from which the big data is gathered to be used in healthcare too has been discussed. As big data is majorly about the data, so if the gathered data gets more organized and precise then it would be more efficient and effective to process and analyze the and use.

REFERENCES:

1. Bmbang Leo Handoko, Archie Nathanael Mulyawan, Jonathan Tanuwijaya, Fransiska Tanciady, Big Data in Auditing for the Future of Data Driven Fraud Detection, IJITEE, Volume-9, Issue-3, January 2020
2. R Rawat and R Yadav, Big Data: Big Data Analysis, Issues and Challenges and Technologies, ResearchGate,2021 IOP Conf. Ser.: Mater. Sci. Eng. 1022 012014
3. Harry E. Pence, What is Big Data and Why is it Important?, Journal of Educational Technology Systems, ResearchGate,43(2),2014-15
4. Sachchidanand Singh, Nirmala Singh, Big Data Analytics, IEEE,ICCICT, Oct. 19-20,2012
5. Seref Sagiroglu, Durga Sinanc, Big Data: A review, IEEE, CCT,,20-24 May, 2013
6. Nada Elgendy, Ahmed Elragal, Big Data Analytics: A Literature Review Paper, lecture notes in computer science, vol8557, Springer, ICDM 2014
7. Ahmed Oussous, Samir Belfkih, Fatima-Zahra Benjelloun, Ayoub Ait Lahcen, Big Data technologies: A survey, sciencedirect, Vol 30, issue 4, oct 2018
8. Ishwarappa, J. Anuradha, A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology, ScienceDirect, vol 48, 2015
9. Rabie Ramadan, Big Data Tools-An Overview, ResearchGate, IJCSSE, dec 29, 2017
10. Deepak Kumar Verma, A Review on Big Data Analytics Tools in Context with Scalability, ResearchGate, IJCSE,Feb 2019
11. Memon, Mashooque & Soomro, Safeeullah & Jumani, Awais & Kartio, Muneer. (2017). Big Data Analytics and Its Applications. Annals of Emerging Technologies in Computing. 1. 10.33166/AETiC.2017.01.006.
12. Popli, Sheena J. and Mahore, Tushar R. and Karale, Nikhil E. and Pande, Dr. Sagar, A Survey on Bigdata in Healthcare (July 12, 2021). Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2021, Available at SSRN: https://ssrn.com/abstract=3884830 or http://dx.doi.org/10.2139/ssrn.3884830
13. V. B. Bhagat, V. M. Thakare, Review on Big Data Analysis Tools (19 -20 February 2022). 5th International Conference On Innovative Computing & Communication (ICICC-2022), Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4031901
14. D. P. Acharya, Kauser Ahmed P. "A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools" published in International Journal of Advanced Computer Science and Applications in 2016
15. V. B. Bhagat, V. M. Thakare, A NOVEL APPROACH FOR BIGDATA ANALYSIS FRAMEWORK FOR HEALTHCARE. First International Conference on recent trends on Management, Engineering & Science (ICMES) – 2021)