OPEN ACCESS

Volume: 3

Issue: 3

Month: September

Year: 2024

ISSN: 2583-7117

Published: 18.09.2024

Citation:

Mrs. Elavarasi Kesavan "Big Data Analytics: Tools, Technologies, and Real-World Applications – A Review" International Journal of Innovations in Science Engineering and Management, vol. 3, no. 3, 2024, pp. 120–126.

DOI:

10.69968/ijisem.2024v3i3120-126

Big Data Analytics: Tools, Technologies, and Real-World Applications – A Review

Mrs. Elavarasi Kesavan[1]

[1]*Full Stack QA Architect, Cognizant*

Abstract

The need for data analytics and storage has grown exponentially due to the Internet's and the digital economy's quick expansion, and IT departments are finding it very difficult to safeguard and analyze these larger amounts of data. Big Data analytics optimizes client retention and creates new business possibilities by gleaning value from these massive amounts of data. Examine the many studies on tools, methods, and practical applications of big data analytics in this page. It concluded that Big Data Analytics has revolutionized multiple sectors, particularly healthcare, by enabling real-time, data-driven decision-making. Leveraging tools like Hadoop, Map Reduce, and cloud computing, organizations can efficiently manage and analyze vast volumes of structured and unstructured data. Languages such as Python and R, along with techniques like text mining and machine learning, enhance data integration and insight generation. The COVID-19 pandemic highlighted the critical role of big data in driving smarter decisions across industries. As data continues to grow exponentially, Big Data Analytics remains central to innovation, operational efficiency, and informed strategies in the digital age.

Keywords; Big Data Analytics (BDA), Data analysis, Cloud computing, Apache Hadoop, Data Mining, Natural Language Processing (NLP), etc.

INTRODUCTION

Imagine living in a world without data storage, where every transaction, every piece that is recorded, and every information about a person or organization are erased immediately after use. Organizations may become less able to get critical data and insights, conduct thorough analyses, and provide new opportunities and advantages [1]. Names and addresses of clients, things that are available, transactions, employing employees, and other details are increasingly essential to continuity, even on a regular basis. Data is the foundation upon which all successful businesses are built [2]. Now think about how much information and data has been made available by technical and Internet advancements, and how detailed the details are. Massive amounts of data are readily accessible because to technological advancements in data collection and storage. The amount of data generated is increasing every second, and it must be kept and assessed in order to be used for value extraction. Furthermore, data is less costly to maintain, enabling businesses to fully use the vast amounts of data gathered [3].

"Big data analytics" is required to manage and analyse data proactively. A user has the ability to evaluate and derive insights that are beneficial for the development of new discoveries and to streamline the selection process within an organisation. Thanks to the cloud's scalable infrastructure, elastic resource allocation, and pay-as-you-go model for managing data applications and big data storage [4]. Cloud computing helps with big data analytics by enabling "quicker analysis, quicker outcomes, and more valuable data". "High-performance and scalable computing systems" are used under data and knowledge discovery methodologies to get over the cloud's constraints and gather crucial data and knowledge [5]. Through forecasting, text mining, predictive analytics, high-performance data mining, and optimisation, big data analytics helps businesses to easily and quickly assess their data.

Big data analytics has the potential to improve corporate marketing and advertising results, uncover untapped financial possibilities, boost customer happiness, improve efficiency, reduce risks, and achieve other goals. It may be difficult to deal with more data, even if it yields better results [6], [7].

Big data analytics

Big data analytics, or BDA, is the critically important process of identifying information that could be advantageous to enterprises, including "concealed models, unidentified correlations, market trends, and consumer preferences". In order to increase operational efficiency and achieve competitive advantages, this analytics process necessitates the deployment of tools and technology [8]. Strategic potential also generates new income sources. In addition to assisting us in comprehending the data, BDA also helps us identify the information that is most crucial for the company and future choices. These days, businesses keep their data on-site, in multi-cloud settings, and in data storage collections that don't fully adhere to different security regulations [9]. By dismantling old storage facilities and enabling the collecting of data from distant sources, cloud computing makes it feasible to prepare data for analysis : The high-performance data extraction, analytics, text categorisation", predictions, and improvement aspects of Big Data Analytics assist businesses in efficiently assessing their data or information and providing real-time analysis. There are several technologies and strategies that may assist save time and money while offering insightful business information, but no one technology can fully capture big data analytics [10].

Figure 1 Process of Big Data Analytics [11]

Big data analytics tools

Domo: In order to provide real-time insights and intuitive visualizations, Domo consolidates data from various sources into a single dashboard. Domo facilitates seamless collaboration and fast decision-making for teams of all sizes by emphasizing user-friendly design.

Apache Hadoop: Designed to distribute duties across multiple systems in order to manage and process large datasets, Hadoop is an open-source framework. An excellent choice for large data storage and processing, it is highly flexible and supports a variety of data types.

Apache Spark: Fast and in-memory, Apache Spark is an excellent tool for machine learning and real-time analytics. Businesses can swiftly extract insights from data, evaluate streaming data, and perform complicated queries with its assistance.

Big Query: Big Query is a server less, fully managed data warehouse from Google Cloud that is intended for quick SQL-based analytics. It is an excellent option for companies that want rapid, affordable access to large data sets because of its scalable nature and adjustable price.

Big data analytics techniques

A/B Testing: To determine which variable adjustments will provide the greatest results, A/B testing compares a control group to a test group. In email marketing, for instance, this kind of data analysis technique is often used to test various subject lines or graphics. Whichever option the algorithm determines is producing the highest degree of interaction is selected [12].

Data Mining: Data mining uses statistical techniques and machine learning to identify patterns in massive data collections. Finding the best items for a certain subset of your complete client database by analyzing customer behavior is a fantastic application of data mining.

"Regression Analysis": An independent variable's influence on a dependent variable is determined through regression analysis, a statistical method. It determines how changes in the independent variable affect a dependent variable. For example, it may be utilised to ascertain the relationship between client loyalty and customer satisfaction levels.

Natural Language Processing (NLP): It is within the domains of artificial intelligence and computer science that algorithms are implemented in "natural language processing (NLP)". There are several applications for this, including text-to-talk, where phone assistants, such as Apple's Siri, may record what people say aloud and then provide a suitable response [13].

Cluster Analysis: When trying to find patterns in a lot of data, this kind of analysis is used. It works by combining

data pieces that are somewhat similar to one another. Additionally, it may be used to contextualize trends. It's challenging to comprehend every individual's behavior while dealing with a big clientele. Cluster analysis is a method of grouping information according to certain criteria, such as buying patterns, demographics, or other factors [14].

Time Series Analysis: Data is analyzed within a predetermined time frame through time series analysis. It is beneficial to observe the impact of various variables on outcomes at various time intervals. It is highly effective in comprehending the effects of seasonality on consumer decisions.

Real world application of big data analytics

"Tracking Customer Spending Habit, Shopping Behavior": Data regarding client purchasing patterns, browsing behaviour, and their most favourite goods must be maintained by the executive team of large retail stores, including "Amazon, Walmart, and Big Bazar". This information is critical for the retention of these products in the store. According to the data, the production/collection rate of the product that is being searched/sold the most is determined [15].

Recommendation: In order to provide suggestions to their clients, many retail enterprises keep track of their spending and purchase habits. Flipkart, Walmart, and Amazon are among the e-commerce platforms that endorse products. In order to recommend that kind of product to a customer, they monitor of the items that the customer is interested in.

"Smart Traffic System": Roadside cameras, GPS devices in vehicles (like those in Ola or Uber cabs), and cameras at the city's entry and exit points are used to collect data on traffic conditions on different highways. All of these data are examined, and methods that are jam-free and less time-consuming are suggested. The city might develop a smart transportation system using big data analysis. Another advantage is lowering fuel consumption [16].

Secure Air Traffic System: There are sensors at many flying locations (such as propellers, etc.). Data like as temperature, moisture content, flight speed, and other environmental conditions are recorded by these sensors. The analysis of such data is used to set up and modify an environmental parameter during flight [17].

LITERATURE REVIEW

(Al-Sai et al., 2022) [18] This article reviews previous and post-COVID-19 research on large data applications. The four well-known industrial sectors of banking, healthcare, education, and transportation are included to the comparison. In light of this, this article explains the significance of big data applications in the COVID-19 age and aligns them with the appropriate big data analytics frameworks. This research indicates that the major constraints that companies experienced during one of the most devastating pandemics in history may be overcome by using the best big data apps and the data analytics models that go along with them. Furthermore, future research will examine the critical obstacles that "Big Data Analytics and applications" face during the COVID-19 pandemic.

(Batko & Ślęzak, 2022) [11] This document endeavors to evaluate the potential of Big Data Analytics in the healthcare sector. The research successfully verified that medical facilities are currently engaged in the processing of both structured and unstructured data. One may differentiate between a number of different kinds and sources of data, including transaction data, unstructured email and document content, data from sensors and devices, and database data. Nevertheless, they use less social media data since they are looking for analytics for their operations, which include financial and administrative tasks as well as therapeutic uses. This plainly demonstrates that medical facilities make decisions that are heavily influenced by data. The investigation's results support the benefits of data-based healthcare and the literature's interpretation that medical institutions are making the switch.

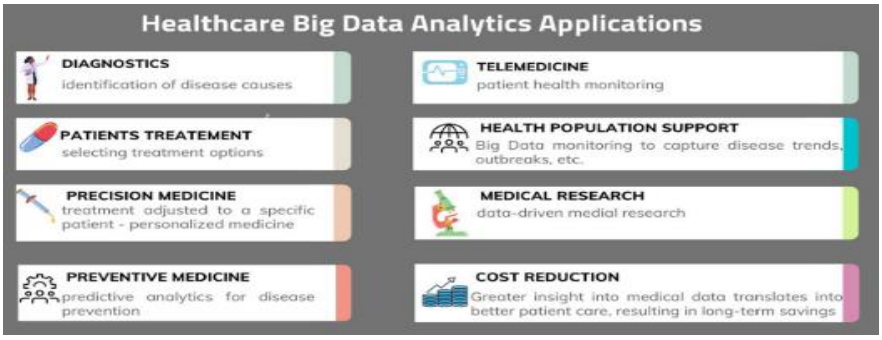

Figure 2 Healthcare Big Data Analytics applications [11]

(Khan et al., 2022) [19] The current literature is examined and evaluated in this study in order to find any gaps that might impair the functionality of the healthcare apps that are already on the market. It also seeks to provide improved ways to close these gaps. Despite the diligent efforts made in the fields of healthcare big data analytics, the results of this systematic mapping suggest that "the more recent hybrid machine learning-based systems and cloud computing-based models" should be utilised to reduce treatment costs, simulation times, and improve care quality. The ability of physicians, practitioners, researchers, and policymakers to utilize this study as support for further research will also be improved by this methodical mapping.

Table 1 Examples of Big Data applications by field before [19]

Field in Charge	Application Name	Description
Health	Ebola Open Data Initiative HealthMap	An open-source worldwide model has been developed using data from West Africa to monitor the number of Ebola cases in 2014.
	HealthMap	An early warning system for the appropriate reaction and a platform for visualizing illness trends
	Proactive listening, mobile phone-based system	Brazil should oversee the problem of bribery in the healthcare system, deal with any associated problems, and act swiftly and decisively to combat corruption.
Education	ENOVA	Mexican students' interactions can be predicted and analyzed through the application of data and data analytics. This, in turn, enhances the educational strategies and the instruments and techniques that are employed in the teaching-learning process.
	(PASS) Personalized Adaptive Study Success	The Open University Australia offers a more individualized learning environment in addition to course content prediction. Analyzing each student's unique attributes in addition to other student-related data gathered from other systems forms the foundation of the predictive data analytics model. Providing a more personalised environment is the primary objective of the application in order to ensure student engagement, retention, and participation in an internet-based educational setting.
Transportation	OpenTraffic platform	An application that analyzes and visualizes data from cellphones and cars to inform choices about urban infrastructure in both historical and current traffic scenarios.
		The program facilitates the journey from the starting point to the destination for night bus drivers in Seoul, South Africa. Data will be collected from a vast number of calls, text messages, and corporate and private taxi data sources in order to do this.
Banks	Avaloq, Finnova, SAP, Sungard and Temenos	In Singapore, OCBC is the biggest bank by market capitalization. It has operations in over fifteen countries worldwide. It serves as an example of successful BDA use. For example, the bank reacted to consumer behavior, demographic profiles, and customized events. As a result, OCBC Bank is able to increase customer happiness and engagement by 20% when compared to a control group.

(Yadav, 2022) [10] For an enterprise's decision-making process, big data analytics is utilized to uncover valuable insights buried in the data. Big data analysis skills are required to learn from unstructured online data, “such as text, photos, videos, and social media postings”. There is a growing demand for big data analytics, and in response, a plethora of technologies and tools are being created to assist enterprises in gaining insights. The most relevant information on big data analytics tools, methods, and supporting technology is thus included in this article. We also highlight the capabilities of a number of tools and methods that may be used to explain big data from a cloud computing viewpoint, with new advancements that provide a deeper comprehension to address real-world problems.

(Goundar et al., 2021) [20] Big data is covered in this study, together with the available techniques for analyzing it and the need of using it, along with its advantages and disadvantages. In order to ascertain which organisations and companies in a small island state employ analytical tools to manage big data and what benefits these tools provide, this investigation implements survey questionnaires, interviews, business process observation, and secondary research methods. The reasons why these organizations are reluctant

to employ these technologies were also described in a study of businesses and organizations that do not use them.

(Nazia Tazeen & Rani, 2021) [21] Fast data transfer between devices is generating a lot of data on several industries, including commerce, entertainment, research, healthcare, media, and engineering. A major challenge is the ability to acquire and store data. A repository of open-source software, Apache Hadoop is used to store large amounts of data, conduct analytics, and carry out other big data-related tasks. Big Data must be appropriately categorized and evaluated as many businesses rely on information extracted from vast and complicated data to inform their choices. Numerous Big Data technologies and tools have been developed to resolve the intricate challenges that Big Data encounters. This article offers a fundamental introduction to the technology, tools, and applications of big data.

(Gujar & Ajmera, 2020) [22] Big data analytics will become more important in daily life, both personally and professionally, as big data begins to develop and spread. Massive quantities of data, equivalent to terabytes, are generated daily by contemporary digital technologies and information systems, including "cloud computing and the Internet of Things". It takes a lot of work on many levels to analyze this vast amount of data in order to extract information for decision-making. As a result, research and development in big data analysis is ongoing. This paper's main goal is to examine the possible effects of big data technologies and the difficulties they present. This article offers a framework for investigating large data at various phases. It also gives academics a fresh perspective on how to tackle the problems and unresolved research questions.

(Raheem & Arachchige, 2020) [23] Big data has been very successful in a variety of application sectors, such as social media, economics, finance, healthcare, and agriculture. This article outlines a few platforms and technologies that are used in the process of data analytics. It also discusses the many difficulties that “big data analytics” presents. A few critical issues, such as data representation, data administration, and data confidentiality, have been identified by big data analytics. The paper also aims to investigate the application of "big data analytic technologies" in the Internet of Things, cloud computing, and Hadoop. Lastly, the report summarizes the main areas where researchers are using big data analytics.

(Suresh et al., 2018) [24] Data analysis refers to the continuous observation, collection, organization, and analysis of information about a certain subject for both current and future development. Generally speaking, big data is a large collection of data that requires very sophisticated manipulation or analysis using conventional data mining or analytics ideas. Therefore, big data analytics requires the use of contemporary big data analytics technologies to tackle several difficult tasks, including data collecting, processing, searching, storing, sharing, privacy, and visualization. In order to address the aforementioned issues, this study examines contemporary big data analytics technologies.

(Srinivasan & Thirumalai Kumari, 2018) [25] The most popular word on the internet and in the world right now is "big data." The vast volumes of data that they have gathered and stored in enormous memory storage systems are being used by major companies. Big data like this cannot be handled by conventional data mining methods. Initially, Hadoop and MapReduce were the techniques used to manage large amounts of data. With the help of several plug-ins from both open source and for-profit big data analytics (BDA) companies, Hadoop and MapReduce can now do extensive "data mining and machine learning tasks". Additionally, large corporations have developed BDA tools and are currently developing more, the most of which are expensive.

CONCLUSION

With its ability to handle both organised and unstructured data, improve patient care, and facilitate real-time decision-making, big data analytics has become a disruptive force in a number of sectors, most notably healthcare. Its role extends to sectors such as banking, insurance, telecommunications, FMCG, fraud detection, and e-commerce. The integration of tools like Hadoop, MapReduce, and cloud computing platforms has enabled efficient data storage, management, and analysis at scale. Additionally, programming languages such as R, Python, and techniques like machine learning and text mining have advanced data-driven insights. The need of Big Data in facilitating prompt, well-informed decision-making was further highlighted by the COVID-19 pandemic. From tracking disease outbreaks to optimizing resource allocation, Big Data has proven essential in crisis management. As organizations continue to leverage analytics tools—covering data collection, cleaning, integration, and visualization—the demand for scalable, cost-effective solutions grows. With evolving technologies, Big Data Analytics will remain pivotal in uncovering patterns, predicting trends, and driving innovation across domains.

REFERENCES

[1] U. Sivarajah, S. Kumar, V. Kumar, S. Chatterjee, and J. Li, "A study on big data analytics and innovation: From technological and business cycle perspectives," Technol. Forecast. Soc. Change, vol. 202, no. February, p. 123328, 2024, doi: 10.1016/j.techfore.2024.123328.

[2] H. Zhang, S. Lee, Y. Lu, X. Yu, and H. Lu, "A Survey on Big Data Technologies and Their Applications to the Metaverse: Past, Current and Future," Mathematics, vol. 11, no. 1, 2023, doi: 10.3390/math11010096.

[3] A. Gupta, "A Review Paper on Big Data Analytics," Int. J. Innov. Res. Comput. Sci. Technol., no. 2, pp. 63–69, 2022, doi: 10.58532/v3bact5p3ch1.

[4] A. Islam Jony, "Applications of Real-Time Big Data Analytics," Int. J. Comput. Appl., vol. 144, no. 5, pp. 975–8887, 2016.

[5] I. A. Ajah and H. F. Nweke, "Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications," Big Data Cogn. Comput., vol. 3, no. 2, pp. 1–30, 2019.

[6] A. Mehta, K. Solanki, and K. Solanki, "Big Data Analytics: Current Research Trends, Applications, Prospects and Challenges," Spec. Issue - 2021 Int. J. Eng. Res. Technol., vol. 9, no. 5, pp. 187–192, 2021, [Online]. Available: https://trends.google.com/trends/explore?date=all&q=%2Fm%2F0bs2j8q

[7] A. Singh and N. Shanker, "Redefining Cybercrimes in light of Artificial Intelligence : Emerging threats and Challenges," pp. 192–201, 2024, doi: 10.69968/ijisem.2024v3si2192-201.

[8] H. P. Jones, "Big Data Analytics Tools: a Bibliometric Literature Review," 2016. [Online]. Available: http://thescholarship.ecu.edu/handle/10342/5615

[9] N. Singh, "A Comprehensive Study on Big Data Techniques, Tools and Challenges," Int. J. Cybern. Informatics, vol. 5, no. 4, pp. 71–79, 2016, doi: 10.5121/ijci.2016.5409.

[10] R. Yadav, "BIG DATA ANALYTICS TOOLS AND TECHNOLOGIES USED IN CLOUD ENVIRONMENT," nternational J. Creat. Res. Thoughts, vol. 10, no. 9, pp. 714–723, 2022.

[11] Batko and A. Ślęzak, "The use of Big Data Analytics in healthcare," J. Big Data, vol. 9, no. 1, 2022, doi: 10.1186/s40537-021-00553-4.

[12] U. KEKEVİ and A. A. AYDIN, "Real-Time Big Data Processing and Analytics: Concepts, Technologies, and Domains," Comput. Sci., no. 2, pp. 111–123, 2022, doi: 10.53070/bbd.1204112.

[13] R. Rawat and R. Yadav, "Big Data: Big data analysis, issues and challenges and technologies," IOP Conf. Ser. Mater. Sci. Eng., vol. 1022, no. 1, 2021, doi: 10.1088/1757-899X/1022/1/012014.

[14] H. B. Abdalla, "A brief survey on big data: technologies, terminologies and data-intensive applications," J. Big Data, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00659-3.

[15] S. Fakheri, "Big Data and Computing Visions A Comprehensive Review of Big Data Applications," Big Data Comput. Visions, vol. 2, no. 1, pp. 9–17, 2022, [Online]. Available: http://dx.doi.org/10.22105/bdcv.2022.325256.1041

[16] M. S. Kumar and J. Prabhu, "Recent development in big data analytics: Research perspective," Res. Anthol. Artif. Intell. Appl. Secur., vol. 3–4, no. November, pp. 1640–1663, 2020, doi: 10.4018/978-1-7998-7705-9.ch072.

[17] A. Badshah, A. Daud, R. Alharbey, A. Banjar, A. Bukhari, and B. Alshemaimri, Big data applications: overview, challenges and future, vol. 57, no. 11. Springer Netherlands, 2024. doi: 10.1007/s10462-024-10938-5.

[18] Z. A. Al-Sai et al., "Explore Big Data Analytics Applications and Opportunities: A Review," Big Data Cogn. Comput., vol. 6, no. 4, 2022, doi: 10.3390/bdcc6040157.

[19] S. Khan, H. U. Khan, and S. Nazir, "Systematic analysis of healthcare big data analytics for efficient care and disease diagnosing," Sci. Rep., vol. 12, no. 1, pp. 1–21, 2022, doi: 10.1038/s41598-022-26090-5.

[20] S. Goundar, A. Bhardwaj, S. Singh, M. Singh, and G. H L, "Big Data and Big Data Analytics: A Review of Tools and its Application," no. May, pp. 1–19, 2021, doi: 10.4018/978-1-7998-6673-2.ch001.

[21] Nazia Tazeen and K. S. Rani, "A Survey on Some Big Data Applications Tools and Technologies," Int. J. Recent Technol. Eng., vol. 9, no. 6, pp. 239–242, 2021, doi: 10.35940/ijrte.f5575.039621.

[22] C. R. Gujar and D. S. R. Ajmera, "A Study on Big Data Analytics: Technologies & Tools," Int. J. Comput. Appl. Technol. Res., vol. 9, no. 3, pp. 098–101, 2020, doi: 10.7753/ijcatr0903.1002.

[23] F. S. F. A. Raheem and I. U. G. Arachchige, "A Study on Big Data Analytics: Platforms and Tools, Challenges, Technologies and Key Applications," J. Inf. Syst. Inf. Technol., vol. 9, no. 3, pp. 098–101, 2020, doi: 10.7753/ijcatr0903.1002.

[24] A. Suresh, R. Kumar, E. Kannan, and Sagunthala, "A study on modern big data analytics tools," J. Adv. Res. Dyn. Control Syst., vol. 10, no. 1 Special Issue, pp. 282–285, 2018.

[25] S. Srinivasan and T. Thirumalai Kumari, "Big data analytics tools a review," Int. J. Eng. Technol., vol. 7, no. 3.3, p. 685, 2018, doi: 10.14419/ijet.v7i2.33.15476