iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: III Month of publication: March 2025

DOI: https://doi.org/10.22214/ijraset.2025.67766

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue III Mar 2025- Available at www.ijraset.com

Big Data Analytics

Mohammad Aaqib[1], Maharib Khan[2], Daksh Chandra[3]
[1]Dept of B.Tech, Parul University, Vadodara, India
[2]Dept of BCA, Makhanlal Chaturvedi University, Bhopal, India
[3]Dept of B.Tech, Parul University, Vadodara, India

Abstract*: Decision-makers are inundated with data in the current digital world. Big data presents difficulties for traditional management tools and techniques since it includes datasets that are not only large but also varied and changing quickly. Exploring and putting into practice ways for efficiently managing and extracting insights from large datasets is essential given the growing volume of such data. Decision-makers must extract valuable information from a range of data sources, including social media activity, everyday transactions, and customer contacts. Big data analytics is the process of applying sophisticated analytical methods to big datasets in order to extract insightful information. In addition to discussing the potential advantages of applying big data analytics to a variety of decision-making domains, this paper will examine many analytics methodologies and tools appropriate for big data analysis.
***Keywords*: big data, analytics, data mining, and decision-making.**

I. INTRODUCTION

Imagine a future in which data on individuals, organizations, transactions, and other recorded facts are all lost as soon as they are utilized. Under such circumstances, companies would be unable to obtain important information, carry out in-depth studies, and seize fresh chances. Everything is essential to day-to-day operations, including staff data, product specifications, and customer information. The cornerstone of any organization's growth and success is data.

Now think about the enormous quantity of material and knowledge that is currently accessible because of the internet and technological breakthroughs. Massive amounts of data are produced every second as a result of expanding storage capacity and different data-gathering techniques. It is now crucial to store and analyze this data in order to derive value. Additionally, as data storage has become less expensive, businesses are trying to get the most out of their enormous data repositories.

New approaches to big data analytics, storage, and analysis techniques are required due to the enormous volume, diversity, and quick evolution of this data. Extracting pertinent information from such large databases requires careful analysis. The purpose of this study is to examine the body of research on big data analytics, going over different tools, techniques, and technologies that are relevant to this field. It investigates their possible uses and advantages in several areas of decision-making. The studied literature covers the years 2008–2013, with a particular emphasis on big data talks from 2011–2013, which reflects the topic's recent prominence.

Research from respectable journals, conferences, and industry white papers are among the sources that were chosen. The majority of conversations on big data analytics, tools, techniques, and applications are found in conference papers and industry publications since academic journals have a drawn-out review procedure. Even while academic institutions contribute to the study of big data analytics, industry publications tend to cover the majority of developments and new technology.

II. LITERATURE REVIEW

"Big Data" describes datasets that are harder to handle with conventional database systems as they get bigger. These datasets are so large that they are difficult to handle, store, manage, and process in a reasonable period of time using standard software tools and storage systems.

The size of big data sets, which can range from a few dozen gigabytes to many petabytes, is constantly growing. The problems of working with big data include gathering, storing, finding, sharing, analyzing, and displaying the information. Businesses nowadays are searching through enormous volumes of precise data in an effort to find fresh possibilities and insights. Big data analytics is the process of applying sophisticated analytical methods to massive data sets in order to extract insightful information and spur business transformations. But it gets harder and harder to manage bigger data volumes.

We'll talk about the features and significance of big data in this part. Although it calls for new data structures, analytical techniques, and tools, analyzing bigger and more complicated data sets may have major commercial benefits.

We'll examine big data analytics tools and techniques, beginning with administration and storage and progressing to analytical processing. Lastly, we'll go over a few big data analytics that have grown in popularity as big data has grown.

III. CHARACTERISTICS OF BIG DATA

The main goal of big data is to manage enormous volumes of data that are too large, too diverse, and too dynamic for conventional systems to manage well. Volume, diversity, and velocity are the three primary components that define big data.

Volume is just the amount of data that we are discussing here. The quantity of records, transactions, tables, or files is equally as important as size. Big data is gathered from a variety of sources, including social media and website logs. Not only is it well-structured, but it also contains a wide variety of data forms, including text, audio, and video.

Variety refers to the many forms and kinds of data. We're discussing anything from tweets to sensor readings to photos; it's not just numbers in spreadsheets. An additional layer of complexity is introduced by the combination of semi-structured, unstructured, and structured data. The rate at which data is created and processed is known as velocity. Data is arriving more quickly than ever before due to the growth of real-time data streams from sources like social media and Internet of Things devices. To stay up, this calls for systems that can process and evaluate data instantly.

A fourth V that some people discuss is truthfulness. It all comes down to the data's quality. Is it true? Is it finished? Is it trustworthy? Making sure the data is reliable is essential since large data sometimes has a lot of noise mixed in with the signal. The requirement for new tools and technology to manage large data is another significant factor. Conventional databases and analysis techniques are no longer sufficient. To make sense of it all, we require tools like sophisticated analytics, machine learning, and distributed computing.

In summary, big data is about addressing the issues that arise from having a lot of data, not merely about having a lot of data. It involves figuring out how to glean insightful information from a vast amount of data and applying that information to influence innovation and commercial choices.

A. Big Data Analytics Tools and Methods

The need for faster and more efficient methods of data analysis is rising as technology develops and businesses handle ever-increasing amounts of data on a daily basis. It's no longer enough to just have a lot of data; you also need to analyze it well and use the insights you obtain to make choices quickly. These enormous datasets can no longer be handled using traditional data administration and analysis methods. For this reason, specific tools and techniques designed for big data analytics are required, as are the infrastructures required for organizing and storing such data. Big data's growth affects all stages of the data lifecycle, including gathering, processing, and decision-making. The Big Data, Analytics, and Decisions (B-DAD) framework was put out in order to solve these issues. Big data analytics techniques and technologies are included into the decision-making process using this approach. The B-DAD architecture matches key phases of the decision-making process with different big data management, processing, analytics, visualization, and assessment technologies. This guarantees that businesses may successfully use big data to guide their choices. Three key areas demonstrate the changes brought about by big data analytics:

Big data architecture and storage: Businesses require reliable architectures and storage solutions that can effectively manage enormous volumes of data. This covers scalable infrastructure and distributed storage options.

Data and analytics processing: To glean valuable insights from massive datasets, sophisticated data processing methods are needed. This calls for methods like distributed computing, parallel processing, and real-time analytics.

Analyzing big data to find information and make decisions: To find important insights and aid in decision-making, big data analytics approaches, including sentiment analysis, predictive modeling, and machine learning, are employed.

Although the main sectors impacted by big data analytics are summarized in this section, it's crucial to remember that the field is always changing. There are always new discoveries, instruments, and technologies coming along that provide businesses more chances to use big data.

B. Big Data Storage

Big data storage entails managing and classifying enormous volumes of data coming from digital platforms, social media, and Internet of Things devices. Big data administration is difficult because of its enormous volume, wide range of formats, and quick rate of development. The phrase "big data" describes data that is larger, more complicated, and more diverse than what can be handled by conventional storage and administration methods.

A flexible and scalable infrastructure is necessary to handle the high velocity, volume, and diversity of big data in order to store it efficiently. Big data may be stored on a variety of platforms and technologies. HDFS, a distributed file system that can store and manage massive data volumes across several cluster nodes, is one example of such a technology. HDFS can grow with the volume of data and is particularly made to manage big data workloads. To provide high-performance data access across large Hadoop clusters, it makes use of a NameNode and DataNode architecture. NoSQL databases, which are non-relational databases skilled in handling unstructured and semi-structured data, provide an additional choice. Big data workloads may be supported by NoSQL databases, which also provide scalability and flexibility. They offer specialized search databases for analytics on semi-structured data and accommodate a variety of data access patterns, including low-latency applications. Furthermore, NoSQL databases include a range of data structures that are geared for scalability and performance, including document, graph, and key-value.

C. Big Data Processing

Analytic processing comes next after massive data storage. Four essential prerequisites must be met in order to handle large amounts of data. First and foremost, quick data loading is necessary to minimize loading times impacted by disk and network traffic while queries are being executed. Second, in order to satisfy the needs of high workloads and real-time queries, quick query processing is required. As query volumes rise, the data placement structure should continue to handle queries at fast rates. Thirdly, the quick increase in user activity necessitates scalable storage capacity, making effective use of storage space essential. In order to optimize space use, limited disk space requires carefully controlled data storage during processing. Finally, since different applications and users evaluate massive data sets for diverse objectives, it is imperative to have great adaptability to changing workload patterns. The system must be extremely flexible to adjust to unforeseen modifications in data processing.

The foundation of Hadoop is MapReduce, a parallel programming paradigm derived from functional languages that works well for processing large amounts of data. To cut down on job completion time, it divides work into phases that are carried out concurrently. The MapReduce algorithm divides the incoming values into smaller jobs, assigns them to the relevant pairings, and maps them to key/value pairs. The "Reduce" function uses the output as input, gathering and combining values that share a key to get the end result.

The Job Tracker and Task Tracker nodes are the two nodes that Hadoop uses for its MapReduce operation. The Job Tracker keeps track of outcomes while allocating mapper and reducer functions to available Task Trackers. In order to minimize inter-node communication via HDFS files and directories, Task Tracker nodes carry out jobs and report findings to the Job Tracker.

Large datasets are stored in HDFS over many Data Nodes, as seen in Figures 1 and 2. Tasks are assigned to mappers for data processing when a MapReduce job begins on a Tracker. Results are finally combined using reducers. Data is arranged into distributed files by Hadoop so that MapReduce may examine it while processing. Because it can adapt to different data sources, this system is preferred for big data analytics. By using algorithms to identify patterns, correlations, and information in recorded data, decision-makers may employ analytics to get insights that have a big influence on corporate operations.

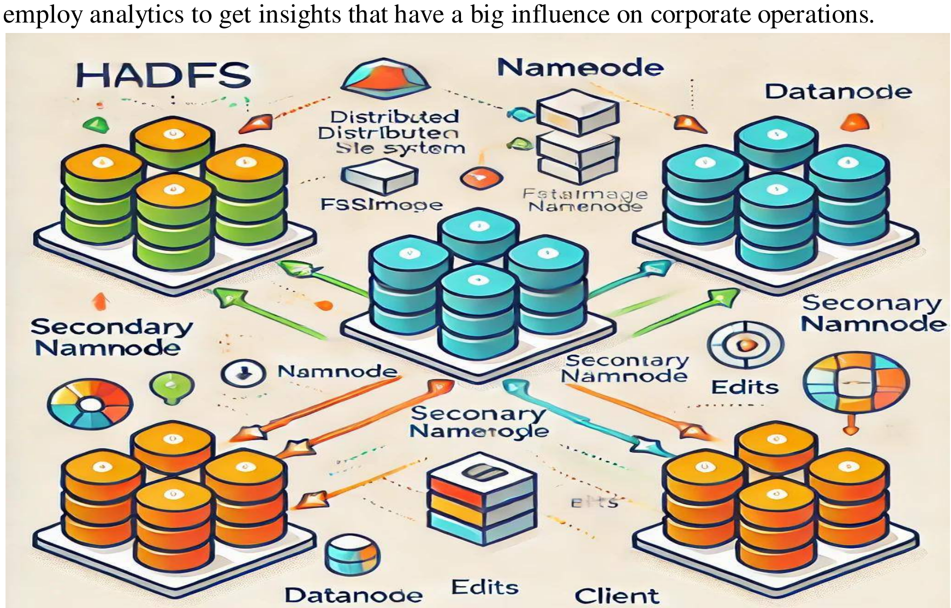


Fig. 1

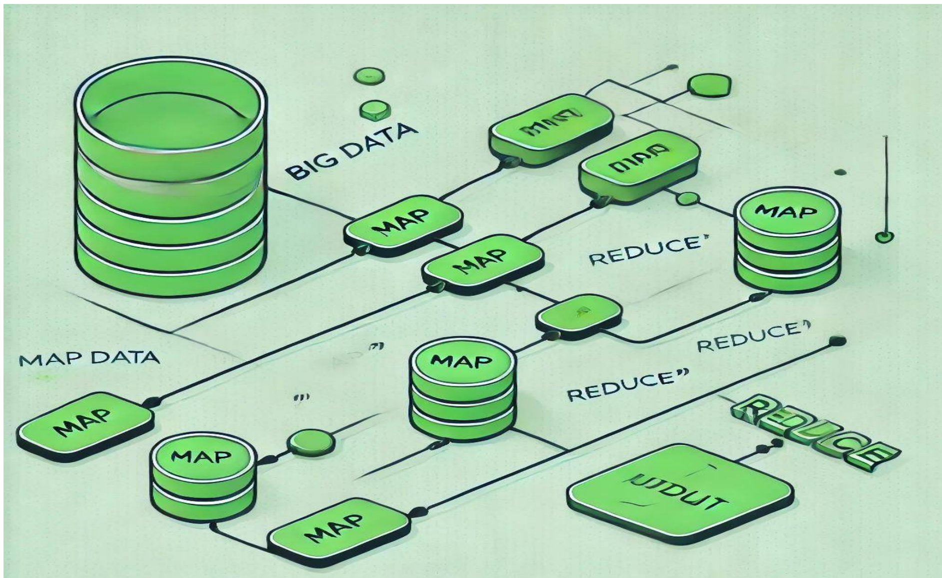


Fig. 2

D. Big Data Analytics and Decision Making

Big data is extremely important from the perspective of decision-makers as it provides useful information and knowledge that can be used to guide choices. Numerous studies have examined the management decision-making process throughout time, highlighting its critical function. With its ability to provide enormous amounts of complex data from a variety of sources, including scanners, mobile devices, loyalty programs, the internet, and social media platforms, big data is steadily becoming a crucial tool for decision-makers.

Extensive analysis is necessary to find important insights in order to derive significant advantages from this abundance of data. Decision-makers may then use historical and current data from a variety of activities, including supply chains and consumer behavior, to take advantage of these insights. Although businesses are used to closely examining internal data, such as sales and inventories, there is an increasing need to examine external data sources, such as supply chain dynamics and consumer demographics. Big data offers the chance to glean information and cumulative value from the data.

Frameworks such as the B-DAD framework have been developed to address the problems that big data presents. This paradigm improves the quality of big data decision-making by incorporating big data tools and approaches into the process. The intelligence phase, which gathers information from both internal and external sources to identify issues and opportunities, is where the decision-making process starts. After then, a variety of big data management and storage techniques are used to process, store, and arrange this data.

Using model planning, data analytics, and analysis, possible strategies are developed and examined during the design process. While the implementation phase involves putting the selected solution into practice, the choosing phase evaluates the implications of suggested solutions from the design phase. Organizations from a variety of industries are becoming more and more focused on handling and evaluating the ever-increasing amount of big data. By analyzing massive databases to find trends, attitudes, and consumer insights, they are using big data analytics to unlock economic value and make quicker, better-informed choices.

IV. CUSTOMER INTELLIGENCE

Big data analytics may greatly help sectors like retail, finance, and telecoms and has a lot of promise for consumer intelligence. Big data may increase transparency and facilitate stakeholders' quick and easy access to pertinent data. Businesses may improve customer satisfaction and retention rates by using big data analytics to profile and segment their clientele according to various socioeconomic factors. In addition to identifying sales and marketing possibilities, this may help businesses make better-informed marketing decisions and target different market segments according to their preferences. Additionally, businesses may utilize social media to learn about the preferences and dislikes of their customers. Businesses may take appropriate action by doing sentiment analysis on this data to identify when customers are switching to other goods or turning against them.

Additionally, businesses may respond to trends and carry out direct marketing by employing SNAs to track consumer attitudes about brands and pinpoint important people. The development of prediction models for consumer behaviour and purchasing trends can also be made possible by big data analytics, increasing total profitability. In order to target promotions and advertising, even companies that have long employed segmentation are starting to embrace more advanced big data approaches, such as real-time micro-segmentation of customers. Big data analytics may therefore help businesses by facilitating more focused social influencer marketing, identifying and forecasting trends based on market sentiment, and examining and comprehending consumer behaviour such as churn.

A. Supply Chain and Performance Management

Big data analytics has the ability to completely transform supply chain management, which would be advantageous for sectors including logistics, retail, manufacturing, and transportation. Businesses may optimize operations and cut costs by using big data to foresee changes in demand and adjust their supply appropriately. Organizations may automate replenishment choices by analyzing variables including delivery patterns and stock use, which reduces lead times, costs, and interruptions.

Big data also makes it easier to choose suppliers intelligently by taking competitive pricing and quality into account. Businesses may increase profit margins and manage inventory more effectively with the help of instantaneous analysis of alternative pricing situations. Better planning and forecasting techniques are promoted by big data's assistance in identifying underlying cost drivers.

Big data analytics has several benefits for performance management, especially for the public and private sectors looking to increase productivity. By enabling the tracking and forecasting of employee performance, predictive analytics solutions help to improve overall efficiency by coordinating strategic goals with service results.

Additionally, operations managers are empowered with insightful information by the availability of big data and performance metrics, which makes it easier to deploy dashboards, balanced scorecards, and predictive key performance indicators (KPIs). These tools improve an organization's general planning and management procedures, goal-setting, transparency, and performance tracking.

B. Risk Management and Fraud Detection

Businesses like retail banking, insurance, and investing stand to benefit greatly from using big data analytics, especially when it comes to risk management. By weighing possible advantages against prospective losses, big data analytics can be extremely helpful in the financial industry, where risk assessment and management are critical. Additionally, organizations can obtain a thorough and dynamic picture of their risk exposure by closely examining both internal and external sources of big data, which will improve risk quantification. By combining heterogeneous risk profiles from different departments into a unified enterprise-wide view, high-performance analytics can further streamline risk management initiatives. Decision-makers can more successfully identify and reduce risks by using this holistic approach, which acknowledges the connections between various risk categories. Additionally, with improved scalability and data-capturing capabilities, big data technologies provide solutions to address database performance issues and deal with the exponential expansion of data generated by networks. To strengthen defences against cyber and network attacks, this involves enhancing cyber analytics and data-intensive computing solutions to take advantage of numerous data streams and automated analyses. Big data analytics offers a potent tool for detecting and stopping fraudulent activity in the field of fraud detection, which is especially common in industries like government, banking, and insurance. Organizations can speed up and improve the accuracy of fraud analytics by utilizing big data's capacity to match electronic data from many sources. Big data analytics-derived consumer intelligence can help simulate normal customer behaviour, enabling the quick identification of unusual or suspicious activity. Furthermore, giving fraud detection systems knowledge about new fraud trends enables them to adjust and successfully counteract changing fraudster threats. By identifying evidence of fraudulent claims and revealing networks of fraudsters working together, social network analysis (SNA) approaches can reduce the likelihood of undiscovered fraudulent activity. Overall, by quickly recognizing and reacting to compliance patterns across many datasets, enterprises can greatly improve their capacity to stop and recover from fraudulent transactions by utilizing big data technologies, methodologies, and governance frameworks.

V. CONCLUSION

Big data is extremely significant from the perspective of decision-makers since it provides important information and knowledge that is necessary for making decisions. Over the years, a great deal of study has examined the managerial decision-making process, highlighting its importance.

Big data, which provides enormous volumes of precise information from several sources like scanners, smart phones, loyalty cards, the internet, and social media platforms, is becoming an increasingly important tool for decision-makers.

In order to fully utilize this abundance of data, careful analysis is essential to revealing insightful information. Then, using both historical and current data produced by operations like supply chains and consumer behavior, decision-makers can profit from these insights. Although businesses are used to examining internal data, such sales and inventories, there is an increasing need to examine external data, such as supplier chains and consumer markets. Big data offers the chance to gather information and cumulative value from a wide range of data sources. Big data presents issues that have been addressed by the development of frameworks such as the B-DAD framework. This paradigm improves the quality of big data decision-making by incorporating big data tools and approaches into the process. The intelligence stage of the decision-making process is where information is collected to find issues and opportunities from both internal and external sources. This data is then processed, stored, and arranged utilizing a variety of big data management and storage techniques. Using model planning, data analytics, and analysis, possible strategies are created and examined during the design phase. While the implementation phase entails putting the selected solution into practice, the choice phase assesses the effects of suggested solutions from the design phase.

Organizations in all industries are becoming more interested in managing and interpreting big data as its volume continues to increase at an exponential rate. By examining vast databases to find trends, attitudes, and consumer insights, they are using big data analytics to generate economic value and make better, quicker choices.

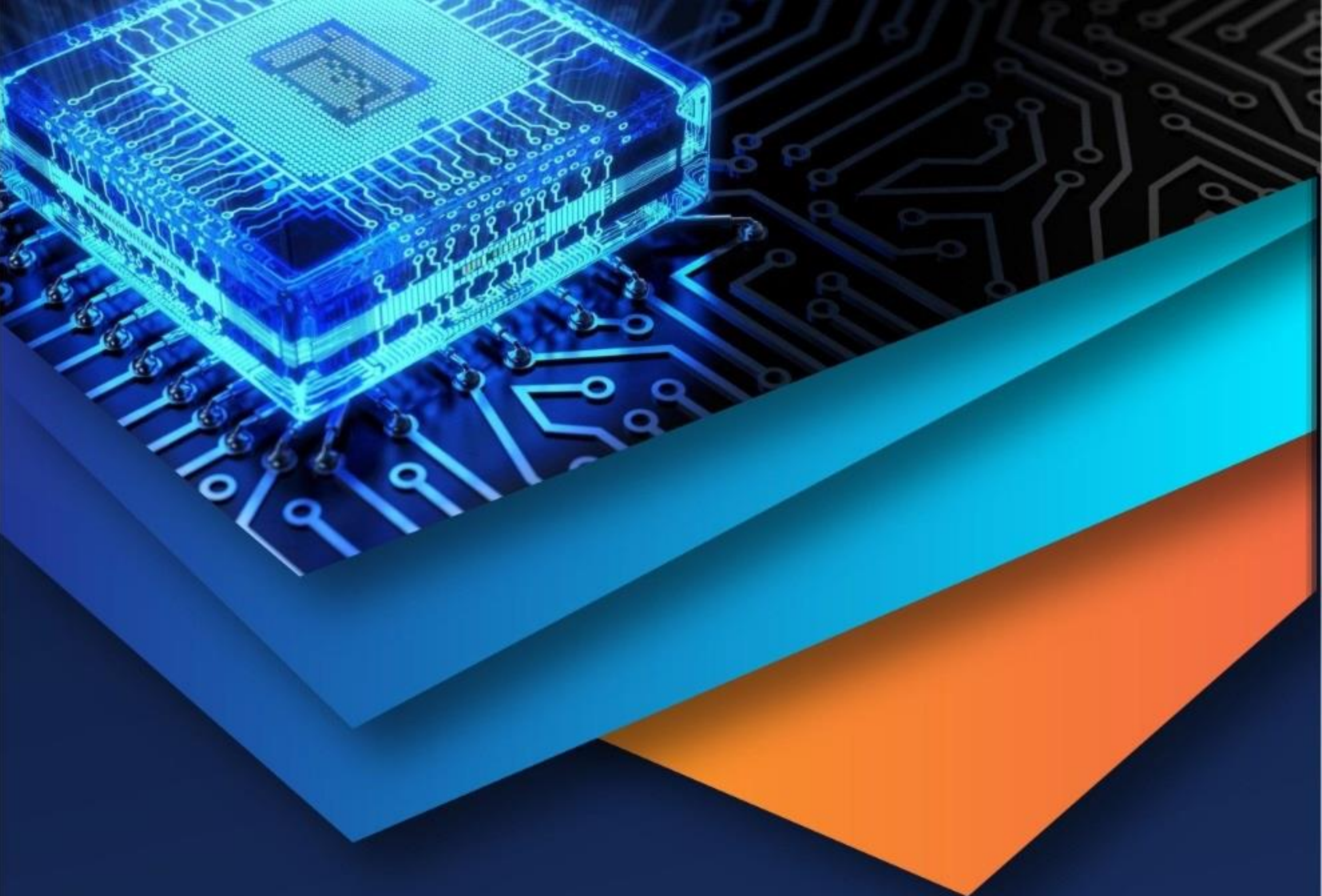
VI. ACKNOWLEGMENT

We would like to express our heartfelt gratitude to our respective institutions, Parul University and Makhanlal Chaturvedi University, for providing us with the necessary resources and support to conduct this research. We extend our sincere appreciation to our professors and mentors for their invaluable guidance and encouragement throughout this study.

Additionally, we would like to acknowledge our colleagues and peers for their insightful discussions and constructive feedback, which have significantly contributed to the improvement of this work. Finally, we are grateful to our families and friends for their continuous support and motivation during the research process.

REFERENCES

[1] Adams, M.N.: Perspectives on Data Mining. International Journal of Market Research 52(1), 11–19 (2010)
[2] Asur, S., Huberman, B.A.: Predicting the Future with Social Media. In: ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 1, pp. 492–499 (2010)
[3] Bakshi, K.: Considerations for Big Data: Architecture and Approaches. In: Proceedings of the IEEE Aerospace Conference, pp. 1–7 (2012)
[4] Cebr: Data equity, Unlocking the value of big data. in: SAS Reports, pp. 1–44 (2012)
[5] Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J.M., Welton, C.: MAD Skills: New Analy-sis Practices for Big Data. Proceedings of the ACM VLDB
[6] Endowment 2(2), 1481–1492 (2009)
[7] Cuzzocrea, A., Song, I., Davis, K.C.: Analytics over Large-Scale Multidimensional Data: The Big Data Revolution! In: Proceedings of the ACM International Workshop on Data Warehousing and OLAP, pp. 101–104 (2011)
[8] Economist Intelligence Unit: The Deciding Factor: Big Data & Decision Making. In: Capgemini Reports, pp. 1–24 (2012)
[9] Elgendy, N.: Big Data Analytics in Support of the Decision Making Process. MSc Thesis, German University in Cairo, p. 164 (2013)
[10] EMC: Data Science and Big Data Analytics. In: EMC Education Services, pp. 1–508 (2012)
[11] He, Y., Lee, R., Huai, Y., Shao, Z., Jain, N., Zhang, X., Xu, Z.: RCFile: A Fast and Space-efficient Data Placement Structure in MapReduce-based Warehouse Systems. In: IEEE International Conference on Data Engineering (ICDE), pp. 1199–1208 (2011)



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY