



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Data Analytics: Big Data analysis and it's methods

Ms. Dinky Kumari^{#1}, Mr. Satyam Raj^{#2}, Mr. Abhishek Kumar Singh ^{#3}

¹ & ^{2&3} Department of MCA, IIMT College of Engineering, Greater Noida, UP, India

Abstract— Managers in this information age possess a treasure trove of data. The term big data relates not only to largeness, but also to the dealing with varied and fast information that is difficult to work with by conventional methods. The need to work with such data, obtain results, and provide valuable understanding becomes important, especially as it continues to grow swiftly. With analytics of big data, managers are able to see everything, like the sales from the daily invoices and information garnered from social cybernetics. The advanced tools and technology provided are of great assistance in analyzing the big data available to give significant information that decision makers require across multiple domains, which is important as the technology grows and the existing data mounts. Data analytics helps provide a coherent means of obtaining value from data, resolving complicated matters, and improving the structure of the organization's decision making.

KEYWORDS: Big Data, Data mining, Analytics, Decision Making

INTRODUCTION

In the information epoch, we often hear the phrase that “data is the new oil.” As a blind-folded person touches an elephant's body and describes it, it undergoes the impressive phenomenon of data sans sophistication, where an explanation is given without a multifaceted understanding of context. Oil can only make sense in reference to techniques of drilling, placing it into barrels, cars, the highway, precise quantities, and cultures dependent upon it in a structured civilization. Various symbols such as ultrasound-like graphical

interfaces have synonyms.

There is an abundance of information, and as oil is a crucial source of energy today, data is essential for every manipulation you can ever think of when removed from bounds of ethics and Terminator-like logic; we oftentimes speak of data as the new oil. But let it reform “an abundance of data is a bane.” Data analytics can thus strengthen governance, improve global security, efficacy in statecraft, and inter-state relations.

WHAT IS BIG DATA??

Big Data is a collection of mixed grouped pieces of information that can be numbers, words, and much more. It is so great in volume that “normal” computer systems are incapable of managing it. Think of every time using a game in your tablet or surfing the internet, a small fragment of information is produced. Now, Big Data is analogous to amassing all such pieces combined. It works like a giant toy box which keeps getting filled with new toys continuously.[2]Because of the massive number of items in the toy box, lack of efficient methods for organizing and “playing” with the toys will arise. However, smart people are constantly trying to solve this problem by developing new ways, which help understand the toys in the big box. These new tools are useful when it comes to solving problems and rationally making decisions like organizing the toys.

Hence, Big Data surmises having an immensely larger toy box of information and sorting out all the parts makes it entertaining to utilize the big box.

CHARACTERSTICS OF BIG DATA

There are five v's of Big Data that explains the characteristics:-

- **Volume (Lots of Data):** Imagine having a huge stack of toys. Volume is like having so many toys that your regular toy box can't hold them all.
- **Velocity (Fast Data):** Think of how quickly you can play a game. Velocity is about having information come in really, really fast, like when you're playing a super-speedy video game.
- **Variety (Different Types of Data):** Consider your toy box having not only toys but also books, drawings, and maybe even some cool rocks. Variety is about having different types of things in your collection.
- **Veracity (Truthfulness of Data):** Picture having a mix of new and old toys, and some are missing parts. Veracity is making sure all your toys are in good shape and not missing pieces.
- **Value (Useful Information):** When you find your favorite toys and learn new things while playing, that's like the value of Big Data. It's about getting useful and cool stuff from all the information you have.[3]

HOW DOES BIG DATA WORK?

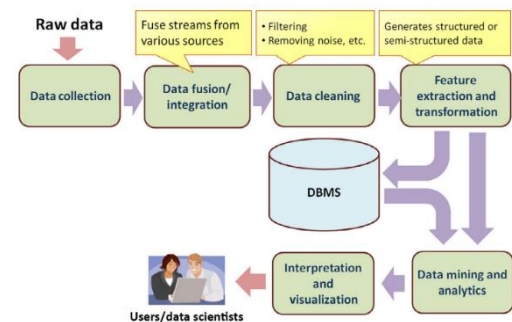
Having greater visibility on any aspect of a business and having access to actionable data helps in making better decisions, identifying growth opportunities, and refining predictive models. This is the fundamental principle on which big data functions.

The work involved in big data can be encapsulated in three broad activities:

Integration: The initial step requires gathering huge volumes of raw data from diverse sources. With the emergence of big data, businesses are now in possession of terabytes of data which can be received processed and transformed into a format that is business ready and usable by the analysts.

Management: Big data needs to be housed, be it in the cloud or on premises, or even both. Regardless of the format, large storage is required. The data must be kept in its required form in order to be processed and made available in real time.

Analysis: Actionable results can be achieved only if the new data is analyzed properly so, big data must be scrutinized. Such actions require several instruments including graphs, charts, and dashboards.



BIG DATA BENEFITS

- Improved Decision -Making
- Increased Agility and Innovation
- Better Agility and motivation
- Continuous Intelligence
- Improved Risk Management
- More efficient operations

BIG DATA ANALYTICS TOOL AND METHODS

Big Data Analytics Tools:

In this table we are defining about the tools which we will use in the analysis of big data such as Hadoop, spark, tableau, Splunk, Apache Flink etc..

Tools used	Uses
Hadoop	Helps organize and process really big amounts of data.
Spark	Makes it fast and easy to work with large sets of data, like in games or predictions.
Tableau	Makes pictures and charts to help understand big and complicated information.
Splunk	Looks at lots of data quickly to find important things, like checking logs or security data.
Apache Flink	Deals with data that comes in really fast, like from sensors or devices.
R and Python (Programming Languages)	Help people talk to computers and analyze data easily.

Big Data Analytics Methods:

In this table we are going to define the methods which we are going to use to analyze the big data such as the machine learning, data mining, text analytic etc.

Methods	Uses
Machine Learning:	Helps computers learn from data and make predictions without being explicitly told what to do
Data Mining:	Finds hidden patterns and useful information in lots of data.
Text Analytic	Understands and pulls out important

	information from written words.
Predictive Analytics	Looks at past data to make smart guesses about what might happen in the future.
Data Warehousing	Keeps all your data in one organized place for easy access.
Streaming Analytics:	Watches and analyzes data as it happens in real-time, useful for things like quickly responding to changes in information.[5]

BIG DATA STORAGE AND MANAGEMENT

It discusses the complexity of handling large data in companies, both through conventional as well as modern ways. Conventional techniques like relational databases, data warehouses, and data marts are used to store structured data. The data is processed very carefully through Extract, Transform, Load (ETL) or Extract, Load, Transform (ELT) so that it is cleaned, transformed, and cataloged before it is made available for analysis.

It makes a case for MAD analysis skills in the context of big data, calling for Magnetic, Agile, and Deep skills. "Magnetic" suggests bringing in all sources of data, whether or not they are good, reflecting the constantly growing ubiquity of data. "Agile" emphasizes that storage technologies need to be flexible and able to rapidly generate and transform data to keep up with changing requirements. "Deep" emphasizes the need for large data stores to accommodate intricate statistical calculations and facilitate in-depth exploration of large sets of data. It also discusses different storage solutions specific to big data issues. Such solutions include distributed systems and Massive Parallel Processing (MPP) databases, which offer high query performance and scalability. Non-relational databases, i.e., NoSQL, are geared for storing unstructured or non-relational information with emphasis on scalability, flexibility, and ease of application development. In-memory databases, where data is held in server memory rather than disk, remove disk input/output (I/O), allowing for real-time response and enormous performance benefits, particularly for analytics on big data.

The Hadoop system is presented as a complete big data analytics solution that assures reliability, scalability, and manageability. It has two major

components: Hadoop Distributed File System (HDFS) for storage and MapReduce for analysis. HDFS provides a redundant and reliable distributed file system that is optimized for big files with a focus on availability and reliability through replication.

APPLICATION OF DATA ANALYSIS

Banking and Securities: Data analytics in the banking sector is an incredibly powerful tool that serves a variety of purposes, such as detecting fraud, segmenting customers, scoring credit, boosting operational efficiency, and even enhancing predictive marketing strategies.

Media & Entertainment: In the media and entertainment world, data analytics is used to delve into audience behavior, preferences, and content consumption patterns. This allows for personalized content recommendations, targeted advertising, and the optimization of business strategies.

Pharma & Healthcare: Within pharma and healthcare, data analytics plays a crucial role in examining patient records, treatment outcomes, and clinical trial data. This provides invaluable insights for personalized medicine, enhances patient care, optimizes treatment protocols, and speeds up drug development processes.

Education: In the education sector, data analytics is utilized to assess student performance and learning patterns. This helps create personalized learning experiences and supports educational institutions in making data-driven decisions to improve overall student success and educational outcomes.

CHALLENGES OF IMPLEMENTING DATA ANALYSIS

Not Enough Skilled People: There's a shortage of experts who know how to effectively use big data tools. This makes it challenging for businesses to fully leverage

their data. Imagine having a fantastic gadget but not knowing how to tap into all its features.

Data Keeps Growing Fast: Big data is constantly expanding. Without a solid system in place to manage all this information, it can be quite a challenge to handle and utilize it effectively. Think of a library where new books are continuously arriving, and you need a reliable way to organize and locate what you're looking for.

Data Isn't Always Good Quality: Sometimes, the information within big data can be messy or inaccurate. Relying on this kind of data can lead to misguided conclusions and poor decisions. It's like trying to complete a puzzle with damaged or missing pieces—you might not get the full picture.

Rules and Laws About Data: Since big data often contains sensitive information, businesses must adhere to regulations and laws to protect privacy: Think of big data as secret treasure, and there are rules about how you can use and protect that treasure.

Data from Different Places is Hard to Put Together : Businesses usually have data in different places, and making it all work together is tough. It's like trying to solve a puzzle with pieces scattered all over the place. Example: Imagine having pieces of a jigsaw puzzle in different rooms – you need a way to bring them all together to see the complete picture.

Worries about Keeping Data Safe: Big data has important information, and bad people might try to get it. It's hard to make sure it's always safe from hackers and other threats. Example: Think of big data as a valuable treasure chest, and businesses need strong locks and guards to keep it safe from thieves.[7]

CONCLUSION

This research dives into the fascinating realm of big data, a hot topic that's capturing a lot of interest due to its incredible potential. In our information-saturated world, we're constantly generating various types of data at lightning speed. Within this vast sea of information lie intricate details and patterns that hold valuable insights just waiting to be uncovered. That's where big data analytics comes into play; it's like a powerful tool that helps us tap into the potential to revolutionize businesses and enhance decision-making.

Big data analytics employs advanced techniques to make sense of large datasets. It serves as a unique key that reveals hidden insights, empowering businesses to adapt and make informed choices. This is especially vital in our fast-paced, data-driven environment. By identifying patterns and extracting key information from massive datasets, big data analytics not only highlights details we might overlook but also lays the foundation for strategic decision-making across various fields.

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.Simplilearn- <https://simplilearn.com/what-is-big-data-analytics-article>
6. Successive cloud- <https://successive.cloud/top-10-industries-that-benefit-most-from-data-analytics/>
7. Google cloud- <https://cloud.google.com/learn/what-is-big-data>

