International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

BIG DATA ANALYTICS: A LITERATURE REVIEW PAPER

[1]SHREYAS V V REDDY, [2]RAKSHITHA KIRAN
[1]PG SCHOLAR, [2]ASSISTANT PROFESSOR
DEPARTMENT OF MCA
DAYANANDA SAGAR COLLEGE OF ENGINEERING
AFFILIATED TO VTU, BENGALURU, INDIA

Abstract

In the digital era, decision makers have access to a large amount of data. As a result of the increasing volume and velocity of data in the world, conventional data management methods are no longer able to keep pace. As the amount of data grows, new ways to manage and extract value and information must be developed. There must be the ability to extract desired information from a broad variety of continually changing data, such as customer interactions and social network activity. Using the newest analytics methods on a vast amount of data, Big Data Analytics may provide such benefits.

Introduction

It's possible to imagine a future in which every piece of information on a person or a company is erased immediately after usage. When it comes to collecting and analysing vital data and information, organisations will lose their ability to do so. Take a look at how much information is now at our fingertips thanks to advancements in technology and the widespread use of the internet. Due to increased storage capacity and improved data collection technologies, a greater volume of data is now easily available. New data is created every second, and it is important to preserve and progress this data in order to extract value from it. Even as data storage costs have decreased, organisations should be able to extract as much value from large volumes of data as feasible. New forms of big data analytics, new repositories, and new research methodologies are needed because of the data's bulk, diversity, and quick evolution. Large amounts of huge data need careful examination."

2 Big Data Analytics

The term "Big Data" has been given to a set of metadata that has become so large that it is impossible to deal with it using traditional database management systems. Software tools and storage systems may not be able to compile, gather, manage or analyse data sets that are too massive.

From a few dozen terabytes (TB) to several petabytes (PB), the size of big data has increased steadily since its inception (PB). As a result, big data difficulties include data collection, storage, search, sharing, analysis, and visualisation. Businesses are increasingly sifting through enormous amounts of data to find information they didn't know existed.

2.1 Characteristics of Big Data

Because of its size, spread, and variety, big data necessitates new technology infrastructures, analytics, and tools to uncover new sources of business value. The "three Vs" of big data are volume, variety, and velocity, and they're well-known in the industry. The volume of the data is used to

determine the size and scope of the data. As data changes or new data is generated, it's called velocity. Diversity also refers to the varied ways in which data may be used and analysed as well as the various forms and kinds of data they come in.

2.2 Big Data Analytics Tools and Methods

It is imperative that data analysts be able to operate swiftly and effectively in a constantly changing technological and data environment. Making rapid and accurate judgments requires more than just a lot of data.

It would be impossible for traditional data management and research methodologies to handle this amount of data. Big data analytics methods and storage and management infrastructures are so required. Because of this, big data has an impact on every step of the data gathering, processing, and analysis process.

Big Data Storage and Management

When dealing with huge volumes of data, organisations must think about how and where the data will be stored once it has been acquired. Database and warehouse management systems employ ETL or ELT technologies to take data from operational databases and convert it before loading it into a database. Before making data available for online analysis and mining, be sure it has been cleaned, processed, and classified.

In other cases, two databases are not connected to one other at all. In addition to unstructured, or non-relational, data, SQL and other non-relational databases were inspired by the creation of SQL. Data model flexibility, huge scalability, and ease of programme creation and deployment are some of the benefits of using NoSQL databases. SQL databases vary from relational databases in terms of data management and storage features. Thus, they put a greater premium on high performance, scalability, and application-level data management.

Real database responses are possible with in-memory databases, which store data in the server's memory rather than on a disc. The bulk of the database is stored in silicon-based memory rather than on mechanical disc drives. It is common practise to employ in-memory databases to speed up the access and scoring of analytical models in more advanced big data analytics.

Big Data Analytic Processing

Analyzing data is the next step after storing a lot of it. There are four essential conditions that must be fulfilled in order to handle massive amounts of data. First and foremost, the capacity to load data quickly is essential. Due to disc and data traffic interfering with query execution during loading data, it is necessary to minimise data loading time. As a result, inquiries should be resolved quickly. The data placement structure should thus be able to sustain high query processing rates as the amount of inquiries increases.

Map Reduce is a parallel programming paradigm that was inspired by the "Map" and "Reduce" methods of functional languages. This core Hadoop component is in charge of processing and analysing large amounts of data. The MapReduce paradigm relies on growing the storage capacity or potentiality of a single machine rather than adding more processors or sources due to electromagnetic compatibility concerns. It is based on the idea that a process may be broken down into phases and conducted in parallel to minimise the time it takes to finish the job.

3.1 Customer Intelligence

Retail, financial and telecoms companies may benefit from using big data analytics to better understand their consumers. A metadata-based system is being considered as a way to increase openness and make important data more easily accessible to participants in real time. As a result of using big data analytics, organisations may discover and categorise their target audiences based on a wide range of demographic traits, as well as increase customer happiness and retention. Marketers might use this information to make better decisions and target consumers depending on their

preferences. To better understand their consumers' wants and needs, businesses may utilise social media.

3.2 Supply Chain and Performance Management

In supply chain management, big data analytics may be utilised to forecast demand changes and match supply accordingly. Manufacturers and retailers as well as transportation service providers might all benefit from this. If you analyse supplier performance data, you may determine whether or not to switch suppliers based on quality or price competitiveness. The reduction of inventory and the increase in profit margins that follow from simultaneously testing many price scenarios, Big data analytics may help with performance management, which is particularly crucial in public sector and healthcare contexts, although this yet to be shown. Increasing productivity necessitates the use of predictive analytics systems to track and forecast employee performance. Predictive KPIs, balanced scorecards, and dashboards have the potential to benefit an organization's operations by improving transparency, goal setting, and other planning and management functions, thanks to the ease with which operations managers can now access and use big data and performance information. Monitoring and Improving Product Quality 3.3

3.4 Risk Management and Fraud Detection

The analysis of the likelihood of returns in comparison to the possibility of losses is a significant aspect of the financial services industry. Big data analytics could be able to assist in the selection of assets by doing this analysis. For a dynamic and comprehensive risk assessment, it is possible to do analysis on large amounts of data derived from both internal and external sources. Because of this, businesses may get an advantage from big data by being able to do more accurate risk assessments. Combining pre-existing risk profiles with high-performance analytics is another option.

New big data tools and technologies have the potential to assist in managing the exponential growth of network-generated data. This may be accomplished by increasing the database's scalability and the amount of vital data that it collects. Technology breakthroughs in data-intensive computing and cyber analytics have made it possible for organisations to improve their ability to protect themselves against assaults on their networks and online. Analytics performed on large amounts of data might be used in the public sector, the financial sector, and the insurance business to detect and prevent fraud. Despite the widespread use of analytics, businesses of all stripes are increasingly turning to big data in order to facilitate automated fraud detection. This is the case even if analytics are often employed. A risk profile for the whole organisation is created by combining the risk profiles of many different divisions. In addition, customer intelligence may be used to imitate regular customer behaviour and recognise suspicious or deviant activities by properly highlighting instances of outliers. This can be accomplished by the accurate highlighting of outlier occurrences. When fraud detection systems have access to data on existing fraud trends, they may learn about new types of fraud and respond efficiently when criminals alter their techniques. This allows them to learn about new kinds of fraud. SNAs may also be used to find evidence of fraudulent insurance or benefit claims. This helps to lessen the risk that fraudulent activity will go undetected. It is possible that the use of big data technologies, techniques, and governance procedures will considerably accelerate the finding of fraudulent transactions.

Conclusion

It has lately been a hot topic owing to its supposed unrivalled potential and advantages, which we examined in our research. In the digital world we live in, massive amounts of high-speed data are generated on a daily basis, each with inherent characteristics and patterns of hidden information that should be recovered and used.

Organizational transformation and decision-making may be facilitated and enhanced by using sophisticated analytic approaches to massive data sets and uncovering meaningful information from the data.

It is possible to gather and use information from enormous datasets using these kinds of analytics, which may enhance decision-making and assist educated judgements. There are numerous areas where big data analytics may help and enhance decision-making, and some of these areas have been investigated. Big data analytics has been shown to have a variety of uses, including customer intelligence, fraud detection, and supply chain management. There are several sectors and businesses that might profit from its advantages, including healthcare, retail, communication and manufacturing.

There are several benefits and gains to be gained from the effective use of any new technology, especially when it comes to the fascinating field of big data management. On the other side, huge data administration is very difficult. Data storage, administration, integration, federation, cleansing, processing, and analysis are only some of the requirements.

References

[1] https://www.journals.elsevier.com/big-data-research

[2] https://www.researchgate.net/publication/264555968_Big_Data_Analytics_A_Literature_Review_Paper

[3] https://journalofbigdata.springeropen.com/articles/10.1186/s40537-015-0030-3