Recent Research Reviews Journal (ISSN: 2583-7079)
https://irojournals.com/rrrj

Big Data Analytics in Distributed Business Management

Thesnath A/L Velayudhan[1], Dinesh Rajassekharan[2]

[1] Faculty of Computer Science and Multimedia, Lincoln University College.

[2] Associate Professor, Faculty of Computer Science and Multimedia, Lincoln University college, Malaysia.

E-mail: [1] Malaysia.thesnath@lincoln.edu.my, [2] diraja@lincoln.edu.my

Abstract

Large-scale data processing and analysis across several locations in a corporate setting is accomplished through the use of sophisticated analytics techniques, which is known as big data analytics in distributed business management. The data can be spread if it comes from several departments, business divisions, locations, or even outside services. Knowledge, data, and information have always been essential to business. Since there is an expanding quantity of data that may be gathered and stored, businesses require innovative approaches to data processing and evaluation. The notion of 'Big Data' is discussed in this article. The study aims to prove that big data analytics is a useful tool for business management. It also shows which tasks and domains may benefit businesses the most from the application of big data analytics.

Keywords: Big Data, Data processing, Business management, Data Management, Cloud Computing.

1. Introduction

Developing business process improvement is a challenging endeavor that requires complex and dependable support systems [12, 13]. However, because of the massive volumes of event data created by the execution of activities over the course of an organization's existence, business customers are unable to efficiently access timely analytics data. This article provides business analysts with visibility into dispersed operations and company performance through a technological solution powered by big data [14-17].

In the context of Big Data, one of the most important issues is information overload. Accurately locating the information customers need from the massive volumes of available data is getting harder. However, obtained information might be incredibly beneficial if firms are able to handle, analyze, and accumulate massive datasets. Data accessibility has to be accelerated in this era of exponential expansion of corporate information. Accurate knowledge management, often known as business intelligence, is fundamental to modern corporate operations. The advancement of communication technology, together with the widespread and ubiquitous presence of communication networks, have made it possible to collect and analyse data with the goal of obtaining valuable knowledge [18].

In cluster computing systems, distributed data management is a critical technology that makes large-scale data processing and analysis efficient. More specifically, huge data files must be summarized using representative samples that share the same statistical characteristics as the entire dataset in contexts where the quantities of data exceed the capacity of the system.

1.1 Data Management

Techniques, concepts, and strategies for data development modeling approaches and data models; data administration, security, and operations; data integrity and quality; data conversion and migration; universal servers and database servers; object databases, object modeling, and object technologies; web distributed databases; portability, interoperability, and replication of data; information on the Internet and Intranets; and data and its relationship with the Internet and Intranets. The most important aspect to information management is data management, which entails the creation and use of systems, policies, and protocols for precise data processing. It is best accomplished using database management systems and is necessary for making wise management decisions. Using a computerised data system to organise, store, and preserve the integrity for electronic data while emphasizing effectiveness, privacy, and secrecy is known as data management in research. An approach called data analysis management enables analysts to keep track of data analysis events, set milestones, and go back to earlier save states.

1.2 Big Data

Big data is a collection of large, intricate data sets that, because of their volume, diversity, and pace, call for specific processing techniques. In the modern world, this data—which is larger than what conventional databases can hold—is essential. Businesses and organisations may get important insights along with a competitive edge by utilising big data analytics. But there are drawbacks to big data deployment as well, such privacy issues. Big data's ultimate objective is to produce business solutions which can aid an organisation in gaining business solutions. The importance of large data analysis is demonstrated by one factor alone.

Big data is information that can't be processed by conventional databases. A single machine cannot process the volume of data. To handle and store such massive amounts of data, new and creative techniques are needed. An overview of big data, its significance in our lives, and several solutions for handling it are given in this study. Below (Figure 1) are the current Big Data applications in the trending industries.

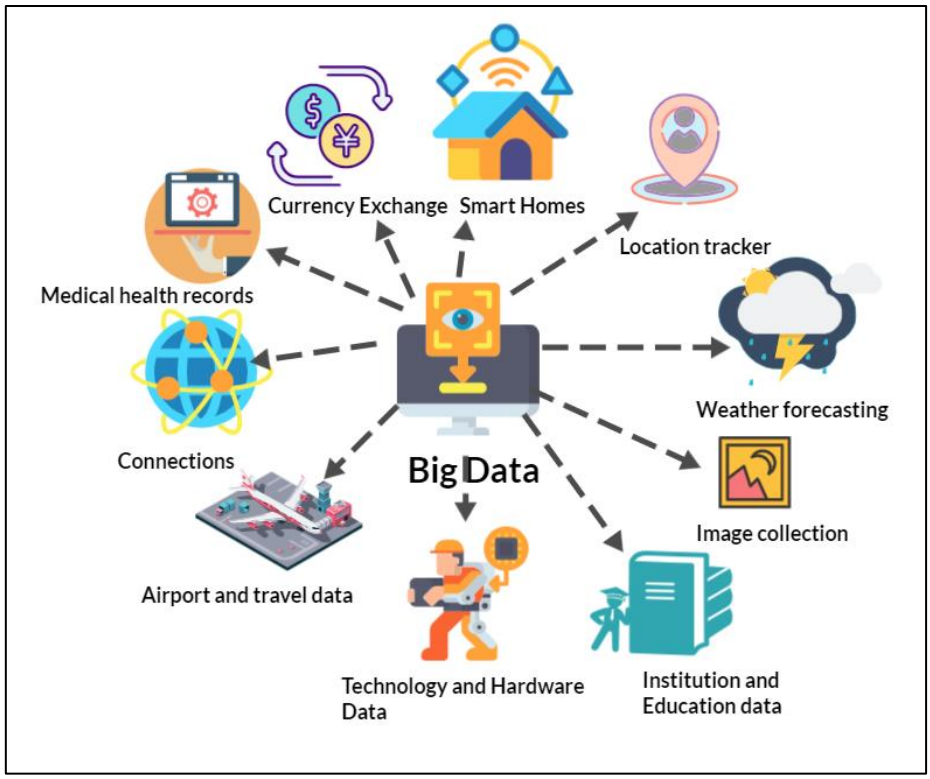

Figure 1. Application of Big Data Infrastructure

1.3 Data Security

Data security in big data management is a complicated problem with many facets. This is especially true when it comes to cloud computing, when data security is crucial. This study highlights the necessity of a thorough security strategy that addresses every phase of the data life cycle. There is a need for more study in this area and for intelligent analytics to improve security.

Big data involves a variety of data formats, data collecting from many sources, and data streaming that might lead to security issues and violate users' right to privacy due to outdated technologies. To solve such complicated problems, rules and regulations must be put in place for data sharing, authentication procedures, the creation of personal information, and system abuse. Data should be exchanged in a safe environment without invading people's privacy.

2. Literature Survey

Big data analytics is a useful tool for business management [4]. The purpose of the article is to show how big data analytics may help with strategic and operational management of businesses.

The findings of an empirical study about the use of big data priorities are given, demonstrating the practical value of analytics for big data across a range of business management domains.

Big Data analytics has shown to be practically effective in many aspects of business management, particularly strategic and stakeholder management. This implies that using big data analytics to your benefit might provide you a competitive edge in the market.

Emerging technology and sophisticated analytics techniques can support corporate users in continually streamlining their operations. Complex and reliable supporting systems are necessary for business process improvement to occur continuously. Emerging technology and sophisticated analytics techniques can support corporate users in continually streamlining their operations. With a quick reaction time, the suggested design enables users to assess business performance in highly dispersed situations [5].

The distributed random data sample blocks serve as the foundation for a large data management system proposal. In addition to providing functionality for block-level administration, the study suggests a large data management system (BDMS) designed around distributed random data samples blocks and demonstrates an appropriate execution time for partition operation in practical applications [6].

Big data is the term used to describe vast amounts of data created from various sources combined with data analysis tools to search through this massive quantity of data for patterns and trends that may be used to forecast future development prospects, acquire insight, make wise decisions, and optimise operations. Big data technology is gaining traction in a number of industries due to its capacity to harvest enormous amounts of data from diverse sources [7].

Business units are using big data to get a competitive edge and achieve growth due to the exponential expansion of data.

2.1 Traditional Data Analysis Vs Big Data Analysis

Table 1 below shows the comparison of traditional and big data analysis

Table 1. Traditional Data Analysis Vs Big Data Analysis

Tradition Data Analysis	Big Data Analysis
Traditional data is superior for regular tasks and organised inquiries, such as calculating monthly sales.	Big data is a master at seeing hidden trends and patterns, such as gleaned from hundreds of social media postings, such as sentiment analysis.
Traditional data refers to the organized information that is mostly kept up to date by businesses of all sizes, from little startups to big organizations.	Big data is a more advanced form of traditional data. Big data applications handle excessively big or complicated data sets, which are challenging to handle with conventional data-processing tools. Large volumes of organised, semi-structured, and unstructured data are handled by it.

Small to medium-sized datasets that are readily stored and analysed with conventional data processing tools are referred to as traditional data.	Big data is the term used to describe incredibly enormous datasets that are difficult to handle or analyse using conventional technology.
Standard system configurations may handle conventional data processing.	Processing large amounts of data requires a high system configuration.

3. Big Data Infrastructure in Business Management

Most businesses utilise a variety of infrastructure technologies in their big data environment. Let us discuss the fundamental methodologies used in big data tasks, how they function, as well as its scenarios.

Hadoop

Hadoop Distributed File System (HDFS): HDFS is a distributed file system which stores data on numerous computers. It divides huge files into smaller chunks and replicates them across several nodes within a Hadoop cluster. HDFS enables fault tolerance by duplicating each block across many nodes. If a particular node fails, that information can still be accessible by other nodes.

MapReduce Programming Model: MapReduce is a programming approach for processing and producing huge datasets that parallelizes computing on a distributed cluster.

The input data is divided into smaller chunks, then a Map function handles each piece separately, resulting in key-value pairs. The framework organizes and arranges the Map phase output according to keys. The Reduce function takes the sorted key-value pairs and produces the final outcome.

Working Scenario of Hadoop

- Raw data from many sources are imported into the Hadoop environment.

- The data is saved in HDFS, which separates it into blocks as well as replicates them among nodes within the Hadoop cluster.
- Developers create MapReduce programmes to process data. MapReduce jobs are sent to the Hadoop cluster.
- The Map jobs execute in parallel on several nodes, processing data and outputting intermediate key-value pairs.The framework shuffles as well as sorts intermediate key-value pairs, arranging them according to key.
- The Reducing tasks process the sorted key-value pairs, combining and providing the desired result.The processed data is saved or utilised for further analysis. Hadoop enables the storing and retrieval of large volumes of organised and unstructured data.

Advantages of Hadoop

Flexibility: Unlike relational databases, data does not need to be processed before it is stored. Hadoop allows you to store as much data as you want and choose how to use it. This also contains unstructured data like photographs, videos, and text.

Speed: It is a substantially faster tool for detailed data analysis since it processes numerous elements of the data set at the same time.

Scalability: Hadoop, unlike traditional systems with restricted data storage, works in a distributed setting and is scalable. Simply adding additional nodes expands the system's capacity to process more data.

Computing Power: Hadoop makes advantage of a distributed computing approach to process large amounts of data fast. Adding more computers can boost processing power.

Cost: The open-source framework is free to be utilized and can store a great deal of data on inexpensive hardware.

NoSQL: NoSQL is a non-relational database technology that enables distributed data processing. Unlike SQL, NoSQL does not require a set structure or a particular query language. This versatility enables NoSQL-based solutions to cope with organised, semi-structured,

including unstructured data from a wide range of sound devices. They can scale quickly because to their schema-agnostic design and distributed architecture.

NoSQL databases were open-source and built to function with massive clusters. MongoDB, Cassandra, and Neo4J are some examples of NoSQL databases as well as database management systems. The Figure 2 depicts the Big Data infrastructure

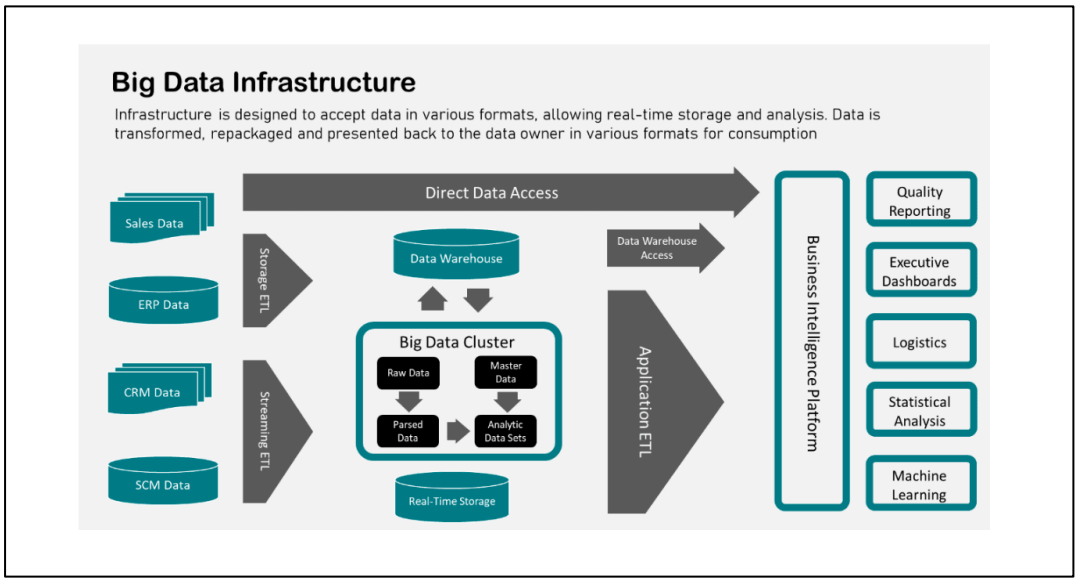

Figure 2. Infrastructure – ETL Technology [10]

4. Market Analysis of BIG DATA in Business Management

Big data analytics software is predicted to have the greatest market share in 2023 due to its revolutionary influence on company decision-making processes. With the exponential expansion of data, organizations are turning to advanced analytics solutions to extract meaningful insights and drive strategic initiatives. The adaptability of big data analytics software enables organizations to leverage machine learning techniques, predictive modelling, and real-time analysis, providing a holistic solution for extracting valuable insight from large and complicated datasets.

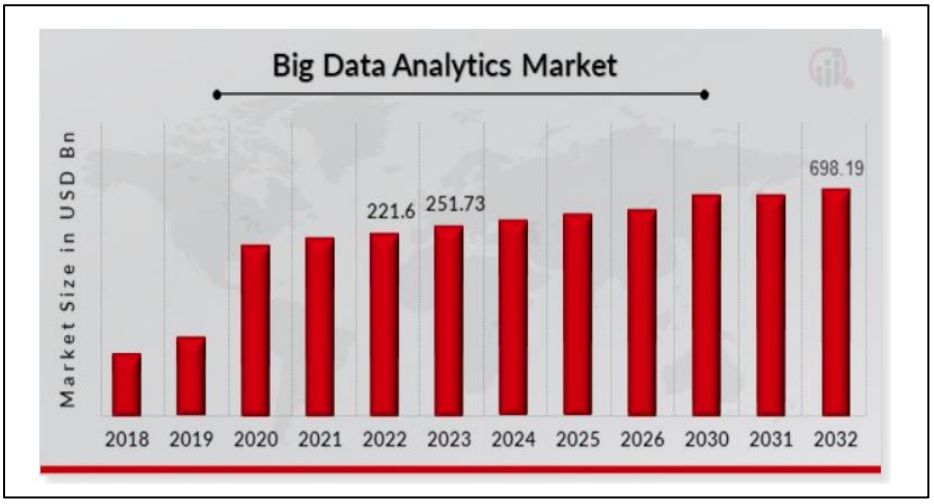

Figure 3. Big Data Market Analysis

5. Benefits of Big Data Analysis

- Companies are able to track the flow of products, spot any bottlenecks, and proactively address interruptions thanks to big data analytics, which gives real-time visibility throughout the whole supply chain.
- Understanding consumer behavior, preferences, and feedback can be assisted by big data analytics.
- Planning and optimising routes may be improved by analysing data on past shipping routes, traffic trends, and transportation expenses.
- This results in lower carrying costs, better order fulfilment, and increased inventory management efficiency all around.
- Distribution managers may stop depending on instinct or conventional techniques and start making well-informed decisions based on real-time information by utilizing data analytics.
- This makes it possible for businesses to create backup plans and efficiently reduce risks, guaranteeing a more robust distribution network.

6. Conclusion

The practical utility of Big Data analytics is obvious in many aspects of business management, particularly strategic management including stakeholder management. Big Data analytics may result in more effective marketing, new income possibilities, increased operational efficiency, advantages in competition for businesses, and other commercial benefits. The practical utility of Big Data analytics is obvious in many aspects of business

management, particularly strategic management and stakeholder management. Big Data analytics may result in more effective marketing, new income possibilities, increased operational efficiency, competitive advantages over competing businesses, and other commercial benefits. In the age of Big Data and new, more complex analytical skills, businesses may obtain a competitive advantage in the market by competing on analytics. The work is part of a growing body of research on data-driven decision-making.

References

[1] Jelonek, Dorota. "Big Data Analytics in the Management of Business." In MATEC Web of Conferences, vol. 125, p. 04021. EDP Sciences, 2017.

[2] [Vera-Baquero, Alejandro, Ricardo Colomo-Palacios, and Owen Molloy. "Business process analytics using a big data approach." It Professional 15, no. 6 (2013): 29-35.

[3] [Emara, Tamer Z., and Joshua Zhexue Huang. "A distributed data management system to support large-scale data analysis." Journal of Systems and Software 148 (2019): 105-115.

[4] Jelonek, Dorota. "Big Data Analytics in the Management of Business." In MATEC Web of Conferences, vol. 125, p. 04021. EDP Sciences, 2017.

[5] Vera-Baquero, Alejandro, Ricardo Colomo-Palacios, and Owen Molloy. "Business process analytics using a big data approach." It Professional 15, no. 6 (2013): 29-35.

[6] Emara, Tamer Z., and Joshua Zhexue Huang. "A distributed data management system to support large-scale data analysis." Journal of Systems and Software 148 (2019): 105-115.

[7] Sandhu, Manbir, and Anuradha Saini Purnima. "A review on Big Data analytics in business." (2018).

[8] Alsghaier, Hiba, Mohammed Akour, Issa Shehabat, and Samah Aldiabat. "The importance of big data analytics in business: a case study." American Journal of Software Engineering and Applications 6, no. 4 (2017): 111-115.

[9] https://segment.com/data-hub/big-data/infrastructure/

[10] https://www.businesstechweekly.com/operational-efficiency/data-management/big-data-infrastructure/

[11] https://www.marketresearchfuture.com/reports/big-data-analytics-market-4503

[12] Elzinga, D. Jack, Tomas Horak, Chung-Yee Lee, and Charles Bruner. "Business process management: survey and methodology." IEEE transactions on engineering management 42, no. 2 (1995): 119-128.

[13] Burlton, Roger. Business process management: profiting from process. Pearson Education, 2001.

[14] Vassakis, Konstantinos, Emmanuel Petrakis, and Ioannis Kopanakis. "Big data analytics: applications, prospects and challenges." Mobile big data: A roadmap from models to technologies (2018): 3-20.

[15] Grover, Varun, Roger HL Chiang, Ting-Peng Liang, and Dongsong Zhang. "Creating strategic business value from big data analytics: A research framework." Journal of management information systems 35, no. 2 (2018): 388-423.

[16] Phillips-Wren, Gloria, Lakshmi S. Iyer, Uday Kulkarni, and Thilini Ariyachandra. "Business analytics in the context of big data: A roadmap for research." Communications of the Association for Information Systems 37, no. 1 (2015): 23.

[17] Schmarzo, Bill. Big Data: Understanding how data powers big business. John Wiley & Sons, 2013.

[18] Rao, T. Ramalingeswara, Pabitra Mitra, Ravindara Bhatt, and Adrijit Goswami. "The big data system, components, tools, and technologies: a survey." Knowledge and Information Systems 60 (2019): 1165-1245.