INTERNATIONAL JOUNAL OF: SCIENCE RESEARCH AND TECHNOLOGY

TIMBOU-AFRICA ACADEMIC PUBLICATIONS
INTERNATIONAL JOURNAL, NOVEMBER, 2022 EDITIONS
VOL. 11 NO. 9 ISSN: 2623-7861

# BIG DATA IN TODAY'S INFORMATION SOCIETY

***UDO, CLEMENT UDO; **AFAHA, IFIOK JAMES; & *OBOH, ANDREW EJE**

*Federal University of Health Sciences, Otukpo, Benue State.
**Department of Library and Information Science, University of Uyo, Uyo.

## ABSTRACT

Understanding what big data is and how it is used has become more critical as the gathering and usage of big data have grown. Big data is a collection of data that is enormous in volume and is rising exponentially over time. Because of its vastness and complexity, no conventional data management solution can store or process it effectively. Big data is just data, except much larger. Therefore, the focus of this study is on the idea of big data, its attributes,

## INTRODUCTION

The importance of creating big data applications has increased over the last few years in society. Several businesses from various industries rely more and more on knowledge gleaned from enormous data quantities peculiar to their services. Traditional data platforms and methodologies, however, are less effective in big data platforms. Large data volumes are now being produced every day from a variety of sources at an unprecedented rate (e.g., health, government, social networks, marketing, and financial). This is brought on by several technical trends, such as the Internet of Things, the growth of cloud computing, and the adoption of smart gadgets (Botta et al., 2016). For example, smart grid systems (Chen et al., 2014), healthcare systems (Kankanhalli et al., 2016), retailing systems like Walmart's (Schmarzo, 2013), government systems (Stoianov et al., 2013), and other systems with multiple connections are supported in the background by robust systems and distributed applications. Before the big data era, businesses could neither keep all of their archives for extended periods nor manage enormous data sets effectively. The storage capacity and administration tools of traditional systems are indeed constrained, and they are also expensive. They lack the scalability, flexibility, and performance required in the context of big data. In actuality, managing big data necessitates

historical growth, advantages, and methods for managing and troubleshooting it.

**Keywords**: Big data, big data management, information society, usage of big data, attributes of big data

enormous resources, innovative approaches, and cutting-edge technologies. More specifically, Big Data necessitates the cleaning, processing, analysis, security, and provision of granular access to vastly expanding data volumes. Organizations are becoming increasingly conscious that data analysis is essential for staying competitive, finding new knowledge, and customized offerings.

Numerous Big Data models, frameworks, and new technologies have been developed as a result of the numerous Big Data projects that have been undertaken globally to increase storage capacity, enable parallel processing, and enable real-time analysis of various heterogeneous sources of data. Additionally, fresh approaches have been created to guarantee data security and privacy. These technologies offer greater flexibility, scalability, and performance when compared to older ones. Additionally, because of sustainable technical improvement, the cost of the majority of hardware stores and processing solutions is continuously declining (Purcell, 2013). Many models, programs, software, hardware, and technologies have been built and proposed to extract knowledge from big data. They work to make sure that big data applications produce more precise and trustworthy outcomes. However, in such a setting, selecting from a variety of technologies can be time-consuming and difficult. Technical compatibility, deployment complexity, cost, efficiency, performance, reliability, support, and security threats are just a few of the numerous factors that need to be taken into account. There are numerous Big Data surveys in the literature, but most of them tend to concentrate more on the techniques and algorithms used to analyze Big Data than on the technology itself (Ali et al., 2016, Chen and Zhang, 2014, Chen et al., 2014). This study, therefore, focuses on the concept of big data, its characteristics, historical development, benefits, troubleshooting, and managing big data.

**Conceptual Framework**

Big data is defined as data that is more diverse, arriving at a faster rate and in larger amounts. Simply put, big data refers to larger, more complicated data volumes, particularly from new data sources. These data sets are so large that conventional data processing technologies just cannot handle them. However, one may use these enormous amounts of data to solve business issues that were previously impossible to solve. Because of more information, big data enables more comprehensive responses. More comprehensive responses increase data confidence, which calls for an entirely different strategy for approaching issues. A definition of big data focuses on three Vs: volume (the amount of data managed), velocity (the rate of incoming data), and variety (the type of data). Sometimes veracity (the quality and accuracy of the incoming data) is also included in the definition.

The phrase "big data" has been in use since the 1990s, and some attribute its inception to John Mashey (Lohr, 2013). Big data often consists of data sets that are too large for frequently used software tools to acquire, curate, manage, and process in a reasonable amount of time (Snijders, 2012). Unstructured data is the primary focus of the big data philosophy, which also includes semi-structured and structured data (Dedic and Stanier, 2017). As of 2012, the "scale" of big data ranged from a few dozen terabytes to many zettabytes of data (Everts, 2016). To extract insights from huge, varied, complicated data sets, a variety of approaches and technologies with novel forms of integration are needed (Ibrahim et al, 2015). Big data describes data sets that are too big or intricate for conventional data-processing application software to handle. Data collection, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy, and data source are just a few of the big data analysis challenges. Volume, diversity, and velocity were the initial three key ideas connected to big data ("The 5 V's of big data", 2016). When handling big data, parallel computing tools are required. Big data uses mathematical analysis, optimization, inductive statistics, and concepts from nonlinear system identification (Billings, 2013) to infer laws (regressions, nonlinear relationships, and causal effects) from large sets of data with low information density ("le Blog ANDSI", 2017) to reveal relationships and dependencies, or to perform predictions of outcomes and behaviors (Les Echos, 2013).

Big data is a term used to describe the enormous, difficult-to-manage amounts of structured and unstructured data that daily saturate enterprises. However, what organizations do with the data matters more than just the type or volume of data; this is what makes a difference. Big data is a large collection of heterogeneous information that comes in ever-increasing volumes and at ever-increasing speeds. There are two types of big data: structured (typically numerical, easily processed, and saved) and unstructured (more free-form, less quantifiable). Big data is the term for the enormous, varied sets of information that multiply at an exponential rate. It includes the quantity of data, the velocity or rate at which it is generated and gathered, and the variety or range of the data points being covered (together referred to as the "three v's" of big data). Data mining is a common source of big data, which can be found in various formats (Segal, 2022). The amount of data generated by internet users is enormous, and only specialized tools and techniques can be utilized to store, comprehend, and make use of it. Big data refers to high-volume, high-velocity, and/or high-variety information assets that necessitate efficient, cutting-edge methods of information processing to improve insight, decision-making, and process automation.

**Characteristics of Big Data**

Big data's primary elements are described as follows in a 2011 McKinsey Global Institute report (Manyika, 2021).

- Techniques for analyzing data, such as A/B testing, machine learning, and natural language processing.
- Big data technologies, like business intelligence, cloud computing, and databases
- Visualization, such as charts, graphs, and other displays of the data

The following qualities are additional ways to characterize big data:

**Volume**: this means the capacity of data created and kept. The value and potential insight of the data, as well as whether it qualifies as big data, are all dependent on its quantity. Big data typically exceeds terabytes and petabytes in size (Sagiroglu, 2013). Data volume is significant. With big data, one needs to process large amounts of unstructured, low-density data. This can be unvalued data from sources like Twitter data feeds, clickstreams from websites or mobile apps, or sensor-enabled hardware. This could amount to tens of terabytes of data for certain firms. Several hundred petabytes might apply to some people, for it to be regarded as big data, the question is; is the amount of data managed growing at an accelerating rate? Must be answered before such data can be regarded as big data.

**Variety:** this implies the type and nature of data. It may also be regarded as the many types of data that are available. Traditional data types were structured and fit neatly in a relational database. With the rise of big data, data comes in new unstructured and semi-structured data types, such as text, audio, and video which require additional preprocessing to derive meaning. The transition from structured to semi-structured or unstructured data, however, presented a problem for the already-in-use techniques and technology. Big data technologies were developed primarily to capture, store, and process the semi-structured and unstructured (variety) data generated at a fast rate of speed and in a massive amount (volume). These Big data technologies such as big data relational database management systems (RDBM) can be explored and used for handling structured data including storage.

**Velocity:** This is the fast rate at which data is received and (perhaps) acted on. Big data is often available in real time. Compared to small data, big data is produced more continually. Two kinds of velocity related to big data are the frequency of generation and the frequency of handling, recording, and publishing (Kitchin, and McArdle, 2016).

**Veracity:** This is the truthfulness or reliability of the data, which refers to the data quality and the data value (Onay and Öztürk, 2018). To be useful in the analysis of it, big data must not only be huge but also trustworthy. An accurate analysis can be hampered by the wide variations in the quality of the data that have been collected.

**Value:** The worth of information that can be achieved by the processing and analysis of large datasets. Value also can be measured by an assessment of the other qualities of big data (IBM Big Data & Analytics Hub, 2021). The value may also represent the profitability of information that is retrieved from the analysis of big data.

**Variability:** The quality of massive data sources, structures, or formats that are constantly changing. Big data may consist of both organized and unstructured data, or it may combine both types of data. The integration of raw data from various sources is possible with big data analysis. Unstructured data conversions into structured data are another aspect of the processing of raw data.

Other possible characteristics of big data according to Kitchin and McArdle (2016) are:

**Exhaustiveness:** Whether or not the complete system is recorded. All of the accessible data from sources may or may not be included in big data.

**Fine-grained and uniquely lexical:** Specifically, the amount of each element's unique data per element that was gathered and whether the element and its properties were correctly indexed or identified.

**Relational:** If the data collected contains common fields that would enable a conjoining, or meta-analysis, of different data sets.

**Extensional:** If adding or modifying new fields in each collected set of data is simple.

**Scalability:** If the big data storage system can quickly grow in size.

**Historical development of Big Data**

Large data sets have their roots in the 1960s and 1970s when the world of data was just getting started with the creation of the first data centers and the relational database. People started to understand how much data, users were producing through Facebook, YouTube, and other online services around 2005. That same year, Hadoop (an open-source framework designed primarily to store and analyze massive data sets) was launched. Due to their ability to make huge data more manageable and more affordable to store, open-source frameworks like Hadoop (and more recently, Spark) were crucial to the development of big data. The amount of big data has multiplied since that time. More products and devices are now online with the advent of the Internet of Things (IoT), which is gathering information on consumer usage trends and product performance. With the development of machine learning, more data has been generated. Big data has gone a long way. The potential uses of big data have been further increased by cloud computing. Additionally, graph databases are gaining importance due to their capacity to show vast volumes of data in a form that facilitates quick and thorough analytics.

**Benefits of Big Data**

Big data may assist with a variety of tasks. A few are listed below:

**Product Development:** They categorize crucial characteristics of previous and present goods and services, analyze the connections between those characteristics and the commercial success of the offerings, and use this information to create predictive models for new goods and services.

**Customer experience**: Big data makes it possible to collect information from phone records, social media, online traffic, and other sources to enhance the engagement process and increase the value given. Reduce customer churn, start sending targeted offers, and take proactive measures to address problems.

**Fraud and compliance:** The security environment and compliance standards are always changing. Big data enables the quick collection of massive amounts of information and the finding of patterns in data that point to fraud.

**Machine learning:** One of the causes for this is data, particularly large data. Big data is readily accessible, allowing for the training of machine learning models. The branch of artificial intelligence (AI) known as machine learning (ML) focuses on creating systems that can learn from the data they are fed and hence enhance their performance.

**Operational efficiency:** Big data analysis and assessment can be used to evaluate production, client feedback and returns, and other aspects to decrease outages and foresee future demands. Big data can also be utilized to make decisions that are better aligned with the demands of the market.

**Promote innovation**: Big data may assist in innovation by examining the connections between people, institutions, things, and processes, and then coming up with new applications for those discoveries. Improve decision-making around monetary and planning factors by utilizing data insights. To deliver innovative products and services, look at trends and what customers want. Implement dynamic pricing.

**Troubleshooting Big Data**

Big data has a lot of potentials, but it also has certain drawbacks, such as;



TIJSRAT
ISSN: 2623-7861

**Huge data volume:** Despite the development of new storage technology, data volumes double in size roughly every day. Organizations could have trouble keeping up with their data and figuring out how to store it efficiently.

**Curability:** merely storing the data is not sufficient. For data to be useful, it must be put to use, and that depends on curation. It takes a lot of effort to create clean data, or data that is pertinent to the client and arranged to allow for insightful analysis. Before using the data, data scientists spend between 50 and 80 percent of their effort filtering and preparing it.

**Big data technology is evolving quickly**: A few years ago, Apache Hadoop was the widely utilized large data management tool. Then, in 2014, Apache Spark was released. The optimum strategy at this time seems to be a blend of the two frameworks. Adapting to big data technology is a constant struggle.

**Inadequate skilled labour**: The lack of skilled labour is one of the largest barriers to realizing the full potential of big data. Making sure big data technologies are incorporated into an organisation's information technology program can help to reduce this risk. Big data solutions and methods enable organizations to examine their skill requirements frequently and proactively detect any potential skill gaps. These can be resolved by utilizing consultancy firms, recruiting additional resources, and training/cross-training existing resources.

**Managing Big Data**

Unstructured or structured big data are two different categories. Information that is already kept up to date by an organisation in databases and spreadsheets is referred to as structured data, and it typically has a quantitative component. Unstructured data is data that is not pre-organised and does not fit into a pre-established model or format. It contains data derived from social media sources that aid organisations in learning more about what people want. Before organisations can utilise big data, they should understand how it moves across a diverse range of sources, platforms, proprietors, and users. The "big data infrastructure," which consists of both structured and unstructured data can be managed in five essential ways which are;

- Developing a big data strategy.
- Recognising the sources of big data.
- Accessing and storing the data.
- Evaluating the data.
- Taking data-driven conclusions

**Developing a big data strategy:** In its broadest definition, a big data strategy refers to a plan developed to help an organisation acquire, store, distribute, and utilise data both inside and outside the organisation. With so much information at hand, a big data strategy opens the door to occupational success. When developing a plan, it is imperative to consider future goals. As a result, big data must be viewed as a valuable corporate asset rather than just an output of submissions.

**Recognising the sources of big data**: Streaming data is generated by the Internet of Things (IoT) and other connected devices that include gadgets, smart automobiles, medical equipment, industrial equipment, and more. This data enters information technology systems. As a huge data comes in, one can examine it, to choose which information should be kept and which needs more in-depth examination. Furthermore, Interactions on sites like Facebook, YouTube, Instagram, and others provide social media data. This comprises enormous amounts of big data that may be used for marketing, sales, and support tasks and comes in the form of photographs, videos, voice, text, and sound. This data frequently takes the form of unstructured or semi-structured data, which presents a special barrier to consumption and interpretation. Also, massive open data sources like the European Union Open Data Portal, and the intelligence-gathering organization, World Factbook, among others provide publicly accessible data. Other big data may originate from operational databases, cloud data sources, distributors, and partners.

**Accessing and storing the data**: Modern computer systems have the power, speed, and adaptability required to quickly retrieve huge volumes and many kinds of big data. Ogranisations need ways to integrate data, create data pipelines, guarantee data quality, provide data governance, store the data, and prepare the data for analysis in addition to reliable access. While some big data may be kept locally in a typical data warehouse, there are also flexible, affordable choices for keeping and managing big data available through cloud solutions, data lakes, etc.

**Evaluating the data:** Organizations can decide whether to deploy all of their big data for analysis using high-performance technologies like grid computing or in-memory analytics. Another strategy is to select the relevant data in advance for analysis. Big data analytics is how businesses derive value and insights from data, regardless of the situation. Big data is increasingly used as fuel for modern advanced analytics projects like artificial intelligence (AI) and machine learning.

**Taking data-driven conclusions:** Data analysis and decisions can only be trusted if the data is well-managed and trustworthy. Organisations must fully utilize the benefits of big data to remain competitive. They must also adopt a data-driven approach, basing choices more on the evidence provided by big data than on guesswork. Being data-driven has several


ISSN: 2623-7861

advantages. Organizations that are data-driven perform better, have more predictable operations and are more productive.

## Conclusion

As a result of realizing how big data supports and facilitates the achievement of an organization's primary goals, huge data sets offer the chance to create novel discoveries in society. To optimize the potential of big data, organizations can also use a knowledge transfer strategy to share expertise. Using this technique can aid in the development of big data capabilities in a more planned and logical manner. In delineating meanings from big data sets, analysts and data scientists have to fully understand the organization's needs and collaboration with the organization is also essential. In the information society, a big data solution includes all types of data, including transactions, master data, reference data, and summary data. To maintain control over the whole data flow, including pre-and post-processing, integration, in-database summarization, and analytical modeling, big data management technology is crucial.

## References

Ali, A. J. Qadir, R. urRasool, R. urRasool, A. Sathiaseelan, A. Zwitter, J. Crowcroft (2016). Big data for development: applications and techniques. Big Data Anal., 1 (2016), p. 2

Billings S.A (2013). Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains. Wiley, 2013

Botta, A. W. de Donato, V. Persico, A. Pescapé (2016) Integration of cloud computing and internet of things: a survey. Future Gener. Comput. Syst., 56 (2016), 684-700

Cappa, Francesco; Oriani, Raffaele; Peruffo, Enzo; McCarthy, Ian (2021). Big Data for Creating and Capturing Value in the Digitalized Environment: Unpacking the Effects of Volume, Variety, and Veracity on Firm Performance. Journal of Product Innovation Management. 38 (1), 49–67.

Chen, M., S. Mao, Y. Liu. (2014). Big data: a survey. Mobile Networks App., 19 (2014), 171-209

Chen, P and C.-Y. Zhang (2014). Data-intensive applications, challenges, techniques, and technologies: a survey on big data. Inf. Sci., 275 (2014), pp. 314-34

Dedić, N.; Stanier, C. (2017). Towards Differentiating Business Intelligence, Big Data, Data Analytics, and Knowledge Discovery. Innovations in Enterprise Information Systems Management and Engineering. Lecture Notes in Business Information Processing. Vol. 285. Berlin; Heidelberg: Springer International Publishing. pp. 114–122. doi:10.1007/978-3-319-58801-8_1

Everts, Sarah (2016). Information Overload. Distillations, 2(2): 26–33.

Ibrahim; Targio Hashem, Abaker; Yaqoob, Ibrar; Badrul Anuar, Nor; Mokhtar, Salimah; Gani, Abdullah; Ullah Khan, Samee (2015). Big data on cloud computing: Review and open research issues. Information Systems. 47: 98–115. doi:10.1016/j.is.2014.07.006.

"le Blog ANDSI » DSI Big Data" (2013). Andsi.fr. Retrieved 8 August 2022

John R. Mashey (1998). "Big Data ... and the Next Wave of InfraStress" (PDF). Slides from an invited talk. Usenix. Retrieved 21 August 2022.

Kankanhalli, A., J. Hahn, S. Tan, G. Gao. (2016). Big data and analytics in healthcare: introduction to the special section. Inf. Syst. Front., 18 (2016), 233-235

Kitchin, Rob; McArdle, Gavin (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. Big Data & Society. 3 (1), 12-17

Kitchin, Rob; McArdle, Gavin (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. Big Data & Society, 3 (1), 205395171663113. doi:10.1177/2053951716631130

Les Echos (2013). "Les Echos – Big Data car Low-Density Data ? La faible densité en information comme facteur discriminant – Archives". Lesechos.fr. Retrieved 8 August, 2022

Lohr, S. (2013). The Origins of Big Data: An Etymological Detective Story. The New York Times. Retrieved 28 September 2022.

Manyika, James; Chui, Michael; Bughin, Jaques; Brown, Brad; Dobbs, Richard; Roxburgh, Charles; Byers, Angela Hung (2011). Big Data: The next frontier for innovation, competition, and productivity"(PDF). McKinsey Global Institute.

"Measuring the Business Value of Big Data | IBM Big Data & Analytics Hub" (2021). Www.ibmbigdatahub.com. Retrieved 20 July 2022

Onay, Ceylan; Öztürk, Elif (2018). "A review of credit scoring research in the age of Big Data". Journal of Financial Regulation and Compliance. 26 (3), 382–405.

Oussous, A. Benjelloun, F. Lahcen, A. and Belfkih, S. (2018). Big Data technologies: A survey. Journal of King Saud University - Computer and Information Sciences, 30(4), 431-448

Purcell, B. (2013). Big Data using cloud computing. Holy Family Univ. J. Technol. Res. (2013)

Sagiroglu, Seref (2013). Big data: A review. International Conference on Collaboration Technologies and Systems (CTS), 42–47.

Schmarzo, B. (2013) Big Data: Understanding How Data Powers Big Business. John Wiley & Sons (2013).

Segal, T. (2022). Big Data. Retrieved from https://www.investopedia.com/terms/b/big-data.asp

Snijders, C.; Matzat, U.; Reips, U.-D. (2012). Big Data: Big gaps of knowledge in the field of the Internet. International Journal of Internet Science. 7: 1–5.

Solnik, Ray (2016). The Time Has Come: Analytics Delivers for IT Operations. Data center Journal. Retrieved on 21 June 2022 from http://www.datacenterjournal.com

Stoianov, N., M. Urueña, M. Niemiec, P. Machnik, G. Maestro. (2013). Integrated security infrastructures for law enforcement agencies. Multimedia Tools App. (2013), pp. 1-16

"The 5 V's of big data" (2021). Watson Health Perspectives. Retrieved on 12 August 2022



TIJSRAT
ISSN: 2623-7861