International Journal of Trend in Research and Development, Volume 12(6), ISSN: 2394-9333
www.ijtrd.com

Big Data: An Overview

[1]Matthew N. O. Sadiku, [2]Uwakwe C. Chukwu and [3]Janet O. Sadiku,
[1]Roy G. Perry College of Engineering, Prairie View A&M University, Prairie View, TX, USA
[2]Department of Engineering Technology, South Carolina State University, Orangeburg, SC, USA
[3]Juliana King University, Houston , TX, USA

Abstract: Big data refers to extremely large and complex datasets that exceed the processing capacity of traditional data management systems. It is essentially about leveraging massive amounts of information to uncover patterns, trends, and insights that can inform better decision-making and drive innovation. Companies use big data in their systems to improve operational efficiency, provide better customer service, create personalized marketing campaigns and take other actions that can increase revenue and profits. By following a structured approach, businesses can extract valuable insights from the vast sea of data they possess. Organizations will be required to manage it appropriately for competitive advantage and durability in the modern digital market. This paper provides an overview of big data and its applications.

Keywords: *Big Data, Big Data Analytics, Big Data Applications*

I. INTRODUCTION

Data is the new gold. We produce a massive amount of data each day through every click on the Internet, every bank transaction, every video we watch on YouTube, and every email we send. Every company generates data, whether consciously or unconsciously. Data can be a company's most valuable asset. The data revolution initiated by the dawn of the Internet and further driven by the expansion of mobile technology is vastly considered as a significant innovation especially in developing nations around the world.

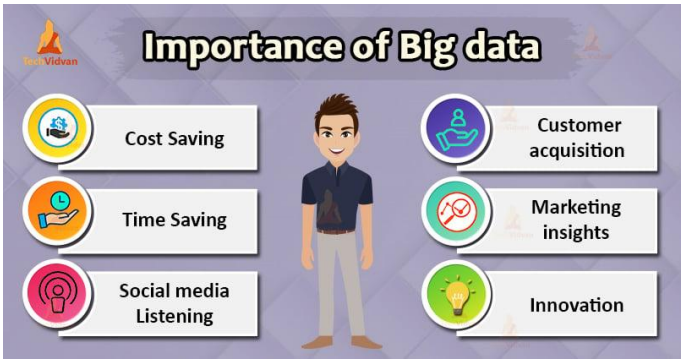

Figure 1: The importance of big data [2].

Big data is a combination of structured, semi-structured, and unstructured data that organizations collect, analyze, and mine for information and insights. These datasets are so huge and complex in volume, velocity, and variety, that traditional data management systems cannot store, process, and analyze them. Big data comes from many sources, including transaction processing systems, customer databases, documents, emails, medical records, Internet clickstream logs, mobile apps, and social networks. With the explosion of devices, sensors, online services, and digital platforms, data is now generated at an unprecedented rate [1]. Organizations that use and manage large data volumes correctly can reap many benefits. For example, big data provides valuable insights into customers that companies can use to refine their marketing, advertising, and promotions to increase customer engagement and conversion rates. An organization can glean important insights, risks, patterns or trends from big data. For example, companies such as Netflix and Procter & Gamble use big data to anticipate customer demand. Figure 1 shows the importance of big data [2].

II. WHAT IS BIG DATA?

Big data applies to data sets of extreme size (e.g. exabytes, zettabytes) which are beyond the capability of the commonly used software tools. It involves situation where very large data sets are big in volume, velocity, veracity, and variability [3]. The data is too big, too fast, or does not fit the regular database architecture. It may require different strategies and tools for profiling, measurement, assessment, and processing. Different components of big data are shown in Figure 2 [4]. The cloud word for big data is shown in Figure 3 [5].

Figure 2: Different components of big data [4].

Figure 3: The cloud word for big data [5].

Big Data is essentially classified into three types [6]:

- *Structured Data:* This is highly organized and is the easiest to work with. Any data that can be stored, accessed, and processed in the form of fixed format is known as a structured data. It

may be stored in tabular format. Due to their nature, it is easy for programs to sort through and collect data. Structured data has quantitative data such as age, contact, address, billing, expenses, credit card numbers, etc. Data that is stored in a relational database management system is an example of structured data.

- *Unstructured Data:* This refers to unorganized data such as video files, log files, audio files, and image files. Any data with unknown form or the structure is classified as unstructured data. Almost everything generated by a computer is unstructured data. It takes a lot of time and effort required to make unstructured data readable. Examples of unstructured data include Metadata, Twitter tweets, and other social media posts.
- *Semi-structured Data:* This falls somewhere between structured data and unstructured data, i.e., both forms of data are present. Semi-structured data can be inherited such as location, time, email address, or device ID stamp.

The different types of big data are depicted in Figure 4 [7].

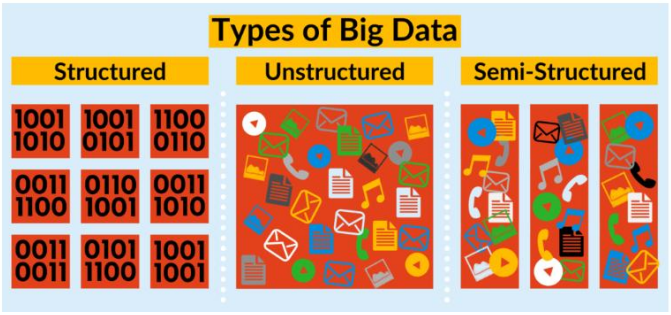

Figure 4: Types of big data [7].

The process of examining big data is often referred to big data analytics. It is an emerging field since massive computing capabilities have been made available by e-infrastructures [8]. Big data analytics is the application of advanced analytic techniques to large, heterogeneous data sets that comprise structured, semi-structured, and unstructured data from many sources with sizes ranging from terabytes to zettabytes.

It enables predictive analytics, which involves using historical data to forecast future outcomes. Analytics include statistical models and other methods that are aimed at creating empirical predictions. Data-driven organizations use analytics to guide decisions at all levels. Several techniques have been proposed for analyzing big data. These include the HACE theorem, cloud computing, Hadoop, and MapReduce [9]. Figure 5 shows big data analytics [10].

Figure 5: Big data analytics [10].

III. CHARACTERISTICS OF BIG DATA

Big data is growing rapidly and expanding in all science and engineering, including physical, biological, and medical services. Different companies use different means to maintain their big data. As shown in Figure 6 [11], big data is characterized by 42 Vs. The first five Vs are volume, velocity, variety, veracity, and value .

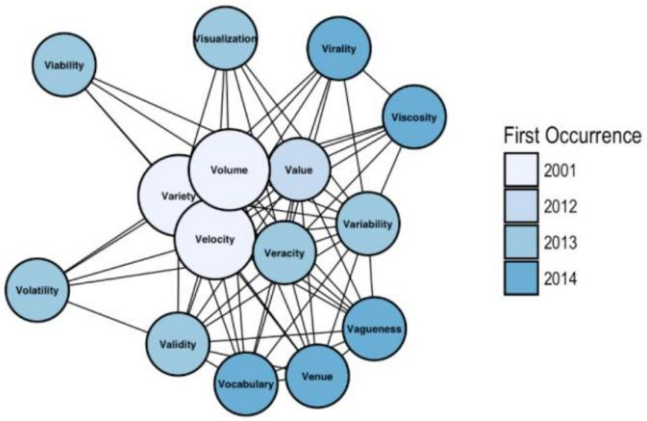

Figure 6: The 42 V's of big data [11].

- *Volume*: This refers to the size of the data being generated both inside and outside organizations and is increasing annually. Some regard big data as data over one petabyte in volume.
- *Velocity*: This depicts the unprecedented speed at which data are generated by Internet users, mobile users, social media, etc. Data are generated and processed in a fast way to extract useful, relevant information. Big data could be analyzed in real time, and it has movement and velocity.
- *Variety*: This refers to the data types since big data may originate from heterogeneous sources and is in different formats (e.g., videos, images, audio, text, logs). BD comprises of structured, semi-structured or unstructured data.
- *Veracity*: By this, we mean the truthfulness of data, i.e. weather the data comes from a reputable, trustworthy, authentic, and accountable source. It suggests the inconsistency in the quality of different sources of big data. The data may not be 100% correct.
- *Value*: This is the most important aspect of the big data. It is the desired outcome of big data processing. It refers to the process of discovering hidden values from large datasets. It denotes the value derived from the analysis of the existing data. If one cannot extract some business value from the data, there is no use managing and storing it.

On this basis, small data can be regarded as having low volume, low velocity, low variety, low veracity, and low value. Additional five Vs has been added [11]:

- *Validity:* This refers to the accuracy and correctness of data. It also indicates how up to date it is.
- *Viability:* This identifies the relevancy of data for each use case. Relevancy of data is required to maintain the desired and accurate outcome through analytical and predictive measures.
- *Volatility:* Since data are generated and change at a rapid rate, volatility determines how quickly data change.

- *Vulnerability:* The vulnerability of data is essential because privacy and security are of utmost importance for personal data.
- *Visualization:* Data needs to be presented unambiguously and attractively to the user. Proper visualization of large and complex clinical reports helps in finding valuable insights.

Instead of the 10V's above, some suggest the following 5V's: Venue, Variability, Vocabulary, Vagueness, and Validity) [12].

Industries that benefit from big data include the healthcare, financial, airline, travel, restaurants, automobile, sports, agriculture, and hospitality industries. Big data technologies are playing an essential role in farming: machines are equipped with sensors that measure data in their environment. The analysis of both structured and unstructured data is crucial in the shipping industry to gain insights into customer behavior, improve operational efficiency, and make informed business decisions.

IV. APPLICATIONS OF BIG DATA

The potential applications for big data are essentially unlimited, and it is already being used across multiple industries. Common applications include the following [4,13-16]:

- *Healthcare:* The healthcare industry can combine numerous data sources internally, such as electronic health records, patient wearable devices, and staffing data, and externally, including insurance records and disease studies, to optimize both provider and patient experiences. Medical researchers use big data to identify disease signs and risk factors. Doctors use it to help diagnose illnesses and medical conditions in patients. In addition, a combination of data from electronic health records, social media sites, the web, and other sources gives healthcare organizations and government agencies up-to-date information on infectious disease threats and outbreaks. For example, big data facilitated in the storage mechanism of the vaccine, since they were kept and stored within precise temperature range.
- *Education:* The educational system is one of the main civilization pillars. Its development can characterize the advancement level of any society. Big data has played a vital role in restructuring the educational system. It enabled educational institutes and professionals to personalize the educational experience for students. Big data can analyze, find correlation between the data, highlight patterns, provide insights and predict for the ultimate teaching-learning process. Hence, educators and professionals in the educational field will provide intelligent decisions to enhance the educational regime. For example, big data can improve the learning process, by optimizing the selection, of the prior teaching techniques and newly proposed ones to meet the student actual needs and interests.
- *Government:* Government digital archiving rates and data generation are on the rise. Government offices can potentially collect data from many different sources, such as DMV records, traffic data, police/firefighter data, public school records, and more. This can drive efficiencies in many different ways, such as detecting driver trends for optimized intersection management and better resource allocation in schools. Governments can also post data publicly, allowing for improved transparency to bolster public trust. Government agencies can leverage big data insights in inventive ways. They can leverage big data and analytics to unlock key information, and improve transparency and efficiency in public management. The use of big data in government and public sector is illustrated in Figure 7 [16].
- *Finance:* Financial institutions leverage big data for fraud detection and risk management. By analyzing transaction patterns and detecting anomalies, banks can prevent fraudulent activities in real time. Big Data is revolutionizing the finance industry by providing real-time insights into market trends and customer behavior. In the financial sector, big data analytics helps in fraud detection, risk management, customer segmentation, personalized banking services, and trading analytics.
- *Manufacturing*: Big Data is being used in the manufacturing industry to improve efficiency, reduce costs, and optimize supply chain operations. Big data applications in manufacturing include predictive maintenance, quality control, and supply chain optimization. By monitoring machinery and equipment data, manufacturers can predict failures and schedule maintenance proactively, reducing downtime and improving operational efficiency.
- *Transportation:* Millions of citizens use public roads every day, whether driving or walking. Many factors contribute to safety on the road, such as the state of the roads, police officers, vehicle safety, and weather conditions. With these factors in play, it is almost impossible to control everything that might lead to an accident. Big data allows governments to oversee the transportation sector to ensure safer roads. Route optimization and fleet management are driven by big data. Analyzing traffic patterns and fleet data helps optimize routes, improve fuel efficiency, and schedule maintenance, leading to cost savings and improved service delivery. For example, big data plays a vital role in predicting the cause effect relation between the driving restriction policies and traffic congestion.
- *Fraud Detection:* One of the most feared challenges of running a business is encountering financial frauds and claims. This was an alarming global problem among organizations before the advent of big data. However, with the emergence of big data, companies can now detect, prevent, and also eliminate any fraudulent risks. Data analysts utilize artificial intelligence and machine learning algorithms to find abnormalities and transaction trends. These irregularities in transaction patterns show that something is out of place or that there is a mismatch, providing us with hints regarding potential frauds. By spotting fraud before they cause problems, a company may provide superior customer service, avoid losses, and stay compliant.
- *Accounting:* The accounting industry is using big data, especially in auditing. Big data is incredibly valuable for accounting firms that need to sell their services; providing accurate and actionable information to clients is a great way to boost firm value. At the moment, accounting firms typically use "audit sampling" to detect issues or trends in transactions or invoices. However, big data analytics

can excel at identifying exceptions and outliers within a larger trend. Accounting firms can then focus their efforts on those exceptions for further analysis. For example, big data sets can allow accounting firms to aggregate performance metrics across an entire industry and present them to a client, pointing out specific reasons the competition may be outperforming the client rather than relying on outdated methods such as ratios or guesswork. Even better, big data can allow accounting professionals to look at the big picture of a particular industry and see shifts in consumer behavior or trends.

- *Marketing:* Big data is notable in marketing due to the constant "datafication" of everyday consumers of the Internet, in which all forms of data are tracked. The datafication of consumers can be defined as quantifying many of or all human behaviors for the purpose of marketing. The increasingly digital world of rapid datafication makes this idea relevant to marketing because the amount of data constantly grows exponentially. Big data in marketing is a highly lucrative tool that can be used for large corporations, its value being as a result of the possibility of predicting significant trends, interests, or statistical outcomes in a consumer-based manner. Big data provides customer behavior pattern spotting for marketers, since all human actions are being quantified into readable numbers for marketers to analyze and use for their research.

Figure 7: Use of big data in government and public sector [16].

V. BENEFITS

Organizations that use and manage large data volumes correctly can reap many benefits. Unlike traditional data management solutions, big data technologies and tools are made to help you deal with large and complex datasets to extract value from them. Big data can be used to identify solvable problems, such as improving healthcare or tackling poverty in a certain area. Other benefits include the following [18,19]:

- *Better Decision-making*: Big data applications and analytics are vital in proposing ultimate strategic decisions. Big data is the key element to becoming a data-driven organization. When you can manage and analyze your big data, you can discover patterns and unlock insights that improve and drive better operational and strategic decisions. An organization can glean important insights, risks, patterns or trends from big data. Large data sets are meant to be comprehensive and encompass as much information as the organization needs to make better decisions. Big data insights let business leaders quickly make data-driven decisions that impact their organizations.
- *Better Insights:* When organizations have more data, they are able to derive better insights. With better insights, organizations can make data-driven decisions with more reliable projections and predictions. Big data that covers market trends and consumer habits gives an organization the important insights it needs to meet the demands of its intended audiences. Product development decisions, in particular, benefit from this type of insight.
- *Better Customer Experience:* Big data has made an unblemished customer experience more relevant and workable. Combining and analyzing structured data sources together with unstructured ones provides you with more useful insights for consumer understanding, personalization, and ways to optimize experience to better meet consumer needs and expectations. Big data allows organizations to build customer profiles through a combination of customer sales data, industry demographic data, and related data such as social media activity and marketing campaign engagement.
- *Cost Savings:* Big data can be used to pinpoint ways businesses can enhance operational efficiency. For example, analysis of big data on a company's energy use can help it be more efficient.
- *Higher Efficiency:* Every company generates data. Using big data analytics tools and capabilities allows you to process data faster and generate insights that can help you determine areas where you can reduce costs, save time, and increase your overall efficiency.
- *Competitive Advantage*: In today's competitive landscape, organizations that harness the power of big data analytics gain a significant advantage. This competitive advantage enables organizations to stay ahead of the curve and drive sustainable growth. By making data-driven decisions, businesses can stay ahead of the curve and adapt to changing market dynamics quickly.
- *Poverty Eradication:* There is a lot of poverty in the world that many governments have tried to eradicate for many years. Big data gives governments the necessary tools to uncover better and innovative ideas on how to reduce poverty levels across the globe. This data makes it easier to identify the areas with urgent needs and how to meet those needs.

VI. CHALLENGES

While big data has many advantages, it does present some challenges that organizations must be ready to tackle when collecting, managing, and taking action on such an enormous amount of data. Big data is big and complex, making it difficult to manage. Big data technology is changing at a rapid pace. Although the term “big data” has been around for some time now, there is still quite a lot of confusion about what it actually means. Other challenges include the following [12,18-20]:

- *Data Privacy:* Data privacy is a critical issue in big data projects, as large amounts of personal data are often collected and analyzed. Organizations must ensure that they comply with data privacy regulations and take

appropriate measures to protect sensitive data. The big data we now generate contains a lot of information about our personal lives, much of which we have a right to keep private. Increasingly, we are asked to strike a balance between the amount of personal data we divulge, and the convenience that big data-powered apps and services offer.

- *Data Security:* Data security is an important consideration in big data projects, as large amounts of data can be a target for cyber attacks. Big data contains valuable business and customer information, making big data stores high-value targets for attackers. Since these datasets are varied and complex, it can be harder to implement comprehensive strategies and policies to protect them.
- *Data Discrimination:* When everything is known, will it become acceptable to discriminate against people based on data we have on their lives? We already use credit scoring to decide who can borrow money, and insurance is heavily data-driven. We can expect to be analyzed and assessed in greater detail, and care must be taken that this is not done in a way that contributes to making life more difficult for those who already have fewer resources and access to information.
- *Data Growth:* Big data, by nature, is changing and increasing exponentially. Without a solid infrastructure in place that can handle your processing, storage, network, and security needs, it can become extremely difficult to manage.
- *Data Quality*: Ensuring data quality is a critical component of big data projects. Poor data quality can lead to inaccurate analysis and incorrect decisions. Data quality directly impacts the quality of decision-making, data analytics, and planning strategies. Raw data is messy and can be difficult to curate. Having big data does not guarantee results unless the data is accurate, relevant, and properly organized for analysis. This can slow down reporting, but if not addressed, you can end up with misleading results and worthless insights.
- *Data Storage:* The collected data then needs to be stored in a way that it can be easily accessed and analyzed later. Big data needs big storage, whether in the cloud, on-premises, or both. Data must also be stored in whatever form required. It also needs to be processed and made available in real time. Increasingly, companies are turning to cloud solutions to take advantage of the unlimited compute and scalability.
- *Data Governance:* This involves establishing policies and procedures for managing and protecting data. It includes defining data standards, ensuring data quality, and compliance with data privacy regulations.
- *Data Integration*: Big data projects often involve integrating data from multiple sources, which can be a complex and challenging. Data integration requires careful planning and consideration of data formats, schemas, and structures. Big data allows you to integrate automated, real-time data streaming with advanced data analytics. However, the process of integrating sets of big data is complicated, particularly when data variety and velocity are factors. Big data collects terabytes, and sometimes even petabytes, of raw data from many sources that must be received, processed, and transformed into the format that business users and analysts need to start analyzing it. Integrating disparate data sources and making data accessible for business users is complex, but vital, if you hope to realize any value from your big data.
- *Accessibility:* Among the main challenges in managing big data systems is making the data accessible to data scientists and analysts, especially in distributed environments that include a mix of different platforms and data stores. To help analysts find relevant data, data management and analytics teams are increasingly building data catalogs that incorporate metadata management and data lineage functions.
- *Skill Shortage:* One of the biggest obstacles to benefiting from your investment in big data is not having enough staff with the necessary skills to analyze your data. Deploying and managing big data systems also requires new skills compared to the ones that database administrators and developers focused on relational software typically possess. Data scientists, data analysts, and data engineers are in short supply. Lack of big data skills and experience with advanced data tools is one of the primary barriers to realizing value from big data environments.
- *Regulation:* Standardizing your approach will allow you to manage costs and leverage resources. To ensure that they comply with the laws that regulate big data, businesses need to carefully manage the process of collecting it. Controls must be put in place to identify regulated data and prevent unauthorized employees and other people from accessing it. Big data contains a lot of sensitive data and information, making it a tricky task to continuously ensure data processing and storage meet data privacy and regulatory requirements, such as data localization and data residency laws.
- *Scalability:* Scalability is a key consideration in big data projects, as the amount of data being processed and analyzed can quickly grow beyond the capacity of traditional systems. Organizations must select tools and platforms that can scale to meet their needs. Cloud platforms like AWS, Azure, and Google Cloud offer scalable solutions for big data storage and processing.

CONCLUSION

The amount and availability of data is growing rapidly, spurred on by digital technology advancements, such as connectivity, mobility, the Internet of things (IoT), and artificial intelligence (AI). Big data refers to the huge volume of both structured and unstructured data that is being generated around the world and holds humongous information. It describes large and diverse datasets that are huge in volume and also rapidly grow in size over time. Companies that act wisely on the insights gained from big data analytics are better poised to make the most of available opportunities as compared to those that do not apply big data solutions. The future of big data is filled with exciting possibilities, including applications in fields such as agriculture, transportation, and energy. More information about big data can be found in the books in [21-26] and the following related journals:

- *Journal of Big Data*
- *Big Data and Cognitive Computing*

References

[1] "What is big data?" August 2025, https://www.geeksforgeeks.org/data-engineering/what-is-big-data/

[2] "Why big data – Benefits and importance of big data," https://techvidvan.com/tutorials/why-big-data/

[3] M. N. O. Sadiku, M. Tembely, and S.M. Musa, "Big data: An introduction for engineers," *Journal of Scientific and Engineering Research,* vol. 3, no. 2, 2016, pp. 106-108.
[4] "Big data: What it is and why it matters?" August 2024, https://www.inventateq.com/top-stories/big-data-what-it-is-and-why-it-matters/
[5] L. Rembert, "How accounting teams can leverage big data," https://tdwi.org/articles/2020/03/03/adv-all-how-accounting-teams-can-leverage-big-data.aspx
[6] "The complete overview of big data," https://intellipaat.com/blog/tutorial/hadoop-tutorial/big-data-overview/
[7] R. Allen, "Types of big data | Understanding & Interacting with key types (2024)," https://investguiding-com.custommapposter.com/article/types-of-big-data-understanding-amp-interacting-with-key-types
[8] P. Baumann et al., "Big data analytics for earth sciences: The earthserver approach," *International Journal of Digital Earth,* vol. 19, no. 1, 2016, pp.3-29.
[9] X. Wu et al., "Knowledge engineering with big data," *IEEE Intelligent Systems*, September/October 2015, pp.46-55.
[10] "Comprehensive guide to big data analysis," May 2024, https://www.sprinkledata.com/blogs/comprehensive-guide-to-big-data-analysis
[11] "The 42 V's of big data and data science," https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html
[12] P. K. D. Pramanik, S. Pal, and M. Mukhopadhyay, "Healthcare big data: A comprehensive overview," in N. Bouchemal (ed.), *Intelligent Systems for Healthcare Management and Delivery*. IGI Global, chapter 4, 2019, pp. 72-100.
[13] "Big data," https://botpenguin.com/glossary/big-data
[14] M. Kour, "A deep dive into big data fundamentals and uses," June 2024, https://www.applify.co/blog/what-is-big-data
[15] "Big data," *Wikipedia*, the free encyclopedia, https://en.wikipedia.org/wiki/Big_data
[16] A. Subramanian, "Big data analytics in government: How the public sector leverages data insights," November 2024, https://www.datasciencecentral.com/big-data-analytics-in-government-how-the-public-sector-leverages-data-insights/
[17] Z. A. Al-Sai et al., "Explore big data analytics applications and opportunities: A review," *Big Data and Cognitive Computing*, vol. 6, no. 4, 2022.
[18] C. Hashemi-Pour, "Big data," March 2024, https://www.techtarget.com/searchdatamanagement/definition/big-data
[19] "What is big data?" https://cloud.google.com/learn/what-is-big-data
[20] B. Marr, "What is big data?" https://bernardmarr.com/what-is-big-data/
[21] M. N. O. Sadiku, U. C. Chukwu, and P. O. Adebo, *Big Data and Its Applications.* Moldova, Europe: Lambert Academic Publishing, 2024.
[22] A. E. Hassanien et al. (eds.), *Big Data in Complex Systems: Challenges and Opportunities*. Springer, 2015.
[23] N. Crepalde, *Big Data on Kubernetes: A Practical Guide to Building Efficient and Scalable Data Solutions.* Packt Publishing, 2024.
[24] J. S. Cook and R. S. Segall (eds.), *Handbook of Research on Big Data Storage and Visualization Techniques.* IGI Global, 2018.
[25] L. M. Goyal et al. (eds.), *Big Data Processing Using Spark in Cloud.* Springer, 2018.
[26] S. Govindappa, *Ultimate Big Data Analytics with Apache Hadoop: Master Big Data Analytics with Apache Hadoop Using Apache Spark, Hive, and Python (English Edition).* Orange Education Pvt. Ltd, 2024.

ABOUT THE AUTHORS

Matthew N. O. Sadiku is a professor emeritus in the Department of Electrical and Computer Engineering at Prairie View A&M University, Prairie View, Texas. He is the author of several books and papers. His areas of research interest include computational electromagnetics and computer networks. He is a Life fellow of IEEE.

Uwakwe C. Chukwu is an associate professor in the Department of Industrial & Electrical Engineering Technology of South Carolina State University. He has published several books and papers. His research interests are power systems, smart grid, V2G, energy scavenging, renewable energies, and microgrids.

Janet O. Sadiku holds bachelor degree in Nursing Science in 1980 at the University of Ife, now known as Obafemi Awolowo University, Nigeria and doctoral degree from Juliana King University, Houston, TX in December 2023. She has worked as a nurse, educator, and church minister in Nigeria, United Kingdom, Canada, and United States. She is a co-author of some papers and books.