International Journal of Innovative Research and Scientific Studies, 8(1) 2025, pages: 2610-2617

IJIRSS
International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

What is big data? Building a fuller understanding of the term

Mohammed El-Astal[1*], Mohammad ALLAYMOUN[2], Ahmad Yahia Mustafa[3]

[1]Mass Communication and Public Relations Department, College of Communication and Media Technologies, Gulf University, Sand 26489, Kingdom of Bahrain
[2]Administrative Science Department, College of Administrative and Financial Science, Gulf University, Sanad 26489, Kingdom of Bahrain
[3]Accounting and Financial Science Department, College of Administrative and Financial Science, Gulf University, Sanad 26489, Kingdom of Bahrain

Corresponding author: Mohammed El-Astal (*Email: mohd41@hotmail.com*)

Abstract

A quick review of the literature shows that there are as many definitions of the term ‘big data’ as there are individuals using the term. This paper attempts to come to a fuller understanding of the term. Thirty-three definitions were collected from the literature for analysis using the ‘theme-rheme’ concept. The collected definitions were broken down into basic ideas and condensed into a final definition. The ‘theme-rheme’ formula suggested here was utilized from Halliday [1] systemic functional linguistics (SFL) theory. This formula helps identify the theme or topic of the definition and the part in which the theme was developed—the central ideas or defining features. Based on the analyses conducted, this paper defines big data as a dataset(s) of high volume, variety, velocity, value, and complexity gathered from various real-world sources through devices and virtual-world sources and analyzed through advanced and unconventional platforms and systems to create value. This definition is expected to enhance the understanding of the term ‘big data’ among industries, theorists, researchers, instructors, and students.

Keywords: Big Data Management, Big Data Value, Big Data, Quasi-Big Data, Theme-Rheme.

DOI: 10.53894/ijirss.v8i1.5035
Funding: This study received no specific financial support.
History: Received: 18 January 2025 / **Revised:** 20 February 2025 / **Accepted:** 26 February 2025 / **Published:** 28 February 2025

Competing Interests: The authors declare that they have no competing interests.
Authors’ Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Publisher: Innovative Research Publishing

1. Introduction

The term ‘big data’ (BD) was first coined by NASA scientists in 1997. The scientists used the term to describe extremely large datasets to be stored in a single computer’s memory [2]. With time, it has become ubiquitous in industry and academia [3]. A quick review of the literature on big data reveals that there are as many descriptions of the term as there are individuals using it. This paper aims to collect as many definitions of the term as possible from the literature and then analyze them to develop a broader understanding of it. It is worth mentioning here that this paper is only concerned with big data as a commodity, not a field of study. Big data as a commodity concerns everyone (scientists, researchers, managers, media organizations, companies, governments, etc.), while as a field of study, it only concerns those teaching and learning it.

In the coming pages, we will explain the theoretical concepts used to guide the analysis, followed by the research methodology. Next, the findings will be displayed, analyzed, discussed, and conclusions drawn.

2. Framework for Analysis

Based on this study's inquiry type, the authors will use Halliday [1] theme-rheme concept, which was utilized from Systemic Functional Linguistics (SFL). The theme-rheme concept assists in identifying the part of the definition where the topic or theme is stated and the part where the key ideas are discursively represented (rheme). El-Astal [4] used Halliday's theme-rheme concept to analyze tens of definitions of the term 'curriculum,' trying to build a more comprehensive understanding of the term. El-Astal used the term 'trunk-branch' to refer to Halliday [1] theme-rheme concept. The trunk-branch term is understandable to everyone (in computer science, engineering, social sciences, etc.), but the term 'theme-rheme' can only be comprehensible to those in applied linguistics and discourse studies. Besides using the theme-rheme formula to analyze definitions, El-Astal and El-Youssef [5] used it in an inductive study to develop a definition for the term 'digital communication' From fragmented knowledge.

According to Eggins [6] language can be understood as a semiotic system consisting of three levels: lexico-grammar (wordings), semantics (meanings), and phonology/graphology (sounds/letters). In SFL, language is understood to have three meta-functional meanings: ideational (concerned with conveying ideas and representing the world), interpersonal (related to social interaction and expressing relationships), and textual (related to the organization and coherence of texts). In the context of this study, the focus is on the textual meanings, which are realized through the wordings used in definitions. Eggins identifies two functional components within clauses and sentences: theme and rheme. Halliday and Matthiessen [7] define the theme as the part of the sentence that works as the starting point of the message, indicating what the sentence is about, and Eggins [6] defines the rheme as the part of the sentence where the theme is developed further.

It is essential to mention that this study does not intend to conduct extensive and in-depth analyses based on Systemic Functional Linguistics. Its sole purpose is to investigate the textual meanings present in the texts (definitions) by examining the theme and rheme of the sentences within those texts. In other words, identifying the clause's theme will assist in categorizing and organizing the definitions according to themes or topics, while identifying the rheme will facilitate the exploration of the fundamental ideas, or the 'defining features,' as Clarke and Erickson [8] put it, developed in the portion of the clause where the theme is expanded upon. Together, the theme and rheme contribute to the overall textual meaning of the analyzed definitions.

3. Methodology

This is a theoretical empirical study. It is theoretical as it addresses ontological questions like what big data is and what makes big data, big data. It is empirical, for which tens of definitions of big data were collected from the literature for analysis. All the collected definitions enable theme-rheme analysis as they consist of both or the rheme at least. El-Astal [4] likened the theme to the tree trunk and the rheme to its branches. This theme-rheme or trunk-branch framework, suggested for analyzing definitions, can also be applied when describing concepts and writing definitions.

As previously mentioned, the analysis of the definitions' themes assists in categorizing the collected definitions into specific themes or topics, whereas the analysis of the rhemes helps identify the fundamental ideas within each definition. To elaborate further, the part of the definition that works as the starting point solely represents the theme or topic, and the subsequent part of the definition, where the theme is developed, contains the definitions' main ideas, such as big data characteristics, sources, analysis software, and value. For example, the underlined part of the definition in which Boyd and Crawford [9] described big data as a 'cultural, technological, and scholarly phenomenon' is the theme (topic), and the rest of the definition (see Table 1, item 5) is the rheme where the authors developed the topic referring to the elements on which big data rests. It should be noted that not all definitions found in the literature were analyzed. Definitions that define big data as a field of study were excluded as this study is limited to definitions that see big data as a commodity, and also, many definitions were found repetitive, containing the three Vs only (volume, variety, and velocity). Moreover, we excluded the definitions that cannot be clearly cited.

3.1. The Study Strategy

The study's strategy involved five stages: (a) the authors collected as many definitions as possible from the literature on big data, (b) they divided the definitions into themes and rhemes, (c) they further divided the rhemes (the central part of a definition) into their fundamental ideas, (d) they identified patterns among the fundamental ideas, and finally (e) they condensed the identified patterns into a final working definition.

Finally, the authors' varied backgrounds equipped them with the interdisciplinary knowledge required to carry out this research: the corresponding author has expertise in discourse analysis, public relations, and mass communication; the second author has expertise in computer science; and the third author has expertise in accounting.

4. Findings and Discussion

In this part of the study, the thirty-three definitions collected from the literature will be displayed, analyzed, and discussed. The definitions collected were chronologically displayed in Table 1. As previously mentioned, the theme-rheme analysis would orient analysis. The theme and rheme will be discussed separately in the coming pages.

Table 1.
Definitions Published between 2009 and 2020 in a Chronological Order.

#	Definition
1.	Data whose size forces us to look beyond the tried-and- true methods that are prevalent at that time [10].
2.	Enormous amounts of unstructured data produced by high-performance applications falling in a wide and heterogeneous family of application scenarios [11].
3.	Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis [12].
4.	Datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze [13].
5	Cultural, technological, and scholarly phenomenon that rests on the interplay of technology, analysis, and mythology that provokes extensive utopian and dystopian rhetoric [9].
6.	The data sets and analytical techniques in applications that are so large and complex that they require advanced and unique data storage, management, analysis, and visualization technologies [14].
7.	A big dataset will probably be so large as to not fit on a single hard drive; as a result, it will be stored on several different disks and will be processed on a number of cores [15].
8.	Data that's too big, too fast, or too hard for existing tools to process [16].
9.	Datasets too large to fit in an Excel spreadsheet or be stored on a single machine [17].
10.	A term that describes large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information [18].
11.	Any kind of data source that has at least three shared characteristics: extremely large volumes of data, extremely high velocity of data, and extremely wide variety of data [19].
12.	Collection of data from traditional and digital sources that represents a source for ongoing discovery and analysis." "In defining big data, it's also important to understand the mix of unstructured and multi-structured data that comprises the volume of information [20].
13.	Big data isn't just volume, variety, and velocity, though; it's volume, variety, and velocity at scale [21].
14.	Datasets that are too large for traditional data-processing systems and that therefore require new technologies [22].
15.	Big data can mean big volume, big velocity, or big variety [23].
16.	Big data is a term describing the storage and analysis of large and or complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce and machine learning [24].
17.	Information assets characterized by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value [3].
18.	Set of techniques and technologies that require new forms of integration to uncover large hidden values from large datasets that are diverse, complex, and of a massive scale [25].
19.	Information assets characterized by high volume, velocity, variety, variability with veracity subjected to specific technology and analytical methods for deriving value with virtue [26].
20.	The set of methods and knowledge that are required for the most-recent forms of amalgamation to expose the huge, secreted values from the big datasets which are varied, intricate, and of enormous scale [27].
21.	Large amounts of digital information that companies and governments collect about people and their surrounding environments [28].
22.	Large, diverse, complex, longitudinal, and distributed data sets generated from instruments, sensors, Internet transactions, e-mail, video, click streams, and other digital sources available today and in the future [29].
23.	Massive datasets that are generated through a variety of resources including environmental, sensors, smart devices, electronic medical records, imaging, laboratory studies, and administrative data [30].
24.	A data environment in which scalable architectures support the requirements of analytical and other applications which process, with high velocity, high volume data which may have a variety of data formats and which may include high velocity data acquisition [31].
25.	Big data possess a suite of key traits: volume, velocity and variety (the 3Vs), but also exhaustivity, resolution, indexicality, relationality, extensionality and scalability [32].
26.	The fact that we can now collect and analyze data in ways that was simply impossible even a few years ago [33].
27.	Large collection of multifaceted data sets, which can also be described as being high volume, variety and velocity, making difficult to move and process instantly with the traditional database management systems [34].

#	**Definition**
28.	Large growing data sets that include heterogeneous formats: structured, unstructured and semi-structured data [35].
29.	Extensive datasets, primarily in the characteristics of volume, velocity and/or variety that require a scalable architecture for efficient storage, manipulation, and analysis [36].
30.	Big data is always digital, has a large sample size, and a large volume or variety or velocity of variables that require additional computing power. It can include quantitative, qualitative, observational or interventional data from a wide range of sources (e.g. government, commercial, cohorts) that have been collected for research or other purposes and may include one or several datasets. Specialist skills in computer programming, database management and data science analytics are usually required to analyze big data [37].
31.	It can be concluded that "if data cannot be stored or processed by a common system's capabilities or exceed a common system's capabilities then these data are considered BD" [38].
32.	Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation [39].
33.	A term for massive data sets having large, more varied and complex structure with the difficulties of storing, analyzing and visualizing for further processes or results [40].

4.1. Theme Analysis

A quick look at the starting points (the underlined parts) of the definitions above (Table 1) reveals that big data was described as follows: (a) 'dataset(s)' in twelve (36.36%) items (items 4, 6, 7, 9, 14, 16, 22, 23, 27, 28, 29, and 33); (b) 'data' in seven (21.21%) items (items 1, 2, 8, 10, 26, and 31); (c) 'information assets' in three (9.09%) items (items 17, 19, and 32); and finally (d) as 'technologies and architectures,' 'phenomenon,' 'data source,' 'set of techniques and technologies,' 'set of methods and knowledge,' 'digital information,' 'data environment,' and a 'suite of traits' in one (3.03%) definition only. There was no clear theme in three (9.09%) definitions (items 13, 15, and 30). To conclude, the most frequent theme used to describe big data was 'dataset(s),' followed by 'data,' and 'information assets,' respectively.

4.2. Rheme Analysis

Analyzing the rhemes of the definitions above (the non-underlined parts) assists us in identifying the defining features embedded within each definition. In the following pages, the defining features found in the rhemes of the definitions will be analyzed and discussed.

4.2.1. BD Characteristics

What characterizes big data is an ontological question. The above definitions (Table 1) and the literature show several characteristics—up to sixteen in some publications. Moorthy, et al. [26] portrayed these characteristics as extensions of Vs (referring to volume, variety, velocity, etc.). For O'Leary, IBM's definition of big data suggests that big data could be characterized by all or any of the three Vs (volume, variety, and velocity) [21]. Although big data is not denoted by volume, volume is the central [41] and most prevailing [42] characteristic. This is what the term 'big data' itself suggests. The following were the traits found in the definitions examined here:

4.2.1.1. Volume

Volume is the characteristic of big data that refers to the magnitude [43] or larger amounts of data produced from various sources [21]. Scrutinizing the rhemes of the definitions listed above (Table 1) reveals that the phrase 'high volume' and synonymous words and phrases, such as enormous amounts, massive, extensive, big, too big, large, too large, so large, size beyond the ability, and [size that] exceeds a common system's capabilities, were found in almost all the definitions, except items 1, 5, and 26. None of these definitions explicitly or implicitly mention how big data is measured, whether in gigabytes, terabytes, petabytes, etc. This is also not clear and agreed upon in the literature. For instance, Manyika, et al. [13]; Madden [16]; Schroeck, et al. [44]; Ularu, et al. [45]; O'Leary [21]; Sagiroglu and Sinanc [46]; Ward and Barker [24]; Kitchin and McArdle [32]; Rani [41]; Sivarajah, et al. [47] and Al-Mekhlal and Khwaja [48] claim that big data volumes are reported in terabytes at least; while Gandomi and Haider [43] and Emmanuel and Stanier [31] argue that big data volumes are relative to time and type (structured, unstructured, semi-structured). That's to say, what is deemed to be big data today may not be tomorrow as the storage capacities increase. Interestingly, Kitchin and McArdle [32] views [17] description of big data as a dataset that cannot be stored on a single machine as a trite perception.

4.2.1.2. Variety

Variety refers to data type or structural heterogeneity [43]. Big data can be structured (tabular), semi-structured (e.g., XML documents), or unstructured (like videos, texts, images, etc.). Structured data does not exceed 5% of the existing data [49] and unstructured data ranges between 80 and 90% [2]. A careful examination of the rhemes of the definitions above shows that the word 'variety' and synonymous words like variable, varied, heterogenous, diverse, structured, unstructured, and semi-structured were found in twenty (60.60%) of the definitions (items 2, 3, 10, 11, 12, 13, 15, 17, 18, 19, 22, 23, 24, 25, 27, 28, 29, 30, 32, and 33.

4.2.1.3. Velocity

For Gandomi and Haider [43] velocity refers to the speed at which data is produced and the rate at which that data needs to be analyzed and acted upon. A careful consideration of the rhemes of the thirty-three definitions above tells that the word 'velocity' and synonymous words like fast appeared in fourteen (42.42%) of the definitions (items 3, 8, 10, 11, 13, 15, 17, 19, 24, 25, 27, 29, 30, and 32.

4.2.1.4. Value

Value refers to the countless insights extracted from big data [32]. A careful review of the thirty-three definitions' rhemes indicates that the word 'value' and synonymous words like insight appeared in six (18.18%) of the definitions (items 3, 17, 18, 19, 20, and 32.

4.2.1.5. Complexity

For Gandomi and Haider [43] complexity refers to big data being produced from many sources. Carefully considering the rhemes of the thirty-three definitions above reveals that the word 'complex' and synonymous words like intricate appeared in seven (21.21%) of the definitions (items 6, 10, 16, 18, 20, 22, and 33).

4.2.1.6. Other Traits

Other traits include characteristics other than volume, variety, velocity, value, and complexity—the prevailing traits found in the definitions above. In addition to these five characteristics, Al-Mekhlal and Khwaja [48] found the following eleven characteristics of big data in the literature: variability, veracity, vocabulary, vagueness, venue, exhaustive, fine-grained, visualization, vulnerability, validity, and volatility. Out of these eleven characteristics, three were found in the definitions as follows: (a) veracity and variability, each appeared once in item 19, and (b) exhaustivity, which appeared once in item 25. In addition, five traits (other than all the traits discussed in this paragraph) were found in item 25. These characteristics were resolution, indexicality, relationality, extensionality, and scalability.

Considering all this, the five characteristics found prevailing in the thirty-three definitions analyzed here were volume, variety, velocity, value, and complexity (4 Vs and a C). Volume stood first (30 patterns found), variety came second (20 patterns found), velocity was third (14 patterns found), complexity came fourth (7 patterns found), and value stood fifth (6 patterns found).

4.2.2. BD Sources

A source refers to where big data originates and comes from. This is another ontological question. A cursory look at the rhemes of the thirty-three definitions listed in Table 1 shows that only five (15.15%) definitions referred explicitly or implicitly to the sources from which big data is gathered (items 12, 21, 22, 23, and 30). All these definitions mentioned that the sources are digital, except one (item 12), which indicated that big data comes from traditional and digital sources. Two (items 22 and 23) of the five definitions explicitly referred to digital sources such as sensors (embedded in all types of objects), Internet transactions, e-mails, videos, click streams, smart devices, images, and electronic medical records. The other definitions refer to the entities that usually produce big data, such as companies and governments. In addition to the digital sources referred to in the definitions analyzed, the literature documents other sources such as blogs [24], Websites [32] digital logs, social networks [35] and Internet of Things [48]—known as IoT and refers to devices embedded in all types of physical objects and connected to the Internet.

4.2.3. BD Management

Kaufmann [50] defines big data management (BDM) as the "process of controlling the flows of large-volume, high-velocity, heterogeneous, and/or uncertain data to create value (p. 6)." The management process involves data collection, storage, and analysis. A quick review of the rhemes of the definitions collected for analysis here reveals that sixteen (48.48%) definitions referred to managing data or words that are more or less synonymous such as storing and analyzing, processing, handling, and deriving value (items 4, 6, 7, 8, 10, 14, 16, 17, 19, 24, 26, 27, 29, 30, 31, and 32). One of the definitions (item 16) referred to NoSQL, MapReduce, and machine learning as examples of the tools and techniques used to store and analyze big data. Zheng, et al. [51] indicated that the most common platforms and systems used to analyze big data are Hadoop, Spark, Apache Storm, Lambda architecture, and NoSQL databases.

4.2.4. What Is BDM For?

This is the last central idea found in the rhemes of the definitions analyzed. It refers to the values created from collecting, storing, and analyzing big data. Such values reflect the benefits and purposes of managing data. For governments, businesses, organizations, etc., such values include but are not limited to identifying growth opportunities, building partnerships, expanding their presence in the market, growing revenue, enhancing competitive advantage, managing relationships, and many more. A careful reflection on the rhemes of the definitions above shows that the values behind managing big data were explicitly or implicitly mentioned in eight (24.24%) of the definitions (items 3, 17, 18, 19, 20, 30, 32, and 33). These items (Table 1) referred to value as follows: (a) '[data] technologies and architectures [are] designed to economically extract value (item 3),' (b) 'specific technology and analytical methods [are required] for its [data] transformation into value (item 17),' (c) 'to uncover large hidden values from large datasets (item 18),' (d) specific technology and analytical methods [are required] for deriving value with virtue (item 19),' (e) 'to expose the huge, secreted values from the big datasets (item 20),' (f) '[data] have been collected for research and other purposes (item 30),' (g) '[data] processing that enables enhanced insight, decision making, and process automation (item 32),' and (h) 'storing, analyzing and visualizing [data] for further processes or results (item 33).'

The four central ideas (BD characteristics, sources, management platforms, and values) found in the rhemes of the analyzed definitions address the ontological questions about the nature of big data and what makes it big data. As mentioned above, five prevailing traits of BD (i.e., volume, variety, velocity, value, and complexity) were found in the definitions, although the literature documented many more. Interestingly, most debates in the literature were on these five, particularly on volume and variety. However, this study and all previous studies that tried to define big data Ward and Barker [24]; De Mauro, et al. [3]; Gandomi and Haider [43]; Emmanuel and Stanier [31]; Al-Mekhlal and Khwaja [48] and Yaseen and Obaid [40] could not provide clear, precise answers to questions about whether big data is measured in gigabytes, terabytes, or petabytes; whether the data produced from a single source (not complex enough) is considered big data; and whether unstructured data alone (not varied enough) is seen as big data. Put differently, is a digital dataset whose volume is less than a terabyte, not complex enough, and not varied enough considered big data? This dataset is not traditional as it cannot be managed using traditional platforms and systems, and if it is not big data, what is it? The term 'quasi-big data' fits here.

5. Summary

The findings obtained from the analyses of the definitions (Table 1), employing the theme-rheme analysis, can be summarized as follows:

(1) What is BD? It was seen as a dataset(s), data, information assets, technologies and architectures, phenomenon, data source, set of techniques and technologies, set of methods and knowledge, digital information, data environment, and a suite of traits;

(2) What does it look like? It was characterized as high in volume, variety, velocity, value, and complexity;

(3) Where does it come from? It is gathered from a range of real-world sources through devices (sensors and actuators) and virtual-world sources such as governments and organizations' repositories, websites, social networks, etc.;

(4) How is it managed? It is managed through advanced and unconventional platforms and systems such as Hadoop, Spark, NoSQL databases, etc.; and

(5) What is it managed for? Big data is collected, stored, and analyzed to create values that include, but are not limited to, managing relationships, growing revenues, and proper decision-making.

All these central ideas can be condensed into a working definition of big data as follows: Big data refers to datasets of high volume, variety, velocity, value, and complexity gathered from a range of real-world sources through devices (sensors and actuators) and virtual-world sources (e.g., governments and organizations' repositories, websites, social networks, etc.); and analyzed through advanced and unconventional platforms and systems (e.g., Hadoop, Spark, Apache Storm, etc.) to create value.

This working definition is expected to deepen the understanding of the term 'big data' among industries, theorists, researchers, instructors, and students. However, it would be more insightful if the definition developed here were discussed in a Delphi study with experts in the field. A Delphi study will help us reach a consensus on the ontological questions raised at the end of the previous section above.

References

[1] M. A. K. Halliday, *An introduction to functional grammar*, 1st ed. London: Arnold, 1985.
[2] M. Bydon *et al.*, "Big data defined: A practical review for neurosurgeons," *World Neurosurgery*, vol. 133, pp. e842-e849, 2020. https://doi.org/10.1016/j.wneu.2019.09.092
[3] A. De Mauro, M. Greco, and M. Grimaldi, "What is big data? A consensual definition and a review of key research topics," in *AIP Conference Proceedings*, 2015, vol. 1644, no. 1: American Institute of Physics, pp. 97-104.
[4] M. El-Astal, "What is Curriculum? Building a broader understanding of the term," *Journal of Curriculum and Teaching*, vol. 12, no. 6, pp. 188-196, 2023. https://doi.org/10.5430/jct.v12n6p188
[5] M. El-Astal and H. El-Youssef, "What is digital communication? Developing a conceptual definition of the term," *Forum for Linguistic Studies*, vol. 7, no. 1, pp. 471-479, 2025. https://doi.org/10.30564/fls.v7i1.7766
[6] S. Eggins, *An introduction to systemic functional linguistics*, 2nd ed. New York: Continuum, 2004.
[7] M. A. K. Halliday and C. Matthiessen, *An introduction to functional grammar*, 3rd ed. London: Arnold, 2004.
[8] A. Clarke and G. Erickson, "Self-study: The fifth commonplace," *Australian Journal of Education*, vol. 48, no. 2, pp. 199-211, 2004. https://doi.org/10.1177/000494410404800208
[9] D. Boyd and K. Crawford, "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon," *Information, Communication & Society*, vol. 15, no. 5, pp. 662-679, 2012. https://doi.org/10.1080/1369118X.2012.678878
[10] A. Jacobs, "The pathologies of big data," *Communications of the ACM*, vol. 52, no. 8, pp. 36-44, 2009. https://doi.org/10.1145/1536616.1536632
[11] A. Cuzzocrea, I.-Y. Song, and K. C. Davis, "Analytics over large-scale multidimensional data: The big data revolution!," in *Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP*, 2011, pp. 101-104.
[12] J. Gantz, "Extracting value from Chaos. The IDC iView," Retrieved: http://idcdocserv.om/1142. [Accessed 2011.
[13] J. Manyika *et al.*, "Big data: The next frontier for innovation, competition, and productivity," Retrieved: http://dln.jaipuria.ac.in:8080/jspui/bitstream/123456789/14265/1/mgi_big_data_full_report.pdf, 2001.
[14] H. Chen, R. H. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact," *MIS Quarterly*, vol. 36, no. 4, pp. 1165-1188, 2012. https://doi.org/10.2307/41703503
[15] D. Fisher, R. DeLine, M. Czerwinski, and S. Drucker, "Interactions with big data analytics," *Interactions*, vol. 19, no. 3, pp. 50-59, 2012. https://doi.org/10.1145/2168931.2168943
[16] S. Madden, "From databases to big data," *IEEE Internet Computing*, vol. 16, no. 3, pp. 4-6, 2012. https://doi.org/10.1109/MIC.2012.50

[17] D. Strom, "Big data makes things better. Dice," Retrieved: https://www.dice.com/career-advice/big-data-makes-things-better/, 2012.

[18] TechAmerica Foundation's Federal Big Data Commission, "Demystifying big data: A practical guide to transforming the business of government," Retrieved: http://www.techamerica.org/Docs/fileManager.cfm?f=techamerica-bigdatareport-final.pdf, 2012.

[19] H. Judith, N. Alan, H. Fern, and K. Marcia, *Big data for dummies*. New Jersey: John Wiley & Sons, Inc, 2013.

[20] A. Lisa, *Big data marketing: Engage your customers more effectively and drive value*. Hoboken, NJ: John Wiley & Sons, Inc, 2013.

[21] D. E. O'Leary, "Artificial intelligence and big data," *IEEE Intelligent Systems,* vol. 28, no. 2, pp. 96-99, 2013. https://doi.org/10.1109/MIS.2013.39

[22] F. Provost and T. Fawcett, "Data science and its relationship to big data and data-driven decision making," *Big Data,* vol. 1, no. 1, pp. 51-59, 2013. https://doi.org/10.1089/big.2013.1508

[23] M. Stonebraker and J. Robertson, "Big data is' buzzword du jour;'CS academics' have the best job'," *Communications of the ACM,* vol. 56, no. 9, pp. 10-11, 2013. https://doi.org/10.1145/2500468.2500471

[24] J. S. Ward and A. Barker, "Undefined by data: a survey of big data definitions," *arXiv preprint arXiv:1309.5821,* 2013. https://arxiv.org/abs/1309.5821

[25] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, "The rise of "big data" on cloud computing: Review and open research issues," *Information Systems*, vol. 47, pp. 98-115, 2015. https://doi.org/10.1016/j.is.2014.07.006

[26] J. Moorthy *et al.,* "Big data: Prospects and challenges," *Vikalpa,* vol. 40, no. 1, pp. 74-96, 2015. https://doi.org/10.1177/0256090915575450

[27] V. Padmapriya, J. Amudhavel, V. Gowri, K. Lakshmipriya, S. Vinothini, and K. P. Kumar, "Demystifying challenges, opportunities and issues of Big data frameworks," in *Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015),* 2015, pp. 1-5.

[28] V. Vivekanand and B. Vidyavathi, "Security challenges in big data: Review," *International Journal of Advanced Research in Computer Science,* vol. 6, no. 6, pp. 199-202, 2015. https://doi.org/10.26483/ijarcs.v6i6.2561

[29] W. Wang, X. Zhou, B. Zhang, and J. Mu, "Anomaly detection in big data from UWB radars," *Security and Communication Networks,* vol. 8, no. 14, pp. 2469-2475, 2015. https://doi.org/10.1002/sec.745

[30] A. J. Barton, "Big data," *Journal of Nursing Education,* vol. 55, no. 3, pp. 123-124, 2016. https://doi.org/10.3928/01484834-20160216-01

[31] I. Emmanuel and C. Stanier, "Defining big data," in *Proceedings of the International Conference on Big Data and Advanced Wireless Technologies*, 2016, pp. 1-6.

[32] R. Kitchin and G. McArdle, "What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets," *Big Data & Society,* vol. 3, no. 1, p. 2053951716631130, 2016. https://doi.org/10.1177/2053951716631130

[33] B. Marr, *Big data in practice: How 45 successful companies used big data analytics to deliver extraordinary results*. West Sussex: John Wiley & Sons Ltd, 2016.

[34] S. Grandhi and S. Wibowo, "A multi-criteria group decision making method for selecting big data visualization tools," *Journal of Telecommunication, Electronic and Computer Engineering,* vol. 10, no. 1–8, pp. 67–72, 2018.

[35] A. Oussous, F.-Z. Benjelloun, A. A. Lahcen, and S. Belfkih, "Big Data technologies: A survey," *Journal of King Saud University-Computer and Information Sciences,* vol. 30, no. 4, pp. 431-448, 2018. https://doi.org/10.1016/j.jksuci.2017.06.001

[36] W. L. Chang and N. Grady, "NiST big data interoperability framework: Volume 2, big data taxonomies," 2019. https://doi.org/10.6028/NIST.SP.1500-2r2

[37] C. Vogel, S. Zwolinsky, C. Griffiths, M. Hobbs, E. Henderson, and E. Wilkins, "A Delphi study to build consensus on the definition and use of big data in obesity research," *International Journal of Obesity,* vol. 43, no. 12, pp. 2573-2586, 2019. https://doi.org/10.1038/s41366-018-0313-9

[38] H. B. Alwan and K. R. Ku-Mahamud, "Big data: Definition, characteristics, life cycle, applications, and challenges," in *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 769, no. 1: IOP Publishing, p. 012007.

[39] G. Gartner, "Big data, Gartner," Retrieved: https://www.gartner.com/en/information-technology/glossary/big-data. [Accessed 10-Apr-2024], 2020.

[40] H. K. Yaseen and A. M. Obaid, "Big data: Definition, architecture & applications," *International Journal on Informatics Visualization,* vol. 4, no. 1, pp. 45-51, 2020. https://dx.doi.org/10.30630/joiv.4.1.292

[41] E. S. Rani, "Big data analytics: Prospects, challenges, and barriers," *International Journal of Electrical Engineering,* vol. 9, no. 1, pp. 231–239, 2017.

[42] M. Volk, D. Staegemann, and K. Turowski, "Providing clarity on big data: Discussing its definition and the most relevant data characteristics," in *Proceedings of the 14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2022)*, 2022, pp. 141-148.

[43] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *International Journal of Information Management,* vol. 35, no. 2, pp. 137-144, 2015. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

[44] M. Schroeck, R. Shockley, J. Smart, D. Romero-Morales, and P. Tufano, "Analytics: The real-world use of big data," *IBM Global Business Services,* vol. 12, no. 2012, pp. 1-20, 2012.

[45] E. G. Ularu, F. C. Puican, A. Apostu, and M. Velicanu, "Perspectives on big data and big data analytics," *Database Systems Journal,* vol. 3, no. 4, pp. 3-14, 2012.

[46] S. Sagiroglu and D. Sinanc, "Big data: A review," in *Proceedings of the 2013 International Conference on Collaboration Technologies and Systems (CTS)*, 2013: IEEE, pp. 42-47.

[47] U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of Big Data challenges and analytical methods," *Journal of Business Research,* vol. 70, pp. 263-286, 2017. https://doi.org/10.1016/j.jbusres.2016.08.001

[48] M. Al-Mekhlal and A. A. Khwaja, "A synthesis of big data definition and characteristics," in *Proceedings of the 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, 2019: IEEE, pp. 314-322.

[49] K. Cukier, "Data, data everywhere: A special report on managing information. The Economist," Retrieved: http://www.economist.com/node/15557443, 2010.

[50] M. Kaufmann, "Big data management canvas: a reference model for value creation from data," *Big Data and Cognitive Computing*, vol. 3, no. 1, p. 19, 2019. https://doi.org/10.3390/bdcc3010019

[51] X. Zheng, M. Fu, and M. Chugh, "Big data storage and management in SaaS applications," *Journal of Communications and Information Networks*, vol. 2, no. 3, pp. 18-29, 2017. https://doi.org/10.1007/s41650-017-0031-9