An Overview of Big Data Concepts, Methods, and Analytics: Challenges, Issues, and Opportunities

Mahshad Mahmoudian
Department of Computer Engineering,
Najafabad Branch, Islamic Azad University,
Najafabad, Iran.
mah.mahmoudian@ieee.org

S. Mohammadali Zanjani*
Smart Microgrid Research Center,
Najafabad Branch, Islamic Azad University,
Najafabad, Iran.
sma_zanjani@pel.iaun.ac.ir

Hossein Shahinzadeh
Department of Electrical Engineering
Amirkabir University of Technology
Tehran, Iran
h.s.shahinzadeh@ieee.org

Yasin Kabalci
Department of Electrical Engineering
Nigde Ömer Halisdemir University
Nigde, Turkey
yasinkabalci@ohu.edu.tr

Ersan Kabalci
Department of Electrical Engineering
Nevsehir Haci Bektas Veli University
Nevsehir, Turkey
kabalci@nevsehir.edu.tr

Farshad Ebrahimi
Department of Electrical and Computer Engineering
University of Houston
Houston, TX 77004, USA
febrahimi@uh.edu

***Abstract*— In recent years, data generation is increasing on a large scale and fast pace, and the development of Internet applications, mobile applications, and network-connected sensors has also increased widely. These applications and extensive internet connections continuously produce a large volume of data, with a wide diversity and different structures, which is called big data. At the same time, technologies related to big data are also developing. The rapid growth of cloud computing and the Internet of Things (IoT) is accelerating the dramatic growth of data generation. Sensors around the world are collecting and transmitting data that will be stored and processed in the cloud, and the era of big data is coming. In this article, first, an overview of big data and the definitions of its features-are explained, and then the applications of big data in different fields are examined and the challenges facing it are discussed. Finally, technologies related to big data in the field of big data analysis, data storage technologies, and visualization tools are proposed and cloud computing, IoT, and data center are examined as new technologies that are closely related to big data. The main goal of this article is to provide a comprehensive overview of big data and examine and explain various aspects of its applications and implementation.**

Keywords*— *Big Data, Data mining, Social networking, Big Data analytics, Decision making, Information technology, Internet of Things, Cloud computing.

I. INTRODUCTION

"Big data" is a relatively new topic in the field of information technology. There are a lot of researchers working on research and studies in this sector right now, and at the same time, a lot of corporations have gotten interested in it for a variety of reasons [1]. As a result of the considerable applications offered by big data analysis, a variety of businesses and fields of study, most notably those in the fields of power distribution, healthcare, social sciences, insurance, and finance, as well as governmental institutions, have also begun to utilize it [2]. The analysis of large amounts of data has become more important in modern research as well as in modern business. These data are generated as a result of online transactions, emails, movies, music, photos, click streams, logs, postings, search queries, medical records, interactions on social networks, scientific data, sensors, mobile phones, and the programs that run on them [3]. Big data is stored in databases that grow incrementally and contain a large volume of information, making storage, management, sharing, analysis, and data visualization complex tasks that require software tools and complex databases. Throughout the course of the previous two decades, there has been a significant expansion of data in a variety of domains [4]. According to a report that was published by the International Data Corporation (IDC) in 2011, the total volume of data that was produced and copied in the world was 1.8 zettabytes (1.8*1024 exabytes) [5]. This figure has since increased to 40 zettabytes and is projected to reach 175 zettabytes by the year 2025. The progression of big data's expansion from 2010 to its anticipated level of development in 2025 is seen in Figure 1.

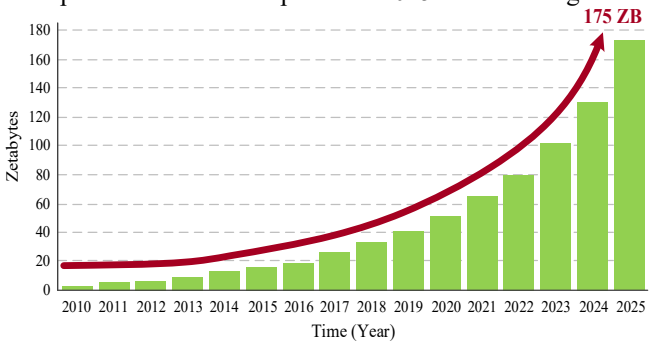

Fig. 1. Annual Growth Rate of Global Data [6]

According to this report, the volume of data has grown more than 20 times in less than a decade, and this figure will double every two years in the near future. With the growth of global data, the phrase "big data" has evolved to characterize such huge collections. Big data is a very high volume of semi-structured and unstructured data that needs more quick analysis compared to other typical datasets and their associated procedures [7]. Additionally, big data creates new opportunities to discover new values and helps to gain a deeper understanding of hidden data values, and of course, comes with new challenges such as how to efficiently manage and organize such data [8-9]. The generation of data has been simpler as a result of advancements in information technology; nowadays, big data originates from the everyday activities of individuals, particularly in connection to the services provided by internet companies. For instance, Google analyzes hundreds of petabytes of data, Facebook generates over ten petabytes of new material each month, and on YouTube, an average of three hundred and fifty hours of video are posted every minute. In addition, the rapid expansion of cloud computing and the Internet of Things has contributed to an increase in the volume of data. Computing in the cloud offers a standardized method for storing and accessing the digital assets of an organization [10]. As part of the Internet of Things, sensors located all over the world are gathering and transferring data that is then saved and processed in the cloud. This volume of data creates many issues and challenges in storing and retrieving heterogeneous and massive datasets, which require hardware and software infrastructure and new technologies to manage and leverage them. In this article, we will review big data, its challenges, and related technologies.

* Assistant Professor, Department of Electrical Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran.

979-8-3503-0198-4/23/$31.00 ©2023 IEEE

First, the definition of big data, its features, and its applications in various fields are explained. Then, the challenges of big data in different areas such as data storage, data visualization, data analysis, data privacy, performance and scalability will be discussed. Finally, the technologies related to big data in the field of data analysis, data storage, and data virtualization, as well as the connection of big data with cloud computing, the Internet of Things, and data centers will be discussed [11-12].

II. Overview of Big Data

The term big data refers to a rapidly growing collection of massive and heterogeneous data in structured, unstructured, and semi-structured formats. Due to their complex nature, big data require powerful technologies and advanced algorithms for management and analysis, and traditional business tools are not effective for dealing with big data [13]. The definition of big data is a topic on which different people disagree. Big data, in general, is a group of data that cannot be comprehended, gathered, managed, and processed simultaneously using conventional hardware/software tools and information technologies. Because of the importance of the topic, technology companies, researchers, and data analysts have different definitions of big data, which will be discussed further below [14].

A. Definition of Big Data Characteristics

Big data refers to data assets that are both enormous and complicated, and which need analysis in order to comprehend and get information from them [15]. In 2010, Apache Hadoop defined big data as "A dataset that has a high volume, velocity, or variety, and traditional methods are limited in their ability to efficiently analyze it." Based on this definition, in May 2011, McKinsey & Company (a global consulting organization) introduced big data as the "next frontier of innovation, competition, and productivity." The National Institute of Standards and Technology (NIST) defines big data as "data sets that have such high volume, velocity, or variety that traditional methods for efficient analysis are limited." This definition focuses on the technological aspect of big data. Most data scientists and big data experts define big data with three main characteristics (known as the "3Vs").

* *Volume*: The dataset that conforms to the big data standard is constantly changing and increasing over time. In big data, there is a large amount of data with sizes ranging from terabytes to zettabytes.

* *Velocity*: Big data is characterized by the rapid generation of data, which, in turn, necessitates the rapid processing of that data in order to derive useful insights. The term "velocity" alludes to the real-time aspect of big data, and in order to make the most of the potential benefits of big data for businesses, it is necessary to gather, analyze, and use the data in a prompt and efficient manner.

* *Variety*: Data comes in various types, including structured data such as database data, semi-structured data such as XML data, and unstructured data such as sound, images, videos, web pages, text, etc [16].

However, others, including IDC, which is one of the most influential leaders in big data and its research fields, have different opinions. In 2011, IDC defined big data as follows: "Big data technologies introduce a new generation of technologies and architectures designed to extract value economically from very large volumes of data with a wide range of diversity, received, discovered, or analyzed at high speeds." With this definition, the characteristics of big data can be defined in the form of 4V, meaning volume (large volume), variety (different methods), velocity (fast production), and value (high value but low density), which is widely recognized. The 4V characteristics of big data are shown in Figure 2 [17].

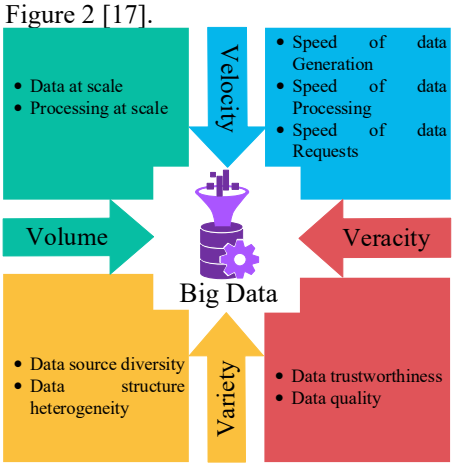

Fig. 2. 4V characteristics of big data

B. Applications of Big Data

There are numerous applications for big data, some of which are illustrated in Figure 3 [18].

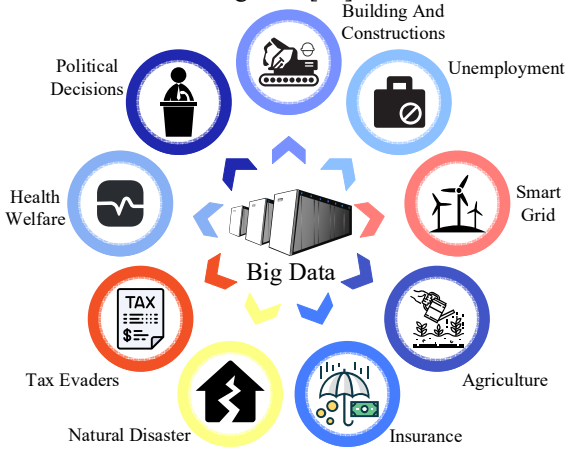

Fig. 3. Applications of Big Data

a) Fraud Detection and Control

In business operations, various types of fraudulent claims or fake data exist, and identifying and controlling these data and fraud in transactions is one of the most important applications of big data. In most cases, fraud is discovered after a long period of its occurrence when data is lost, and in this case, only its effects can be reduced or policies can be implemented to prevent its recurrence. Big data-based platforms can examine and analyze transactions and business operations in real-time and detect inappropriate behavior from a user by examining large-scale patterns for all transactions and deals, thereby changing the way fraud and fake data are detected [19].

b) Call Center Data Analysis

Analyzing call center data is one of the useful applications of big data. In current processes, there are no solutions for processing customer data in the call center, and the information and knowledge that a call center can provide is ignored or presented with delay. Big data-based solutions in call centers can identify recurring problems and behavioral patterns of customers and employees by receiving and processing call content, and help improve organizational performance and increase customer satisfaction [20].

c) Social network analysis

One of the most important applications of big data directly related to users is the analysis of user activity on social networks. Users are widely active on social networks and

record a lot of information about their activities on a daily basis, from expressing interest in a company's products on Facebook to expressing opinions or complaints about other products in the form of a message on Twitter. Social network data can provide useful real-time information about market responses to products and campaigns, enabling companies to prepare and offer their products in line with market and customer opinions [21].

d) Financial data analysis

Big data analysis can also be used for financial analysis and forecasting. For example, big data is used in tools for predicting stock market trends to support decision-making in this area [22].

e) Agriculture

Biotechnology centers use sensor data in agriculture to increase crop productivity. They study and simulate plant reactions in different environmental conditions so that plants can adjust to the environment based on this information. In addition, big data can be used to select the type of crop to be cultivated [23].

III. Big Data Challenges

Data analysis of big data provides attractive and valuable opportunities. However, researchers and experts in this field face multiple challenges when exploring big data and extracting knowledge and value from it. These problems exist at various levels of storage, data display, analysis, lifecycle management, reducing redundancy and compression, etc. In addition, issues related to privacy and confidentiality are especially obstacles and challenges that must be overcome in distributed applications of big data. Some key obstacles and challenges that must be overcome in developing big data applications are described below [24]. Some of the existing challenges for big data are shown in Figure 4, which we will explain below.

Fig. 4. Some of the challenges for Big Data

A. Storage

Today's hard drives have a capacity of terabytes, while the data generated in big data is far beyond that and is increasing exponentially, reaching exabytes. Traditional data management and analysis systems are based on Relational Data Base Management Systems (RDBMS) and are only suitable for managing structured data and are unable to store and process such large amounts of data that are semi-structured and unstructured [25]. The solution to this problem is to use distributed file systems and NoSQL databases, which are designed to manage unstructured data on a large scale.

B. Data Display

The types of datasets, their structures, the meanings of the datasets, their organizations, the granularity of the datasets, and their levels of accessibility can vary widely. The purpose of displaying data is to give it meaning so that it may be interpreted meaningfully by both users and computers. The value of the primary data, on the other hand, is diminished by an unsuitable display of the data, which may even impede an effective study of the data [26]. Displaying data effectively requires taking into account not only the structure, class, and data type, but also the requirements and preferences of the end user.

C. Redundancy reduction and data compression

In most cases, there is a significant amount of duplicate information present in datasets. If the data's potential value is not diminished in the process of decreasing this duplication and compressing the data, the system's overall indirect costs will be reduced to a greater extent than would have been the case otherwise. For instance, the majority of the information that is produced by sensor networks has a significant amount of redundancy. This redundancy may be eliminated, and the quantity of the resulting data can be reduced [27].

D. Data Lifecycle Management

Sensors and ubiquitous computing systems are creating data at an unprecedented rate and scale, and present storage systems are not capable of sustaining such enormous volumes of data. This is in contrast to the comparatively modest advances that storage systems have been making in comparison. The worth of the data is taken into consideration during the process of managing the data lifecycle to determine which data should be kept and which should be discarded.

E. Analysis

The big data analysis process, which has a large volume of unstructured or semi-structured heterogeneous data, requires a lot of resources and time. To address this issue, distributed processing architectures are used, where data is divided into smaller sections and made available for processing by the number of computers in the network, and finally, the processed data are combined [28].

F. Confidentiality of Information

One of the important challenges of big data is the confidentiality and preservation of information. Most big data providers and owners cannot efficiently maintain and analyze their large datasets due to their limited capacity. They rely on data analysis experts and tools that increase potential security risks. Therefore, maintaining the confidentiality of information is a major issue and challenge in big data [29].

G. Energy Management

With the increasing volume of data and demand for analysis, processing, storage, and transfer of big data, inevitably more electrical energy will be consumed for these purposes. Therefore, mechanisms for controlling and managing energy consumption levels for big data must be established.

H. Scalability

The big data analytics system must support current and future datasets. Therefore, the analytics algorithm should be capable of processing increasingly complex datasets that are expanding over time [30].

I. Collaboration

Big data analytics is interdisciplinary research that calls for the cooperation of specialists from diverse domains in order to fully utilize the potential of big data. To enable scientists and engineers from diverse professions to access various types of data and fully utilize their knowledge to

interact with one another in order to achieve the analytics objectives, a comprehensive big data network architecture must be developed [31].

IV. Big Data Management Tools and Technologies

Big Data management involves organizing and utilizing a large amount of data. Assuring data quality and accessibility for use in Business Intelligence (BI) projects and Big Data analytics is the aim of big data management. For analytics, storage, and visualization, a variety of Big Data management solutions are employed, some of which are briefly covered in this section [32]. In addition, the relevant technologies related to Big Data will also be discussed in this section.

A. Data Analysis

* *Hadoop*: An open-source software framework that provides scalable solutions for solving problems with big data on a set of computers. Hadoop is made up of two key components: the MapReduce (MR) framework and the Hadoop Data File System (HDFS). The data storage source for MR is HDFS, a distributed file system created by Google based on the Data File System and running on commercial hardware (DFS).

* *Hive*: An open-source data warehouse for querying and analyzing large sets of data stored in Hadoop files. It features a SQL-like user interface for querying data held in multiple Hadoop-integrated databases and storage systems. It was initially introduced and developed by Facebook and is now offered as an open-source tool.

* *Pig*: An advanced environment for developing MapReduce applications using Hadoop. Pig Latin, a high-level descriptive language that can express huge data gathering and analysis tasks in MR programming, is the language utilized in this platform.

* *Platform*: It is a tool for analyzing and discovering big data. It is a platform that automatically takes user queries to the target and allows users to interact visually with vast amounts of data at a petabyte scale in the shortest possible time. In fact, it creates an abstraction layer that anyone can use to simplify and organize their datasets.

* *Rapidminer*: It is software that offers an integrated platform for business analysis, predictive analytics, text mining, machine learning, and data mining. Rapidminer covers all data mining operations, including data preparation, validation, visualization, and result optimization. It is used for both the development of commercial applications as well as research and education [33].

B. Storage Technologies

For the administration of huge volumes of data, methods of data storage that are both efficient and effective are necessary. This is due to the fact that the size and volume of the data continue to rise at an alarming rate. Both the virtualization of storage and the compression of data have been major contributors to the total development that has been made in this sector.

* *HBase*: The columnar, non-relational database known as HBase is supported by the Apache Hadoop File System (HDFS), which acts as the basis for the database. Users are able to get read and write access in real-time to vast volumes of data that come from a broad range of sources and organizational forms with the help of HBase, which is a free and open-source database system that anyone may download and install on their own computers.

* *SkyTree*: A high-performance platform for machine learning and data analysis that specifically focuses on managing and analyzing big data.

* *Non-Relational Databases*: A strategy for managing and constructing databases that are appropriate for use with vast amounts of data in contexts that are dispersed is referred to as a non-relational database, which is also known as NoSQL. The most widely used of these databases is Apache Cassandra, which was initially developed for Facebook in 2008 before being made available under an open-source license. Additional examples of these databases are SimpleDB, Google BigTable, MongoDB, and Voldemort. Large organizations like Netflix, LinkedIn, and Twitter employ one or more of these databases [34].

C. Data visualization tools

There are numerous open-source data visualization tools, some of which are mentioned below [35].

* *R*: A free and open-source programming language and development environment designed for visualizing and graphically representing data based on graphic and statistical computations. R is a programming language that is often utilized in the statistical software development and data analysis fields.

* *Tableau*: A tool used for visualizing results in the form of charts, maps, graphs, and other graphics. There is also the possibility of connecting Hadoop and Tableau, and interaction between these two products.

* *Infogram*: This tool allows for the easy selection of a wide range of ready-made visual templates. Additionally, there are additional templates such as map charts and videos in this software, and the ability to share created models are also provided.

* *ChartBlocks*: A free online tool that provides the ability to visualize databases and extensive pages without the need for any complex code.

* *Tangle*: This visual tool provides capabilities beyond data visualization and allows designers and developers to design programs interactively for a better understanding of data relationships.

D. Big data-related technologies

Some significant technologies that are closely connected to big data are covered in this section.

a) Cloud computing

Cloud computing has a close relationship with big data. Figure 5 depicts the main components of cloud computing. The term "cloud computing" refers to a type of technology that is capable of storing significant amounts of data. The main goal of cloud computing is to use centralized management of computational resources and capacities to provide various applications by sharing resources in a unified manner and making these applications accessible to users in a transparent and efficient manner [36].

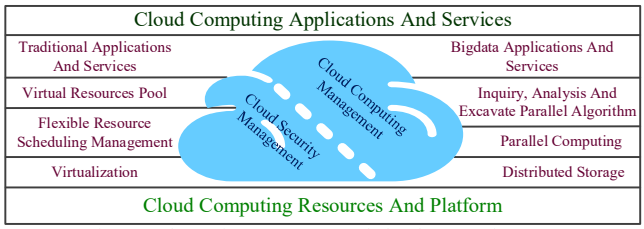

Fig. 5. The main components of cloud computing [37]

The proliferation of cloud-based computing services has opened up new avenues for managing massive datasets. What this means is that the advent of big data serves to hasten the maturation of cloud computing. Cloud computing and its offshoot, cloud storage, have made it possible to effectively handle massive data sets. Big data acquisition and analysis in the cloud may be sped up with the use of parallel computing power [38].

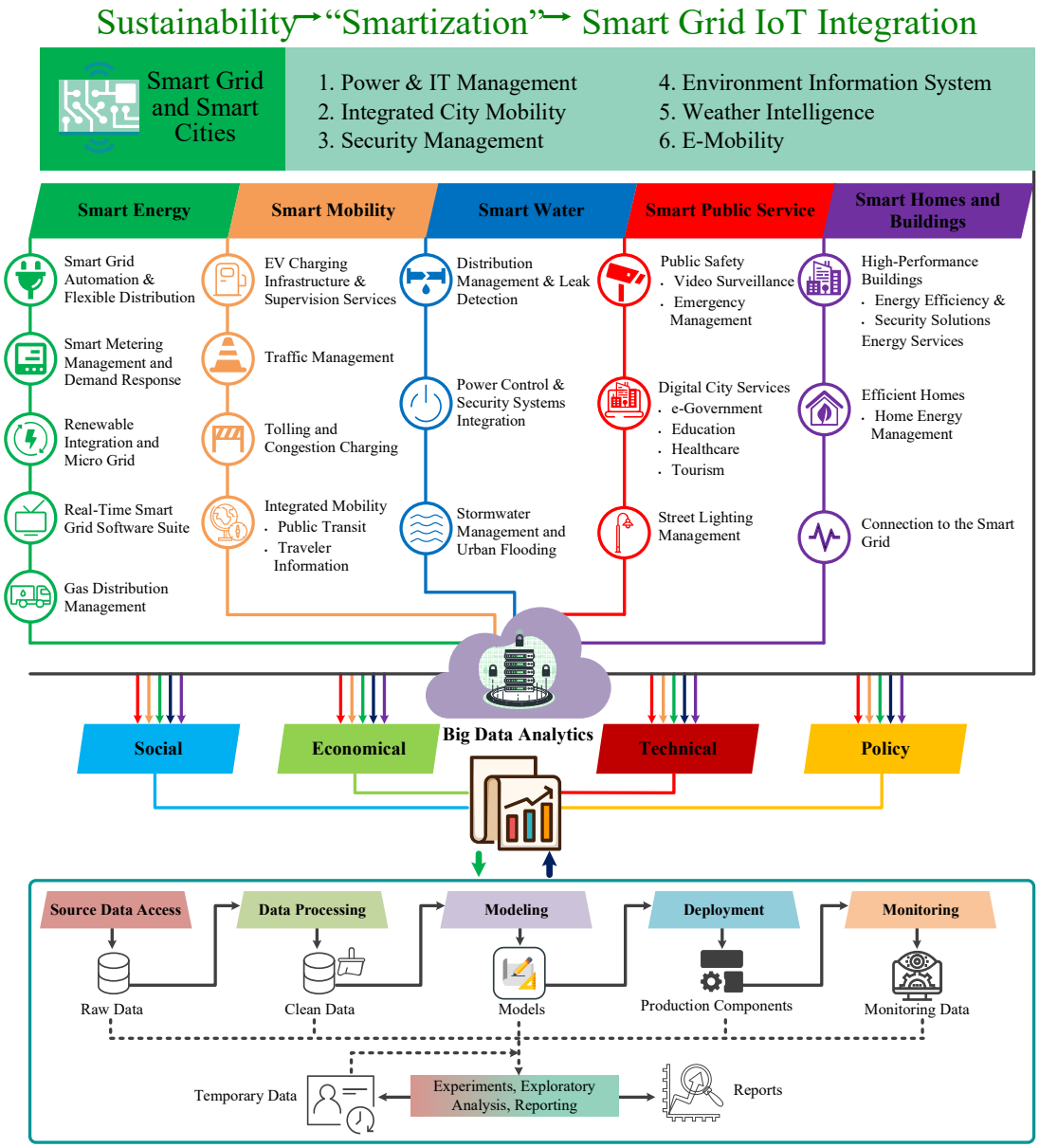

Fig. 6. A visual representation of equipment and methods for collecting data in the smart grid and smart cities Based on IoT

b) Internet of Things

A large number of network sensors are installed in various devices around the world, collecting various types of data such as network communication data, environmental data, geographic data, astronomical data, and more. Since the information resources collected in the Internet of Things are from various environments, the big data generated by the Internet of Things has different characteristics compared to general big data [39]. Heterogeneity, diversity, non-structuredness, redundancy, and rapid growth are some of the characteristics of big data generated by the Internet of Things. Figure 6 shows the equipment and methods for collecting information in the smart grid and smart cities Based on the Internet of Things platform.

An Intel report has mentioned three characteristics of big data in the Internet of Things:

* Abundant terminals that produce large amounts of data.
* The Internet of Things often produces semi-structured or unstructured data.
* Only after analysis is data from the Internet of Things valuable.

c) Data Center

Data centers are not just centralized storage facilities for data by one organization; they also have additional responsibilities. Data collection, data processing, data organization, data value optimization, and operations are all performed in a data center. A data center organizes and maintains a large volume of data in accordance with its primary goal and development route. Big data's rise has presented data centers with both possibilities and obstacles for expansion [40-42].

V. Conclusion

In this review article, big data and related concepts including definitions, features, challenges, and leading issues and technologies in data analysis, storage, and data visualization have been discussed. Additionally, the Internet of Things, cloud computing, and data centers as technologies closely related to big data that contribute to its progress and development have been explained. Despite significant advancements in the field of big data, compared to other technologies, there are still significant shortcomings in this area and many issues remain to be resolved. Standardization, technologies related to big data storage, real-time performance, big data management, search, exploration and analysis of big data, the development of big data applications in various fields, data security, and mechanisms related to big data security and privacy are issues that researchers must address and provide appropriate solutions.

REFERENCES

[1] Escobar, C. A., McGovern, M. E., & Morales-Menendez, R. (2021). Quality 4.0: a review of big data challenges in manufacturing. *Journal of Intelligent Manufacturing*, *32*, 2319-2334.

[2] Kong, L., Liu, Z., & Wu, J. (2020). A systematic review of big data-based urban sustainability research: State-of-the-science and future directions. *Journal of Cleaner Production*, *273*, 123142.

[3] Nti, I. K., Quarcoo, J. A., Aning, J., & Fosu, G. K. (2022). A mini-review of machine learning in big data analytics: Applications, challenges, and prospects. *Big Data Mining and Analytics*, *5*(2), 81-97.

[4] Hajjaji, Y., Boulila, W., Farah, I. R., Romdhani, I., & Hussain, A. (2021). Big data and IoT-based applications in smart environments: A systematic review. *Computer Science Review*, *39*, 100318.

[5] Mallappallil, M., Sabu, J., Gruessner, A., & Salifu, M. (2020). A review of big data and medical research. *SAGE open medicine*, *8*, 2050312120934839.

[6] Janev, V. (2021). Semantic intelligence in big data applications. *Smart Connected World: Technologies and Applications Shaping the Future*, 71-89.

[7] Lampropoulos, G. (2023). Artificial Intelligence, Big Data, and Machine Learning in Industry 4.0. In *Encyclopedia of Data Science and Machine Learning* (pp. 2101-2109). IGI Global.

[8] Moradi, J., Shahinzadeh, H., Nafisi, H., Marzband, M., & Gharehpetian, G. B. (2019, December). Attributes of big data analytics for data-driven decision making in cyber-physical power systems. In *2020 14th international conference on protection and automation of power systems (IPAPS)* (pp. 83-92). IEEE.

[9] Kabalcı, Y., & Ali, M. (2019, June). Energy Internet: A Novel Vision for Next-Generation Smart Grid Communications. In *2019 1st Global Power, Energy and Communication Conference (GPECOM)* (pp. 96-100). IEEE.

[10] Pramanik, S., & Bandyopadhyay, S. K. (2023). Analysis of Big Data. In *Encyclopedia of Data Science and Machine Learning* (pp. 97-115). IGI Global.

[11] Pal, K. (2023). A Review of Big Data Analytics for the Internet of Things Applications in Supply Chain Management. *Applied AI and Multimedia Technologies for Smart Manufacturing and CPS Applications*, 221-245.

[12] Escobar, C. A., McGovern, M. E., & Morales-Menendez, R. (2021). Quality 4.0: a review of big data challenges in manufacturing. *Journal of Intelligent Manufacturing*, *32*, 2319-2334.

[13] Shahinzadeh, H., Zanjani, S. M., Moradi, J., Fayaz-dastgerdi, M. H., Yaïci, W., & Benbouzid, M. (2022, October). The Transition Toward Merging Big Data Analytics, IoT, and Artificial Intelligence with Blockchain in Transactive Energy Markets. In *2022 Global Energy Conference (GEC)* (pp. 241-246). IEEE.

[14] Sharma, A., Singh, G., & Rehman, S. (2020). A review of big data challenges and preserving privacy in big data. *Advances in Data and Information Sciences: Proceedings of ICDIS 2019*, 57-65.

[15] Ahmed, H., & Ismail, M. A. (2020). Towards a novel framework for automatic big data detection. *IEEE Access*, *8*, 186304-186322.

[16] Zanjani, S. M., Zanjani, S. H., Shahinzadeh, H., Rezaei, Z., Kaviani-Baghbaderani, B., & Moradi, J. (2022, November). Big Data Analytics in IoT with the Approach of Storage and Processing in Blockchain. In *2022 6th Iranian Conference on Advances in Enterprise Architecture (ICAEA)* (pp. 1-6). IEEE.

[17] Reda, O., Sassi, I., Zellou, A., & Anter, S. (2020, September). Towards a data quality assessment in big data. In *Proceedings of the 13th International Conference on Intelligent Systems: Theories and Applications* (pp. 1-6).

[18] Shi, Y. (2022). Advances in big data analytics. *Adv Big Data Anal*.

[19] Keskar, V., Yadav, J., & Kumar, A. (2022). Perspective of anomaly detection in big data for data quality improvement. *Materials Today: Proceedings*, *51*, 532-537.

[20] Wang, J., Yang, Y., Wang, T., Sherratt, R. S., & Zhang, J. (2020). Big data service architecture: a survey. *Journal of Internet Technology*, *21*(2), 393-405.

[21] Ghani, N. A., Hamid, S., Hashem, I. A. T., & Ahmed, E. (2019). Social media big data analytics: A survey. *Computers in Human Behavior*, *101*, 417-428.

[22] Ren, S. (2022). Optimization of enterprise financial management and decision-making systems based on big data. *Journal of Mathematics*, *2022*, 1-11.

[23] Cravero, A., Pardo, S., Sepúlveda, S., & Muñoz, L. (2022). Challenges to Use Machine Learning in Agricultural Big Data: A Systematic Literature Review. *Agronomy*, *12*(3), 748.

[24] Demirol, D., Das, R., & Hanbay, D. (2022). A key review on security and privacy of big data: issues, challenges, and future research directions. *Signal, Image and Video Processing*, 1-9.

[25] Mazumdar, S., Seybold, D., Kritikos, K., & Verginadis, Y. (2019). A survey on data storage and placement methodologies for cloud-big data ecosystem. *Journal of Big Data*, *6*(1), 1-37.

[26] Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research*, *98*, 261-276.

[27] Dai, H. N., Wang, H., Xu, G., Wan, J., & Imran, M. (2020). Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies. *Enterprise Information Systems*, *14*(9-10), 1279-1303.

[28] ur Rehman, M. H., Yaqoob, I., Salah, K., Imran, M., Jayaraman, P. P., & Perera, C. (2019). The role of big data analytics in industrial Internet of Things. *Future Generation Computer Systems*, *99*, 247-259.

[29] Talha, M., Abou El Kalam, A., & Elmarzouqi, N. (2019). Big data: Trade-off between data quality and data security. *Procedia Computer Science*, *151*, 916-922.

[30] Gupta, R., Jadav, N. K., Nair, A., Tanwar, S., & Shahinzadeh, H. (2022, September). Blockchain and AI-based Secure Onion Routing Framework for Data Dissemination in IoT Environment Underlying 6G Networks. In *2022 Sixth International Conference on Smart Cities, Internet of Things and Applications (SCIoT)* (pp. 1-6). IEEE.

[31] Han, H., & Trimi, S. (2022). Towards a data science platform for improving SME collaboration through Industry 4.0 technologies. *Technological Forecasting and Social Change*, *174*, 121242.

[32] Rao, T. R., Mitra, P., Bhatt, R., & Goswami, A. (2019). The big data system, components, tools, and technologies: a survey. *Knowledge and Information Systems*, *60*, 1165-1245.

[33] Pavithra, N., & Manasa, C. M. (2021, December). Big Data Analytics Tools: A Comparative Study. In *2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)* (pp. 1-6). IEEE.

[34] Ikegwu, A. C., Nweke, H. F., Anikwe, C. V., Alo, U. R., & Okonkwo, O. R. (2022). Big data analytics for data-driven industry: a review of data sources, tools, challenges, solutions, and research directions. *Cluster Computing*, *25*(5), 3343-3387.

[35] Archana Acharya, T., & Veda Upasan, P. (2020). A stitch in time saves nine: a Big Data analytics perspective. In *Smart Technologies in Data Science and Communication: Proceedings of SMART-DSC 2019* (pp. 227-243). Springer Singapore.

[36] Abualigah, L., Diabat, A., & Elaziz, M. A. (2021). Intelligent workflow scheduling for Big Data applications in IoT cloud computing environments. *Cluster Computing*, *24*(4), 2957-2976.

[37] Srinivas, J., Das, A. K., & Rodrigues, J. J. (2020). 2PBDC: privacy-preserving bigdata collection in cloud environment. *The Journal of Supercomputing*, *76*, 4772-4801.

[38] Bagherzadeh, L., Shahinzadeh, H., Shayeghi, H., Dejamkhooy, A., Bayindir, R., & Iranpour, M. (2020, July). Integration of cloud computing and IoT (CloudIoT) in smart grids: Benefits, challenges, and solutions. In *2020 International Conference on Computational Intelligence for Smart Power System and Sustainable Energy (CISPSSE)* (pp. 1-8). IEEE.

[39] Kabalci, Y., Kabalci, E., Padmanaban, S., Holm-Nielsen, J. B., & Blaabjerg, F. (2019). Internet of things applications as energy internet in smart grids and smart environments. *Electronics*, *8*(9), 972.

[40] Shahinzadeh, H., Mirhedayati, A. S., Shaneh, M., Nafisi, H., Gharehpetian, G. B., & Moradi, J. (2020, December). Role of joint 5G-IoT framework for smart grid interoperability enhancement. In *2020 15th International Conference on Protection and Automation of Power Systems (IPAPS)* (pp. 12-18). IEEE.

[41] Shahinzadeh, H., Moradi, J., Gharehpetian, G. B., Nafisi, H., & Abedi, M. (2019, January). IoT architecture for smart grids. In *2019 International Conference on Protection and Automation of Power System (IPAPS)* (pp. 22-30). IEEE.

[42] Moazzami, M., Sheini-Shahvand, N., Kabalci, E., Shahinzadeh, H., Kabalci, Y., & Gharehpetian, G. B. (2021, October). Internet of things architecture for intelligent transportation systems in a smart city. In *2021 3rd Global Power, Energy and Communication Conference (GPECOM)* (pp. 285-290). IEEE.