e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

# BIG DATA ANALYTICS IN BUSINESS

**S. Ramya*1, R. Sakthidevi*2**

*1,2Assistant Professor RAAK Arts and Science College, Perambai. India.

## ABSTRACT

Big data analytics is the complex process of examining big data to uncover information such as hidden patterns, correlations, market trends and customer preferences that can help organizations make informed business decisions. On a broad scale, data analytics technologies and techniques give organizations a way to analyze data sets and gather new information. Due to the rapid growth of data, solutions need to be studied and provided in order to handle and extract value and knowledge from these datasets. Furthermore, decision makers need to be able to gain valuable insights from such varied and rapidly changing data, ranging from daily transactions to customer interactions and social network data. Such value can be provided using big data analytics, which is the application of advanced analytics techniques on big data. The aim of the paper is to demonstrate that Big Data analytics is an effective support in managing the company. It also indicates the areas and activities where the use of Big Data analytics can bring the greatest benefits to companies and also the challenges in using the bigdata.

**Keyword:** Velocity, Hadoop, Data Ware House, Variability, Visualization.

## I. INTRODUCTION

Organizations can use big data analytics systems and software to make data-driven decisions that can improve business-related outcomes. The benefits may include more effective marketing, new revenue opportunities, customer personalization and improved operational efficiency. With an effective strategy, these benefits can provide competitive advantages over rivals.

## II. TYPES OF BIG DATA

As the Internet age continues to grow, we generate an incomprehensible amount of data every second. So much so that the number of data floating around the internet is estimated to reach 163 zetta bytes by 2025. That's a lot of tweets, selfies, purchases, emails, blog posts, and any other piece of digital information that we can think of. These data can be classified according to the following types:

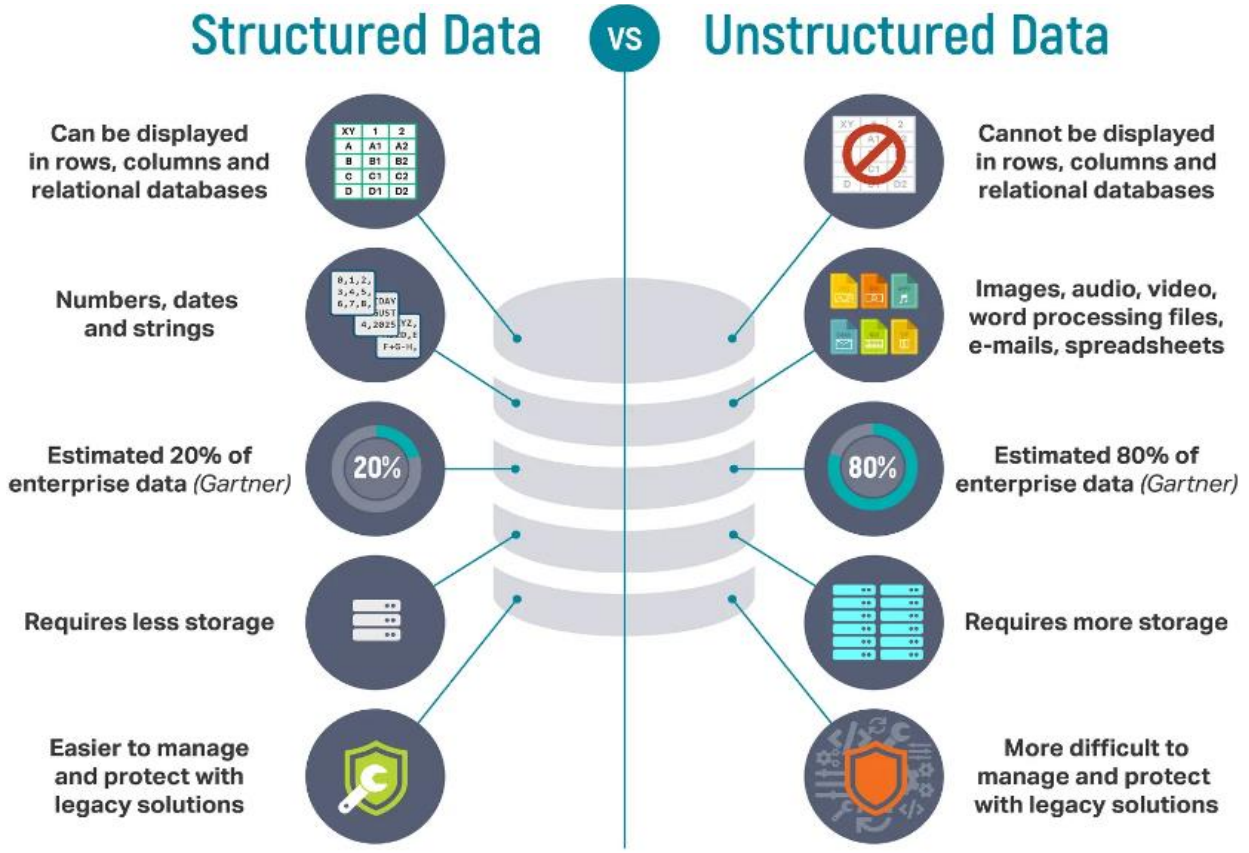
**Structured data**

Structured data has certain predefined organizational properties and is present in structured or tabular schema, making it easier to analyze and sort. In addition, thanks to its predefined nature, each field is discrete and can be accessed separately or jointly along with data from other fields. This makes structured data extremely valuable, making it possible to collect data from various locations in the database quickly.

**Unstructured data**

Unstructured data entails information with no predefined conceptual definitions and is not easily interpreted or analyzed by standard databases or data models. Unstructured data accounts for the majority of big data and comprises information such as dates, numbers, and facts. Big data examples of this type include video and audio files, mobile activity, satellite imagery, and No-SQL databases, to name a few. Photos we upload on Facebook or Instagram and videos that we watch on YouTube or any other platform contribute to the growing pile of unstructured data.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

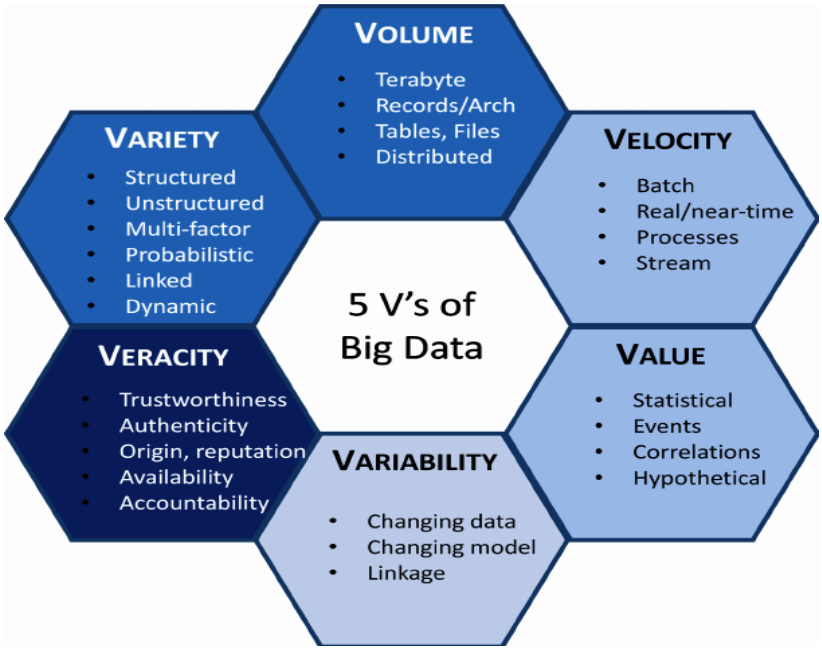



## III. CHARACTERISTICS OF BIG DATA

General characteristics of Big Data can be referred to as the five Vs: Volume, Velocity, Variety, Veracity, and Value. They have been elucidated below:-

1. **Volume:** Volume is the size of a dataset processed and stored in the Big Data System and is known to be its most important and prominent feature. The size of data usually ranges from petabytes to Exabyte's and is processed with advanced processing technology.
2. **Velocity:** Velocity is referred to as the data accumulation rate, which also helps analysts determine if it falls under the classification of regular data or Big Data. Data needs real-time evaluation, which requires well-integrated systems for handling the amount and pace of generated data.
3. **Variety:** Variety is defined as the type of data format and the way it is organized and made ready to be processed. The data accumulation rate also influences whether the data is classified as Big Data or regular data. The speed of data processing essentially means that more data will be available than the previous set and also that the data processing rate will be high.
4. **Veracity:** Veracity is the quality and reliability of the data in concern. Unreliable data devalues the authenticity of Big Data, especially when the data is updated in real-time. Therefore, data authenticity requires regular checks at every level of collection and processing.
5. **Value:** Value is also worth considering in collecting and processing Big Data. More than the amount of data, the value of that data is important for acquiring insights.
6. **Variability:** Variability is the characteristic of Big Data that enables it to be formatted and used for actionable purposes.

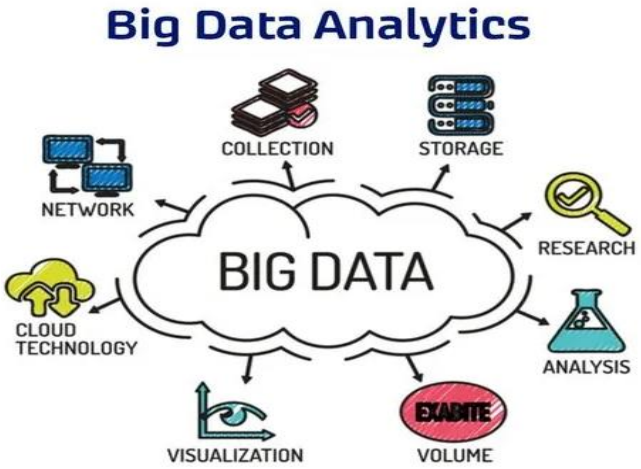
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com




## IV. HOW DOES BIG DATA ANALYTICS WORK?

1. Data analysts, data scientists, predictive modelers, statisticians and other analytics professionals collect, process, clean and analyze growing volumes of structured transaction data as well as other forms of data not used by conventional BI and analytics programs. Here is an overview of the four steps of the big data analytics process:

2. Data professionals collect data from a variety of different sources. Often, it is a mix of semistructured and unstructured data. While each organization will use different data streams, some common sources include internet clickstream data, web server logs, cloud applications, mobile applications, social media content, text from customer emails and survey responses mobile phone records and machine data captured by sensors connected to the internet of things (IoT).




3. Data is prepared and processed. After data is collected and stored in a data warehouse or data lake, data professionals must organize, configure and partition the data properly for analytical queries. Thorough data preparation and processing makes for higher performance from analytical queries.

4. Data is cleansed to improve its quality. Data professionals scrub the data using scripting tools or data quality software. They look for any errors or inconsistencies, such as duplications or formatting mistakes, and organize and tidy up the data.

5. The collected, processed and cleaned data is analyzed with analytics software. This includes tools for data mining, which sifts through data sets in search of patterns and relationships. Predictive analytics, which builds

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

models to forecast customer behavior and other future actions, scenarios and trends. Machine learning, which taps various algorithms to analyze large data sets. Deep learning which is a more advanced offshoot of machine learning. Text mining and Statistical analysis software. Artificial intelligence (AI) and business intelligence software and data visualization tools

## V. KEY BIG DATA ANALYTICS TECHNOLOGIES AND TOOLS

Many different types of tools and technologies are used to support big data analytics processes. Common technologies and tools used to enable big data analytics processes include:

- **Hadoop,** which is an open source framework for storing and processing big data sets. Hadoop can handle large amounts of structured and unstructured data.
- **Predictive analytics** hardware and software, which process large amounts of complex data, and use machine learning and statistical algorithms to make predictions about future event outcomes. Organizations use predictive analytics tools for fraud detection, marketing, risk assessment and operations.
- **Stream analytics** tools, which are used to filter, aggregate and analyze big data that may be stored in many different formats or platforms.
- **Distributed storage** data, which is replicated, generally on a non-relational database. This can be as a measure against independent node failures, lost or corrupted big data, or to provide low-latency access.
- **NoSQL databases**, which are non-relational data management systems that are useful when working with large sets of distributed data. They do not require a fixed schema, which makes them ideal for raw and unstructured data.
- **A data lake** is a large storage repository that holds native-format raw data until it is needed. Data lakes use a flat architecture.
- **A data warehouse,** which is a repository that stores large amounts of data collected by different sources. Data warehouses typically store data using predefined schemas.
- **Knowledge discovery/big data mining** tools, which enable businesses to mine large amounts of structured and unstructured big data.
- **In-memory data fabric**, which distributes large amounts of data across system memory resources. This helps provide low latency for data access and processing.
- **Data virtualization**, which enables data access without technical restrictions.
- **Data integration software,** which enables big data to be streamlined across different platforms, including Apache, Hadoop, MongoDB and Amazon EMR.
- **Data quality software**, which cleanses and enriches large data sets.
- **Data preprocessing software,** which prepares data for further analysis. Data is formatted and unstructured data is cleansed.
- **Spark,** which is an open source cluster computing framework used for batch and stream data processing.

Big data analytics applications often include data from both internal systems and external sources, such as weather data or demographic data on consumers compiled by third-party information services providers. In addition, streaming analytics applications are becoming common in big data environments as users look to perform real-time analytics on data fed into Hadoop systems through stream processing engines, such as Spark, Flink and Storm.

Early big data systems were mostly deployed on premises, particularly in large organizations that collected, organized and analyzed massive amounts of data. But cloud platform vendors, such as Amazon Web Services (AWS), Google and Microsoft, have made it easier to set up and manage Hadoop clusters in the cloud. The same goes for Hadoop suppliers such as Cloud era, which supports the distribution of the big data framework on the AWS, Google and Microsoft Azure clouds. Users can now spin up clusters in the cloud, run them for as long as they need and then take them offline with usage-based pricing that doesn't require ongoing software licenses.

Big data has become increasingly beneficial in supply chain analytics. Big supply chain analytics utilizes big data and quantitative methods to enhance decision-making processes across the supply chain. Specifically, big

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

supply chain analytics expands data sets for increased analysis that goes beyond the traditional internal data found on enterprise resource planning (ERP) and supply chain management (SCM) systems. Also, big supply chain analytics implements highly effective statistical methods on new and existing data sources.

## VI. BIG DATA ANALYTICS USES AND EXAMPLES

Here are some examples of how big data analytics can be used to help organizations:

- **Customer acquisition and retention.** Consumer data can help the marketing efforts of companies, which can act on trends to increase customer satisfaction. For example, personalization engines for Amazon, Netflix and Spotify can provide improved customer experiences and create customer loyalty.
- **Targeted ads.** Personalization data from sources such as past purchases, interaction patterns and product page viewing histories can help generate compelling targeted ad campaigns for users on the individual level and on a larger scale.
- **Product development.** Big data analytics can provide insights to inform about product viability, development decisions, progress measurement and steer improvements in the direction of what fits a business' customers.
- **Price optimization.** Retailers may opt for pricing models that use and model data from a variety of data sources to maximize revenues.
- **Supply chain and channel analytics.** Predictive analytical models can help with preemptive replenishment, B2B supplier networks, inventory management, route optimizations and the notification of potential delays to deliveries.
- **Risk management.** Big data analytics can identify new risks from data patterns for effective risk management strategies.
- **Improved decision-making.** Insights business users extract from relevant data can help organizations make quicker and better decisions.

## VII. BIG DATA ANALYTICS BENEFITS

The benefits of using big data analytics include:

- Quickly analyzing large amounts of data from different sources, in many different formats and types.
- Rapidly making better-informed decisions for effective strategizing, which can benefit and improve the supply chain, operations and other areas of strategic decision-making.
- Cost savings, which can result from new business process efficiencies and optimizations.
- A better understanding of customer needs, behavior and sentiment, which can lead to better marketing insights, as well as provide information for product development.
- Improved, better informed risk management strategies that draw from large sample sizes of data.

## VIII. BIG DATA ANALYTICS CHALLENGES

Despite the wide-reaching benefits that come with using big data analytics, its use also comes with challenges:

- **Accessibility of data.** With larger amounts of data, storage and processing become more complicated. Big data should be stored and maintained properly to ensure it can be used by less experienced data scientists and analysts.
- **Data quality maintenance.** With high volumes of data coming in from a variety of sources and in different formats, data quality management for big data requires significant time, effort and resources to properly maintain it.
- **Data security.** The complexity of big data systems presents unique security challenges. Properly addressing security concerns within such a complicated big data ecosystem can be a complex undertaking.
- **Choosing the right tools.** Selecting from the vast array of big data analytics tools and platforms available on the market can be confusing, so organizations must know how to pick the best tool that aligns with users' needs and infrastructure.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

- With a potential lack of internal analytics skills and the high cost of hiring experienced data scientists and engineers, some organizations are finding it hard to fill the gaps.

## IX. CONCLUSION

Big Data has been extensively used in recent years in every aspect of our lives and every possible sector of global industries. It is one of the most valuable resources in the market, used to optimize any operational process. As an aspirant of data science, it is imperative to have the basic skills and knowledge about fundamental aspects of data analysis and to learn about the different types of data in analytics. You can take your first step into this lucrative career field by pursuing a reliable course or undergoing professional training. A great way to start is by taking part in Knowledge Hut's Big Data training. Big data analytics can lead to more effective marketing, new revenue opportunities, improved operational efficiency, competitive advantages over rival organizations and other business benefits.

## X. REFERENCES

[1] School of Business, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macao 999078, China
[2] The School of Management, Jinan University, Zhuhai 519000, China
[3] The Faculty of Data Science, City University of Macau, Macao 999078, China
[4] The Department of Artificial Intelligence and Big Data Applications, Zhuhai Institute of Advanced Technology Chinese Academy of Sciences, Zhuhai 519000, China
[5] The School of Computer, College of Beijing University of Technology Zhuhai, Zhuhai 519000, China
[6] Abbasi, A., Sarker, S., & Chiang, R. H. (2016). Big Data Research in Information Systems: Toward an Inclusive Research Agenda. Journal of the association for information systems, 17(2).
[7] Chen, D. Q., Preston, D. S., & Swink, M. (2015). How the use of big data analytics affects value creation in supply chain management. Journal of Management Information Systems, 32(4), 4-39.