iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.59806

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

Big Data Analytics

Sumit Suresh Jadhav[1], Mrs. Sujata Patil[2]
[1]*Dept of MCA, Trinity Academy of Engineering, Pune, India*
[2]*Assistant Professor, Trinity Academy of Engineering, Pune, India*

Abstract: ***In today's digital era, decision-makers are presented with an abundance of data. Big data encompasses datasets that are not only extensive but also diverse and rapidly evolving, posing challenges for conventional management tools and methods. With the escalating volume of such data, it's crucial to explore and implement solutions for effectively handling and deriving insights from these datasets. Decision-makers need to glean meaningful insights from a variety of data sources, including daily transactions, consumer interactions, and social media activity . Big data analytics involves employing advanced analytical techniques on large datasets to unlock valuable insights. This paper will explore various analytics approaches and tools suitable for big data analysis, as well as the potential benefits offered by leveraging big data analytics across different decision-making domains.***

Keywords: big data, data mining, analytics, decision making.

I. INTRODUCTION

Imagine a world where information disappears as soon as it's used, with no way to store data about people, organizations, transactions, or any documented details. In such a scenario, organizations would lose the ability to gather valuable insights, perform thorough analyses, and capitalize on new opportunities. Everything from customer information to product details to employee records plays a vital role in daily operations. Data serves as the foundation upon which organizations build and thrive.

Now, consider the vast amount of data and information available today, thanks to technological advancements and the internet. With increased storage capacity and various data collection methods, enormous volumes of data are generated every second. Storing and analyzing this data to extract value has become essential. Additionally, the cost of storing data has decreased, prompting organizations to maximize the value derived from their vast data stores.

The sheer size, diversity, and rapid evolution of this data necessitate new approaches to big data analytics, storage, and analysis methods. Proper analysis of such massive datasets is crucial for extracting relevant information.

This paper aims to analyze existing literature on big data analytics, discussing various tools, methods, and technologies applicable to big data analysis. It explores their potential applications and opportunities across different decision-making domains.

The literature reviewed spans from 2008 to 2013, with a focus on big data discussions from 2011 to 2013, reflecting the recent prominence of the topic.

The selected sources include research from reputable journals, conferences, and industry white papers. Due to the lengthy review process of academic journals, most discussions about big data analytics, tools, methods, and applications are found in conference papers and industry publications. While academia contributes to the research on big data analytics, many advancements and new technologies are primarily discussed in industry papers.

II. LITERATURE REVIEW

The term "Big Data" refers to datasets that become difficult to manage with traditional database systems as they grow larger. These datasets are so massive that commonly used software tools and storage systems struggle to handle, store, manage, and process them within a reasonable amount of time.

Big data sets are continually expanding, ranging from a few dozen terabytes to many petabytes in size. Dealing with big data presents challenges such as capturing, storing, searching, sharing, analyzing, and visualizing the data. Nowadays, businesses are delving into vast amounts of detailed data to uncover new insights and opportunities.

Big data analytics involves using advanced analytical techniques on large data sets to uncover valuable insights and drive business changes. However, managing larger data sets becomes increasingly challenging..

In this section, we'll discuss the characteristics and importance of big data. Analyzing larger and more complex data sets can provide significant business benefits, but it requires new data architectures, analytical methods, and tools. We'll explore big data analytics tools and methods, starting with storage and management, then moving on to analytic processing. Finally, we'll discuss various big data analyses that have become more prevalent with the rise of big data.

©IJRASET: All Rights are Reserved

III. CHARACTERISTICS OF BIG DATA

Big data is all about handling massive amounts of data that are too big, too varied, and too fast-moving for traditional systems to handle effectively. There are three main things that make big data what it is: volume, variety, and velocity.

Volume refers to just how much data we're talking about here. It's not just about size, but also about the number of records, transactions, tables, or files. Big data comes from a wide range of sources like website logs, social media, and more. It's not just neatly organized stuff either; there's all kinds of data formats, from text to video to audio, making it incredibly diverse.

Variety is about the different types and formats of data. It's not just numbers in spreadsheets; we're talking about everything from tweets to sensor readings to images. This mix of structured, unstructured, and semi-structured data adds another layer of complexity.

Velocity is about the speed at which data is being generated and processed. With the rise of real-time data streams from sources like social media and IoT devices, data is coming in faster than ever before. This requires systems that can handle and analyze data in real-time to keep up with the pace.

Some folks also talk about a fourth V: veracity. This is all about the quality of the data. Is it accurate? Is it complete? Is it reliable? With big data, there's often a lot of noise mixed in with the signal, so making sure the data is trustworthy is crucial.

Another important aspect of big data is the need for new technologies and tools to handle it all. Traditional databases and analysis methods just can't cut it anymore. We need things like distributed computing, machine learning, and advanced analytics to make sense of it all.

In conclusion, big data is not just about having lots of data; it's about dealing with the challenges that come with it. It's about finding ways to extract valuable insights from a sea of information and using that knowledge to drive business decisions and innovation.

A. Big Data Analytics Tools and Methods

As technology advances and organizations deal with increasingly large volumes of data every day, there's a growing demand for quicker and more effective ways to analyze this data. Simply having lots of data isn't enough anymore; it's crucial to analyze it efficiently and make timely decisions based on the insights gained. Traditional data management and analysis techniques are no longer sufficient for handling these massive datasets. That's why there's a need for specialized tools and methods tailored for big data analytics, along with the necessary infrastructures for storing and managing such data. The rise of big data impacts every aspect of the data lifecycle, from collection to processing to decision-making. To address these challenges, the Big Data, Analytics, and Decisions (B-DAD) framework has been proposed. This framework integrates big data analytics tools and methods into the decision-making process. The B-DAD framework aligns various tools for big data storage, management, processing, analytics, visualization, and evaluation with different stages of the decision-making process. This ensures that organizations can leverage big data effectively to make informed decisions. The changes brought about by big data analytics are evident in three main areas:

Big data storage and architecture: Organizations need robust storage systems and architectures capable of handling vast amounts of data efficiently. This includes distributed storage solutions and scalable infrastructure.

Data and analytics processing: Advanced data processing techniques are required to extract meaningful insights from large datasets. This involves techniques like parallel processing, distributed computing, and real-time analytics.

Big data analyses for knowledge discovery and decision-making: Big data analytics techniques such as machine learning, predictive modeling, and sentiment analysis are used to uncover valuable insights and support decision-making processes.

While this section provides an overview of the key areas affected by big data analytics, it's important to note that the field is continually evolving. New findings, tools, and technologies are constantly emerging, offering new opportunities for organizations to harness the power of big data.

B. Big Data Storage

Big data storage involves handling and organizing vast amounts of data originating from sources like social media, IoT devices, and digital platforms. The management of big data poses challenges due to its massive volume, varied formats, and rapid generation speed. The term "big data" refers to data that surpasses the capacity of traditional storage and management systems in terms of size, complexity, and diversity. To effectively store big data, a flexible and scalable infrastructure is required to accommodate the data's high velocity, volume, and variety. Various technologies and platforms are available for big data storage. One such technology is HDFS, a distributed file system capable of storing and managing large data volumes across multiple nodes in a cluster. HDFS is specifically designed for handling big data workloads and can scale according to data volume. It utilizes a NameNode and DataNode architecture to enable high-performance data access across extensive Hadoop clusters. Another option is NoSQL databases, which are non-relational databases adept at managing unstructured and semi-structured data. NoSQL databases offer scalability, flexibility, and support for big data workloads. They cater to diverse data access patterns, including low-latency applications, and provide specialized search databases for analytics on semi-structured data. Additionally, NoSQL databases offer various data models like key-value, document, and graph, optimized for performance and scalability.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

C. Big Data Processing

Following big data storage, the next step is analytic processing. There are four crucial requirements for processing big data. Firstly, fast data loading is essential to reduce loading times affected by disk and network traffic during query executions. Secondly, fast query processing is necessary to meet the demands of heavy workloads and real-time requests. The data placement structure should maintain high query processing speeds as query volumes increase. Thirdly, efficient utilization of storage space is vital due to the rapid growth in user activities requiring scalable storage capacity. Limited disk space necessitates well-managed data storage during processing to maximize space utilization. Finally, strong adaptivity to dynamic workload patterns is crucial as big data sets are analyzed by various applications and users for different purposes. The system should be highly adaptive to unexpected changes in data processing.

MapReduce, a parallel programming model inspired by functional languages, is suitable for big data processing and forms the core of Hadoop. It breaks tasks into stages executed in parallel to reduce task completion time. In the MapReduce process, input values are mapped to key/value pairs, partitioned into smaller tasks, and assigned to appropriate pairs. The output serves as input to the "Reduce" function, which collects and combines values sharing the same key to provide the final result.

Hadoop's MapReduce function relies on two nodes: the Job Tracker and Task Tracker nodes. The Job Tracker distributes mapper and reducer functions to available Task Trackers and monitors results. Task Tracker nodes execute jobs and communicate results to the Job Tracker, minimizing inter-node communication through HDFS files and directories.

Figures 1 and 2 depict the storage of large datasets in HDFS, where data is stored across multiple Data Nodes. When a MapReduce job starts on a Tracker, tasks are distributed to mappers for data processing. Finally, reducers combine results. Hadoop organizes data into distributed files, allowing MapReduce to analyze it during processing. This system is favored for big data analytics due to its adaptability to various data sources. Decision-makers use analytics to extract insights from stored data by applying algorithms to find patterns, relationships, and information, significantly impacting business operations.

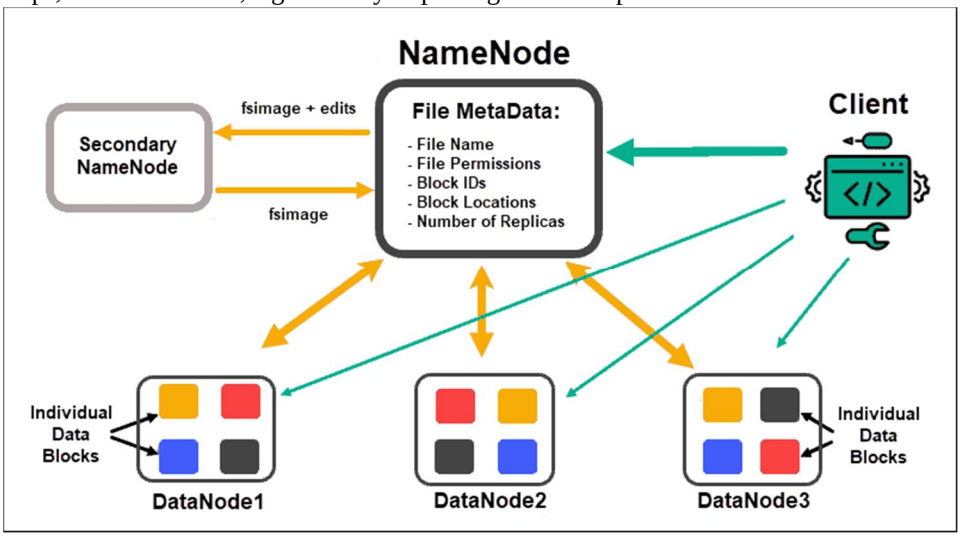


Fig.1

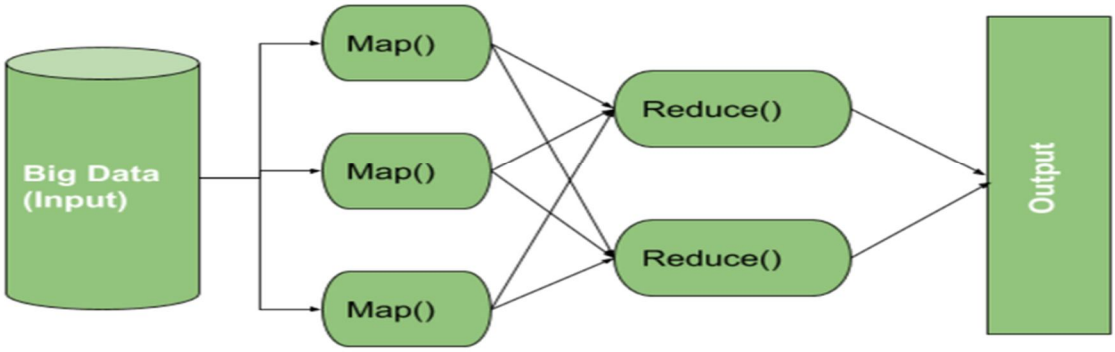


Fig.2

D. Big Data Analytics and Decision Making

From the viewpoint of decision-makers, big data holds immense significance as it furnishes valuable information and knowledge upon which decisions can be founded. Over time, extensive research has delved into the managerial decision-making process, underscoring its crucial role. Big data is progressively emerging as a pivotal asset for decision-makers, offering vast volumes of intricate data from diverse origins like scanners, mobile devices, loyalty programs, the internet, and social media platforms.

To extract substantial benefits from this wealth of data, thorough analysis is imperative to uncover valuable insights. Decision-makers can then capitalize on these insights, leveraging both historical and real-time data stemming from various processes such as supply chains and customer behaviors. While organizations are accustomed to scrutinizing internal data like sales and inventory, there's a mounting necessity to analyze external data sources such as customer demographics and supply chain dynamics. Big data presents an opportunity to extract cumulative value and intelligence from such data.

To tackle the challenges posed by big data, frameworks like the B-DAD framework have been devised. This framework integrates big data tools and techniques into the decision-making process, augmenting the quality of decision-making regarding big data. The decision-making process commences with the intelligence phase, where data is gathered from internal and external sources to pinpoint problems and opportunities. Subsequently, this data is processed, stored, and organized using various big data storage and management tools.

In the design phase, potential courses of action are formulated and scrutinized through model planning, data analytics, and analysis. The choice phase assesses the ramifications of proposed solutions from the design phase, while the implementation phase entails putting the chosen solution into action. With the ever-growing volume of big data, organizations spanning diverse sectors are increasingly focused on managing and analyzing such data. They are embracing big data analytics to unlock economic value and make more informed, prompt decisions by dissecting large datasets to unveil patterns, sentiments, and customer insights.

IV. CUSTOMER INTELLIGENCE

Big data analytics holds much potential for customer intelligence, and can highly benefit industries such as retail, banking, and telecommunications. Big data can create transparency, and make relevant data more easily accessible to stakeholders in a timely manner. Big data analytics can provide organizations with the ability to profile and segment customers based on different socioeconomic characteristics, as well as increase levels of customer satisfaction and retention. This can allow them to make more informed marketing decisions, and market to different segments based on their preferences along with the recognition of sales and marketing opportunities. Moreover, social media can be used to inform companies what their customers like, as well as what they don’t like. By performing sentiment analysis on this data, firms can be alerted beforehand when customers are turning against them or shifting to different products, and accordingly take action.

Additionally, using SNAs to monitor customer sentiments towards brands, and identify influential individuals, can help organizations react to trends and perform direct marketing. Big data analytics can also enable the construction of predictive models for customer behavior and purchase patterns, therefore raising overall profita-bility. Even organizations which have used segmentation for many years are beginning to deploy more sophisticated big data techniques, such as real-time micro-segmentation of customers, in order to target promotions and advertising. Consequently, big data analytics can benefit organizations by enabling better targeted social influencer marketing, defining and predicting trends from market sentiments, as well as analyzing and understanding churn and other customer behaviors.

A. Supply Chain and Performance Management

Big data analytics holds immense potential for revolutionizing supply chain management, benefiting industries like manufacturing, retail, transportation, and logistics. By leveraging big data to forecast shifts in demand, businesses can align their supply accordingly, optimizing operations and minimizing costs. Analyzing factors such as stock utilization and delivery patterns enables organizations to automate replenishment decisions, resulting in reduced lead times, cost savings, and fewer disruptions.

Moreover, big data facilitates informed decisions regarding supplier selection, considering factors like quality and price competitiveness. Instantaneous analysis of alternate pricing scenarios empowers businesses to manage inventories more efficiently and boost profit margins. Big data also aids in identifying underlying cost drivers, fostering better planning and forecasting practices.

In performance management, big data analytics offers significant advantages, particularly for governmental and healthcare sectors seeking productivity improvements. Predictive analytics tools enable the monitoring and forecasting of staff performance, aligning strategic objectives with service outcomes to enhance overall efficiency.

Furthermore, the availability of big data and performance metrics empowers operations managers with valuable insights, facilitating the implementation of predictive key performance indicators (KPIs), balanced scorecards, and dashboards. These tools enhance performance monitoring, transparency, objective setting, and overall planning and management processes within organizations..

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

B. Risk Management and Fraud Detection

Industries like investment, retail banking, and insurance stand to gain significant advantages from leveraging big data analytics, particularly in the realm of risk management. In the financial sector, where assessing and managing risk is paramount, big data analytics can play a crucial role in making informed investment decisions by analyzing potential gains against potential losses. Moreover, by scrutinizing both internal and external sources of big data, organizations can achieve a comprehensive and dynamic understanding of their risk exposure, facilitating better risk quantification. Implementing high-performance analytics can further streamline risk management efforts by integrating disparate risk profiles from various departments into a cohesive enterprise-wide view. This holistic approach enables decision-makers to identify and mitigate risks more effectively by recognizing the interrelationships between different risk types. Furthermore, advancements in big data technologies offer solutions to cope with the exponential growth of data generated by networks while also addressing database performance challenges through enhanced scalability and data capture capabilities. This includes bolstering cyber analytics and data-intensive computing solutions to leverage multiple data streams and automated analyses, thereby fortifying defenses against cyber and network attacks. In fraud detection, particularly prevalent in sectors such as government, banking, and insurance, big data analytics presents a powerful tool for identifying and preventing fraudulent activities. By harnessing big data's capabilities, organizations can streamline the process of matching electronic data across diverse sources, enabling quicker and more accurate fraud analytics. Customer intelligence derived from big data analytics can aid in modeling typical customer behavior, allowing for the swift detection of anomalous or suspicious activities. Additionally, providing fraud detection systems with insights into emerging fraud patterns empowers them to adapt and respond effectively to evolving threats posed by fraudsters. Social network analysis (SNA) techniques can be employed to uncover collaborative networks among fraudsters and identify evidence of fraudulent claims, thereby minimizing the occurrence of undetected fraudulent activities. Overall, by leveraging big data tools, techniques, and governance frameworks, organizations can significantly enhance their ability to prevent and recover from fraudulent transactions by rapidly identifying and responding to compliance patterns across diverse datasets.

V. CONCLUSION

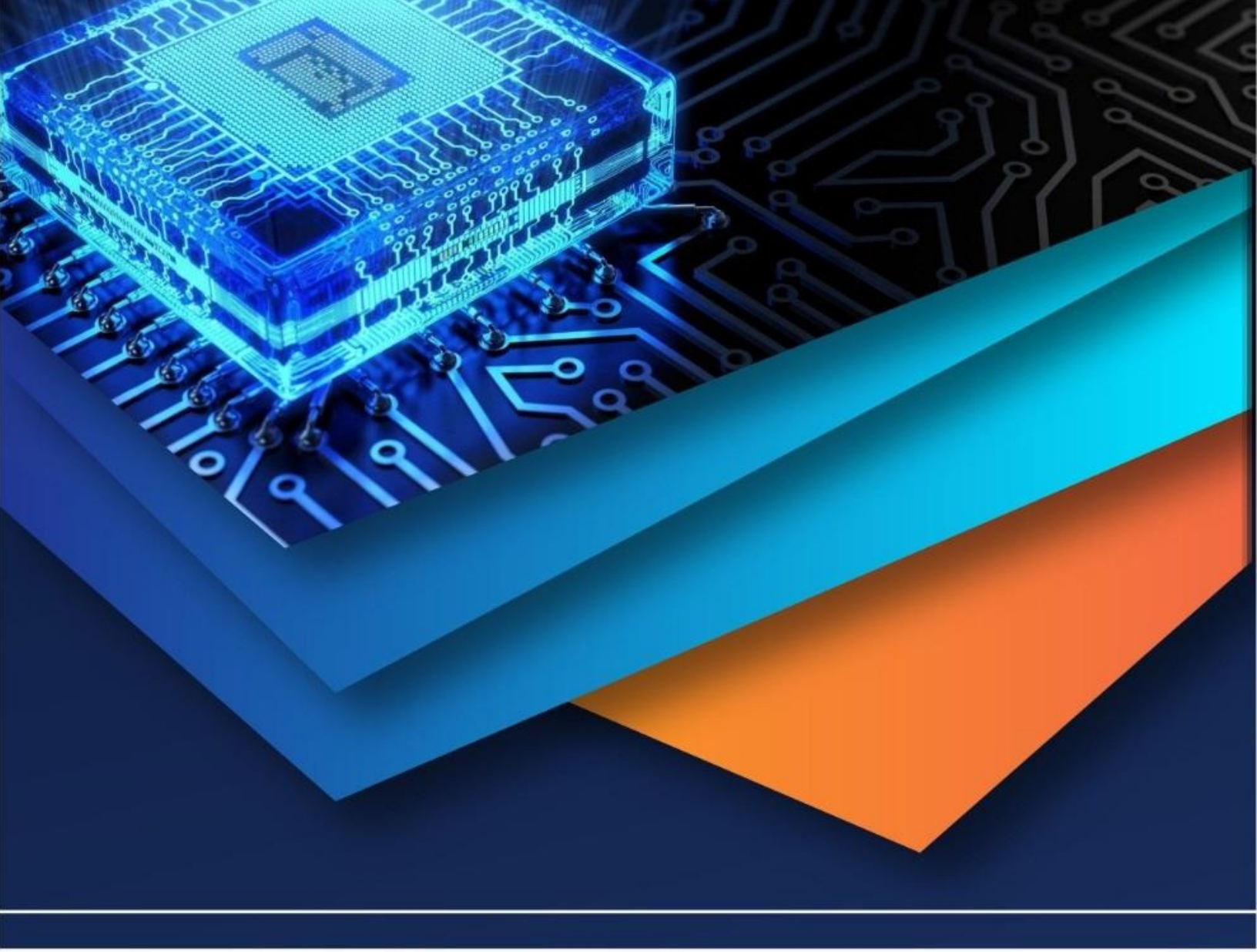
From the viewpoint of decision-makers, big data holds tremendous importance as it furnishes valuable information and knowledge essential for decision-making. Extensive research has explored the managerial decision-making process over the years, underscoring its significance. Big data is emerging as a vital resource for decision-makers, offering vast amounts of detailed data from various sources such as scanners, mobile phones, loyalty cards, the internet, and social media platforms.

To harness the full potential of this wealth of data, thorough analysis is crucial to uncover valuable insights. Decision-makers can then capitalize on these insights, utilizing both historical and real-time data generated through processes like supply chains and customer behaviors. While organizations are accustomed to analyzing internal data such as sales and inventory, there is a growing imperative to analyze external data like customer markets and supply chains. Big data presents an opportunity to extract cumulative value and knowledge from such diverse data sources. To overcome the challenges posed by big data, frameworks like the B-DAD framework have been devised. This framework integrates big data tools and techniques into the decision-making process, enhancing the quality of decision-making regarding big data. The decision-making process begins with the intelligence phase, where data is gathered from internal and external sources to identify problems and opportunities. Subsequently, this data undergoes processing, storage, and organization using various big data storage and management tools. In the design phase, potential courses of action are developed and analyzed through model planning, data analytics, and analysis. The choice phase evaluates the impacts of proposed solutions from the design phase, while the implementation phase involves putting the chosen solution into action.

As the volume of big data continues to grow exponentially, organizations across sectors are increasingly keen on managing and analyzing such data. They are embracing big data analytics to unlock economic value and make better, faster decisions by analyzing large datasets to uncover patterns, sentiments, and customer insights.

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 Endowment 2(2), 1481–1492 (2009)
[6] 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)
[7] Economist Intelligence Unit: The Deciding Factor: Big Data & Decision Making. In: Capgemini Reports, pp. 1–24 (2012)
[8] Elgendy, N.: Big Data Analytics in Support of the Decision Making Process. MSc Thesis, German University in Cairo, p. 164 (2013)
[9] EMC: Data Science and Big Data Analytics. In: EMC Education Services, pp. 1–508 (2012)
[10] 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

Call : 08813907089 (24*7 Support on Whatsapp)