Communications in Mathematics and Applications
Vol. 13, No. 1, pp. 331–349, 2022
ISSN 0975-8607 (online); 0976-5905 (print)
Published by RGN Publications
DOI: 10.26713/cma.v13i1.1974

http://www.rgnpublications.com

Review Article

Challenges and Obstacles Facing Data in the Big Data Environment

Hameed Abdullah Aloufi* and Adel Habib Abdulaziz

Department of Information Science, College of Arts and Humanities, King Abdulaziz University, Jeddah, Saudi Arabia
***Corresponding author:** alofeihumaid@gmail.com

Received: December 5, 2021 **Accepted:** February 21, 2022

Abstract. Big data is dubbed "today's digital oil" and the "new raw resource of the twenty-first century". BD is synonymous with the future of innovation, competition, and productivity. It can produce and find corporate value by analyzing data in ways that older methodologies could not. Regardless of its benefits, the development of big data continues to encounter various barriers, the most important of which are security and privacy concerns. As a result, this study is motivated by the need to address and evaluate big data challenges. Thus, by comparing and contrasting big data difficulties with available and potential solutions, users, developers, and businesses can find pertinent and timely responses to specific dangers, resulting in the best possible big data-based services. The objective of this essay is to highlight the inherent challenges of big data and some essential strategies for overcoming them.The purpose of this article was to extract and analyze significant works in order to contribute to the corpus of literature by emphasizing many critical difficulties in the big data domain and throwing light on how these challenges affect a range of domains, including users, sites, and business. Many issues such as data privacy, information sharing, failures in big data technologies and infrastructure, poor data quality management, managers and policymakers' inability to learn and adapt, the absence of government policies and plans, the lack of successfully implemented big data analytics projects, and the lack of human experience are among the most frequently mentioned issues. Obstacles also exist in the areas of real-time data collection, as well as real-time data processing and visualization, among others.By combining previously identified solutions, this research addressed these concerns. The consequences for both researchers and practitioners have been discussed. Aiming to help scholars gain a comprehensive understanding of these issues and confirm the approaches used to address them, this study's theoretical focus is broad. This study uses tried-and-true solutions to overcome these obstacles. Business and individuals using big data analytics systems will benefit from these solutions.

Keywords. Big data "BD", Challenges, Solutions, Issues, Threats, Information Technology "IT", Data Security

Mathematics Subject Classification (2020). 68P15

1. Introduction

Nowadays, the growth of networks and information has resulted in an explosion of data. Because of the diversity of data generation sources such as *Internet of Things* (IoT) sensors, *Radio-Frequency Identification* (RFID), social media, *Global Positioning System* (GPS), log files, images, videos, texts, and the computerization of all life aspects, the flood of data is increasing exponentially [9]. Due to the exponential growth of data gathered through numerous methods, big data is critical in a wide variety of fields, including business, search engines, and education [2]. The massive volume of data makes processing unfeasible using conventional database and software technologies. It necessitates the execution of massively parallel software over tens, hundreds, or even thousands of servers [1]. The term "big data" is believed to have originated with Internet search companies who had to query very huge, distributed data that was loosely structured. With regards to the notion of big data, IBM summarized the concept using the 5Vs: volume, velocity, variety, value, and veracity ([15], [12]).

The lifetime of big data is shown in Figure 1. At the moment, big data-driven analysis will provide the manufacturing industry with more perfect benefits, thanks to the mutual assistance of associated new technologies in the context of Industry 4.0. The data analysis method attempts to increase decision-making transparency [8]. Decision-making based on big data analysis optimizes the function of the entire manufacturing system in accordance with the enterprise's internal structure. It makes efficient use of production resources to ensure that its economic benefits are maximized [11].

Regrettably, the development of big data continues to face numerous obstacles, with security and privacy concerns ranking among the most serious. For instance, a retail website monitors our shopping history; a social network records our thoughts and private images; and a bank monitors our private property information [2]. It appears as though all of your data and actions are in the hands of others, which is really disturbing. Indeed, sensitive data can readily be leaked if appropriate protection is not in place throughout its lifecycle, which includes data collection, storage and administration, transportation, analysis, and disposal [2].

The challenges of big data appear in several domains. For example, the optimal use of big data sets in healthcare informatics faces numerous challenges, including the volume of data, the speed with which data is generated, the variety of data types, the veracity of data, and the privacy of patients' medical information [6]. When *Wireless Sensor Networks* (WSN) networks grow in size and deployment space, the amount of data collected and processed grows exponentially, necessitating efficient processing and making traditional data processing methods difficult to use [5].

The full potential of smart cities using Big Data requires overcoming a number of obstacles, including the availability of Big Data analytics tools, the accuracy and representation of real-time analytics, privacy concerns, and reliable infrastructure [10]. Big Data security shares numerous characteristics with traditional information system security (where data are structured). However, Big Data security necessitates the development of more powerful tools, proper procedures, and innovative technology for performing rapid data analysis [4]. Additionally, it necessitates a new security management paradigm that manages both internal data (data generated by an organization's internal systems and procedures) and external data (e.g., data collected from other firms or external web sites) [4]. As a result of the aforementioned difficulties, it is necessary to highlight these difficulties and demonstrate how previous research has addressed this conundrum.

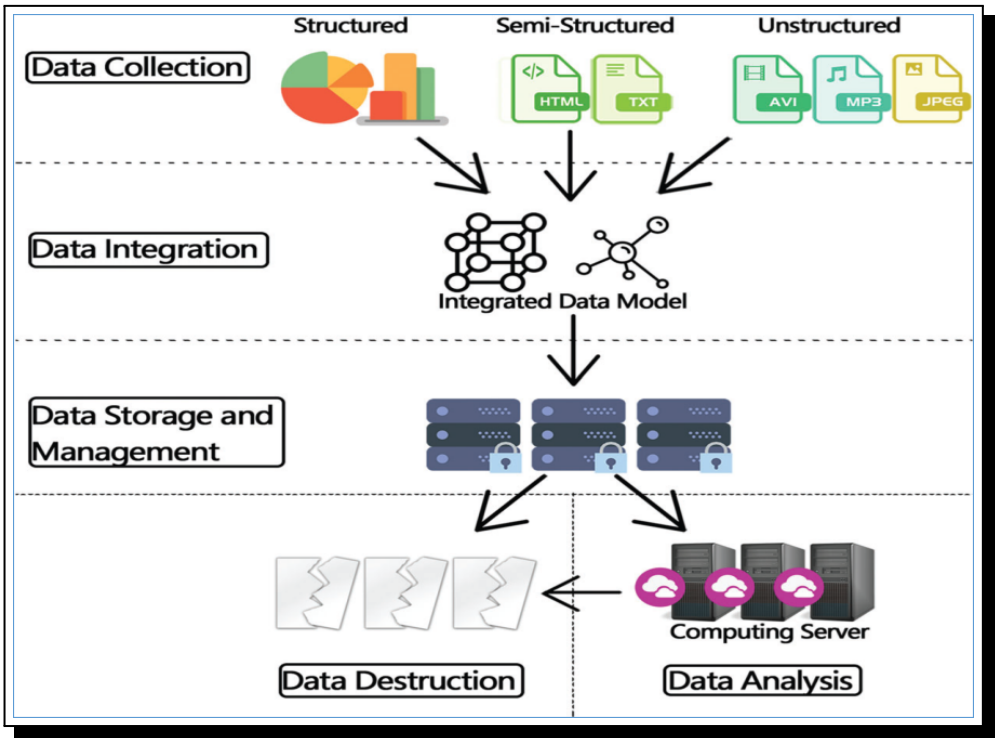

Figure 1. The lifetime of big data [2]

The importance of this study appears theoretically and practically as follow. On the theoretical level, this research focuses on some of the most pervasive obstacles that researchers face across a variety of subjects, supporting them in creating a comprehensive knowledge of these concerns and validating the approaches that are employed to solve them. This study contains previously established solutions to these difficulties that have been previously identified. Furthermore, these solutions would be beneficial to businesses and consumers who interact with big data environments on a daily basis in practice.Security is one of the most fundamental concerns for users, which is why many organizations continue to have misgivings about employing big data analytics. The following question identifies the study's problem: What is the scope of data challenges in the big data environment, and what is the level of security and solution employed to address them?

2. The Details of the Reviewed Articles

By analyzing and critiquing published studies on the topic of challenges and obstacles facing data in the big data environment, the descriptive, analytical, and documentary approach being used. These approaches are based on referring to documents and literature from research, articles, and books, and dealing with them with description and analysis to extract relevant conclusions and implications. A *systematic literature review* (SLR) mechanism was used to select articles that were relevant and related to big data challenges and issues. Several articles were analyzed, and the results are shown in Table 1.

Table 1. Analyzed results of the various reviewed articles

Ref.	Study title	Main objectives	Main results and recommendations
[7]	Quality 4.0: a review of big data challenges in manufacturing	The purpose of this manuscript is to provide an overview of the Quality 4.0 process monitoring project. Then, four critical concerns are identified (paradigm, project selection, process redesign, and relearning challenges) that must be comprehended and addressed in order for implementation to be successful.	A new manufacturing paradigm, project selection, process redesign, and relearning are all explored in this article. Quality 4.0 projects in Process Monitoring for Quality face many problems that must be fully understood and overcome. This study proposes a revolutionary 7-step problem solving technique (identify, acsensorize, discover, learn, predict, redesign, relearn).
[10]	Modeling the Big Data challenges in context of smart cities –an integrated fuzzy ISM-DEMATEL approach	The objective of this study is to identify and analyze the barriers to effective Big Data application in the creation of smart cities.	According to an analysis of the interrelationships between the difficulties, the diverse population in smart cities and a lack of infrastructure are the primary impediments to the integration of Big Data in the creation of smart cities.
[11]	A review of industrial big data for decision making in intelligent manufacturing	This paper proposes a conceptual framework for intelligent decision-making based on industrial big data-driven technologies, which provides useful insights and perspectives on the field's severe obstacles and future research paths.	This article introduced an intelligent decision-making analysis framework based on industrial big data technologies, as well as the role and fundamental design concepts for each component of the framework. This technique establishes a new intelligent manufacturing paradigm in production environments by emphasizing dynamic perception and precise decision-making based on big data analytics.
[14]	Knowledge, attitude, challenges of big data analytics based on information technology staff's point of view in a developing country	The current study's goal is to present the knowledge, attitude, and challenges of big data analytics from the perspective of IT staff in a developing country.	The survey included 120 IT professionals out of 250. 35% of participants had low, moderate, or high knowledge levels. Age and sex were the two most important factors determining participants' knowledge. IT personnel are enthusiastic about big data analytics. Most IT workers agreed that big data analysis may benefit organization managers and that big data management is vital for the country. Also, 35 big data analytics difficulties were discovered.
[16]	Mining in Big Data: Challenges, solutions	Various obstacles encountered during data mining were described in this article, along with adaptive ways for resolving such scenarios.	Various concerns and problems linked to Big Data Mining and approaches (techniques) for viewing Big Data related to map-reduce over Hadoop are articulated in this article. Also, some big data mining technologies may be configured, which helps specify how to extract useful Big Data information.

Contd. Table

Ref.	Study title	Main objectives	Main results and recommendations
[17]	A Review of Big Data Challenges and Preserving Privacy in Big Data	This article discusses big data, the issues associated with big data mining, and the privacy concerns associated with big data.	The authors believe that additional work is necessary to tackle the system's difficulties, which include heterogeneity, scalability, infrastructure failures, timeliness, and privacy. If an individual's privacy is breached, the repercussions might be catastrophic.
[19]	Big data challenges for resource-constrained organizations in a developing economy	This article linked big data challenges to an organization's internal and external resources using the resource-based view. It offered three use examples from Bangladesh's developing economy's contemporary fashion wear, modern footwear, and ethnic clothing businesses.	The findings indicate that a lack of financial resources, followed by a lack of human, complementary organizational, and technological resources, are important problems for resource-constrained enterprises, particularly those in developing countries.
[9]	Big Data Challenges and Opportunities in Healthcare Informatics and Smart Hospitals	With a full explanation of big data and its properties, this paper discussed some of the prospects and problems for big data and its analytical tools in the healthcare informatics domain.	The experimental results indicated that when Spark is used to distribute storage and processing across multiple clusters, J48 and Nave Bayer perform admirably with monitoring datasets.
[13]	Big Data Challenges and Achievements: Applications on Smart Cities and Energy sector	The study analyses the Big Data issues and their resolution.	The article discusses how a variety of data techniques, including Big Data, OLAP, large data, large data transfer, and large data privacy, are used in research and development in the subject of large research.
[4]	Big Data Security: Challenges, Recommendations and Solutions	This study discussed the security concerns associated with Big Data and the state of the art in terms of approaches, procedures, and solutions for protecting data-intensive information systems.	Because securing very large data sets is nearly impossible, it is more practical to protect the data's value and key attributes rather than the data itself, to assess the security risks associated with combining different evolving Big Data technologies, and to tailor security tools to the Big Data project's objectives.
[5]	Big Data Challenges and Data Aggregation Strategies in Wireless Sensor Networks	This paper established the big data paradigm and its main dimensions, which are difficult to grasp. The research also presents large data issues that must be overcome to efficiently manipulate massive data sets, and suggests a new classification based on WSN requirements and challenges.	This article offered a classification scheme for massive sensor data concerns and discussed proposed solutions. As massive sensor data aggregation is our primary focus, we will discuss its proposed strategies in detail.
[2]	Challenges and techniques in Big data security and privacy: A review	Big data security and privacy issues were examined in this study, particularly in light of the many lifetimes of Big data. Homomorphic encryption, multi-party computation, attribute-based encryption, and anonymity protection in social networks were all examined in detail.	This study examined two possibilities for Big data security and privacy development, including regulatory and technical directions. Without a question, existing Big data technology and solutions are severely constrained in their ability to fully address actual Big data security and privacy issues.

Contd. Table

Ref.	Study title	Main objectives	Main results and recommendations
[3]	Big Data Analytics in Real Time – Technical Challenges and its Solutions	An open-source solution for analyzing large amounts of data in real-time, generating alerts in case of business emergencies was presented in this study. The paper also presented open-stack hardware topology.	Suggest an algorithm that provides better security for the data to be stored and transmitted over the communication line, and is fast when uploading and downloading data from the cloud
[18]	Critical analysis of Big Data challenges and analytical methods	This SLR examines historical trends and extant patterns/themes in BDA research, evaluates contributions, synthesizes information, and identifies limitations, implications, and prospective future research routes to help the academic community explore research themes/patterns.	The analysis reported in this work identifies pertinent BD research projects that have contributed conceptually and empirically to the BDA's expansion and accumulation of intellectual riches in the technology and organizational resource management disciplines.
[1]	Big Data Security – The Big Challenge	This article examined the challenges surrounding Big Data. Additionally, it explored the difficulties surrounding the privacy of this data and some of the existing techniques for ensuring data privacy.	In comparison to privacy-preserving aggregation and operations on encrypted data, de-identification is more viable for privacy-preserving big data analytics if efficient and privacy-preserving algorithms are developed to help decrease the danger of re-identification.

While the review of prior studies in Table 1 revealed that data face numerous challenges and hurdles in the big data environment, in the following, each study will be present in detail followed by challenges and solutions in Table 2 as follow:

- The study of [4] discussed the security challenges associated with Big Data and the state of the art in terms of methods, mechanisms, and solutions for protecting data-intensive information systems. Extraction of valuable insight and information from disparate large data sources enables organizations to enhance their competitive advantage. For example, analysis of data streams or archives (e.g., through the use of predictive or identification models) can aid in optimizing production processes, enhancing value-added services, and adapting them to customer needs. However, the sharing and analysis of Big Data raises numerous security concerns and increases privacy threats. This study discussed several significant Big Data security issues and related solutions and recommendations. Because securing very large data sets is nearly impossible, it is more practical to protect the data's value and key attributes rather than the data itself, to assess the security risks associated with combining different evolving Big Data technologies, and to tailor security tools to the Big Data project's objectives.
- The topic of Big Data was addressed in the [1] paper. Additionally, this study addressed the topic of privacy and some of the current strategies for securing this data. It is difficult to secure Big Data because of both its strengths and its weaknesses. Data leaking is a major concern because of the huge volume and velocity of big Data. There are two additional risks that arise from a rise in the development of intelligent terminals, which are related

to privacy and the ability to forecast human behavior. If we can build efficient and privacy-preserving algorithms to reduce the danger of identification, de-identification is more critical for privacy-preserving big data analytics than privacy-preserving aggregation and operations over encrypted data.

- Security and privacy concerns not only afflict individuals and enterprises, but also obstruct the anticipated benefits and advancements of big data. The [2] study began by reviewing the issues associated with big data security and privacy, with a particular emphasis on the challenges associated with the various lifetimes of big data. Then, they discussed four widely used safe technologies: homomorphic encryption (Figure 2), secure multiparty computation (Figure 3), attribute-based encryption, and anonymous social network protection.

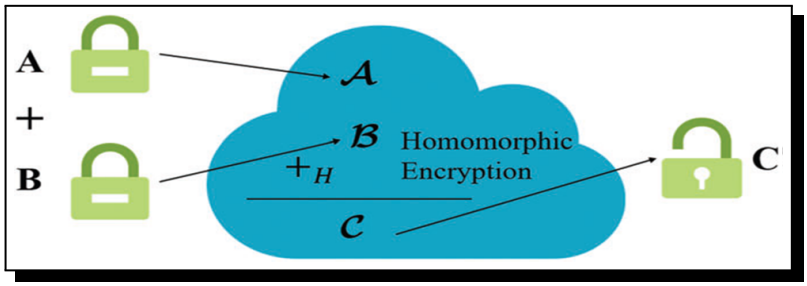

Figure 2. A conceptual example for addition of homomorphic encryption in cloud environment [2]

According to [2] study, the example in Figure 2 demonstrates the addition of homomorphic encryption, in which A and B are sent to a cloud environment. The cipher text is decrypted following the homomorphic addition, and the correct result C is printed out. In general, homomorphic encryption schemes fall into two categories: semi-homomorphic encryption (SHE) and fully homomorphic encryption (FHE). SHE schemes support only a subset of operations, such as addition and multiplication.

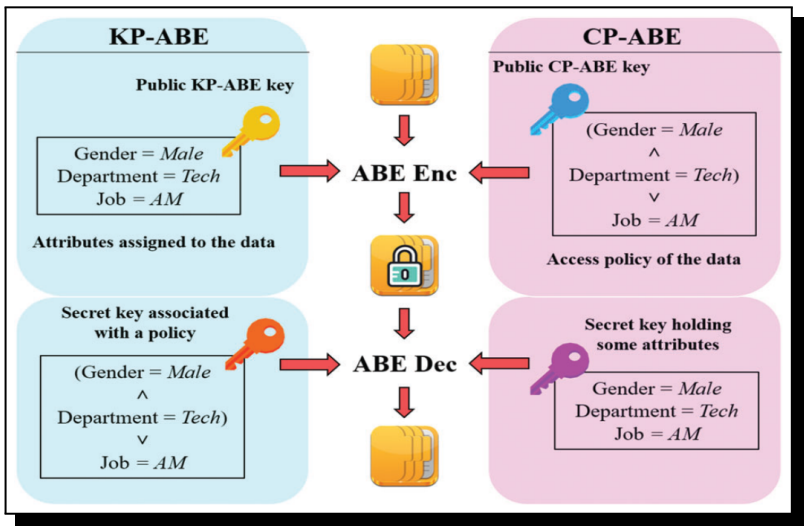

Figure 3. An overview of KP-ABE and CP-ABE for specific attributes of secure multiparty computation [2]

According to [2] study, studies have proposed an effective perfectly secure AMPC scheme with $n = 4t + 1$ parties, where t denotes the number of potentially slow but honest parties. This scheme reduces the time complexity of AMPC. As a result of this design, the communication complexity for each gate is reduced to $O(n^2 \log|F|)$, which is a state-of-the-art complexity for asynchronous verifiable secret sharing (AVSS).

- The issues with big data are storing, processing, visualizing, and ultimately analyzing in real time. The purpose of [3] article was to introduce an open-source solution for analyzing large amounts of data and presenting real-time information on trends and patterns, as well as alerting users in the event of a business emergency. Additionally, the article discussed hardware topology with an open-stack solution Figure 4.

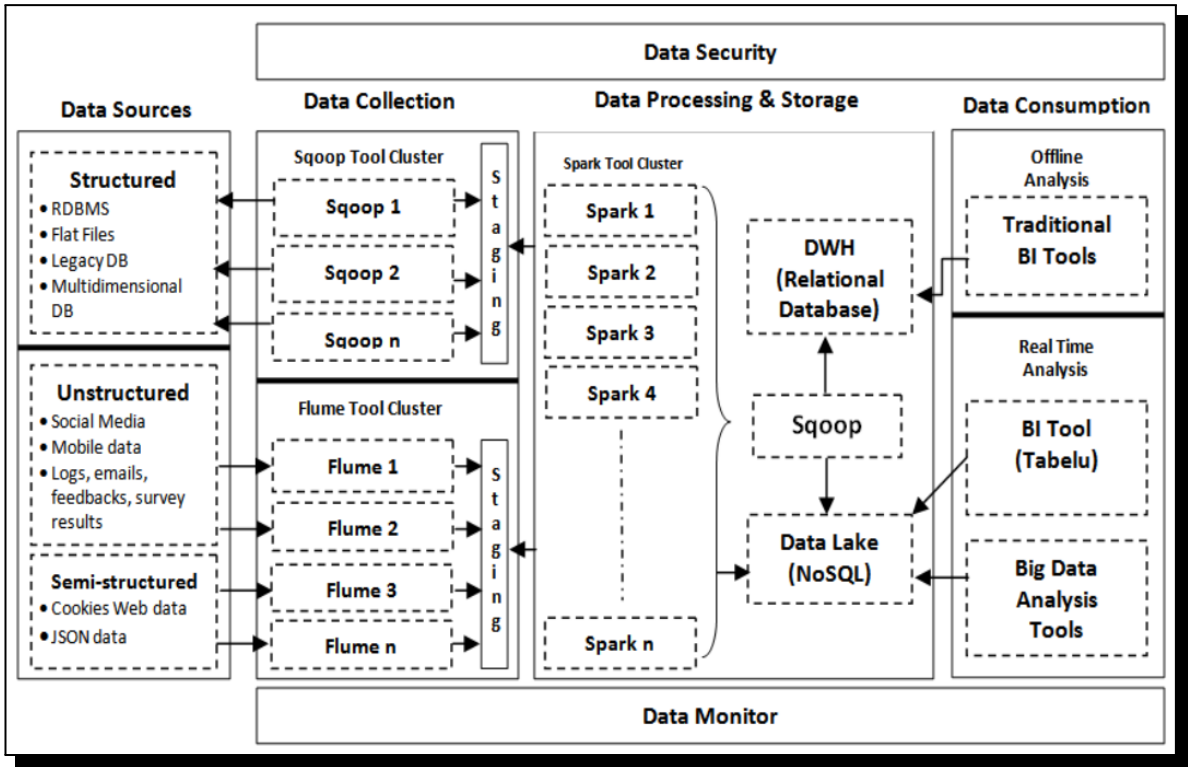

Figure 4. Big Data Solution Stack – Open-source Apache [3]

According to [3] article, using an open source platform to manage large amounts of data in real time results in cost savings, fault tolerance, scalability, high processing power, flexibility, and, in general, a simple coherency model. With visualization tools such as reporting, dash-boarding, ad-hoc queries, drill down/up, slicing/dicing, and predictive analysis, and other evaluation processes, additional data is generated that, when analyzed, helps refine implementation and quality, thereby optimizing various aspects of the business.

- Aggregation of data is a rapidly growing topic of research. It is one of the processing issues inherent with large sensor networks. The [5] article explained the big data paradigm, its primary dimensions, which constitute one of the most difficult ideas to grasp, and its primary analytic tools, which are becoming increasingly integrated into WSNs technology. Additionally, the study discussed the big data issues that must be addressed in order

to handle the enormous data efficiently, and suggests a new classification (Figure 5) of these challenges based on the requirements and challenges of WSNs. As the massive data aggregation challenge is at the heart of our research, this article discussed the proposed solutions in WSNs for optimizing many elements of business.

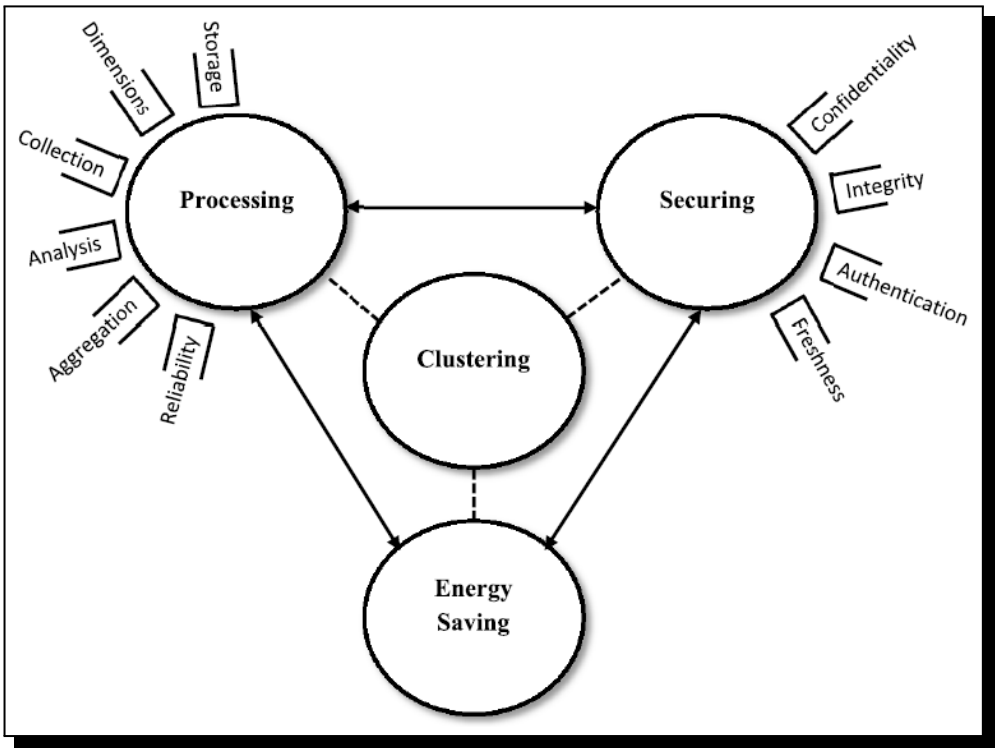

Figure 5. Proposed classification for big data challenges in wireless sensor networks [5]

- According to [7], [11], artificial intelligence and big data in the industrial manufacturing are driving the next generation of production, smart manufacturing. Machine learning algorithms face a very high bar in industrial manufacturing when it comes to quality, compliance, productivity, and creativity. Quality 4.0 is a new concept that aims to solve this problem. In current times of potentially disruptive digital revolution, this moniker was formed from a desire to achieve performance excellence. The Quality 4.0 process monitoring initiative is reviewed in the [7] article. Finally, a paradigm, project selection, process redesign and relearning problems are recognized as the four most important issues for successful implementation. A novel 7-step problem-solving technique based on [7] research is introduced as shown in Figure 6. This Quality 4.0 project is more likely to succeed if the provided strategy is followed. On the other hand, the [16] Paper demonstrated the practicality and internal motivation of big data technology in intelligent manufacturing by providing theoretical analysis of how big data may be used to guide decision making. An intelligent decision framework based on industrial big data-driven technology has been proposed by [11], which provides essential insights and ideas for the severe issues and future research directions in this sector (Figure 7).

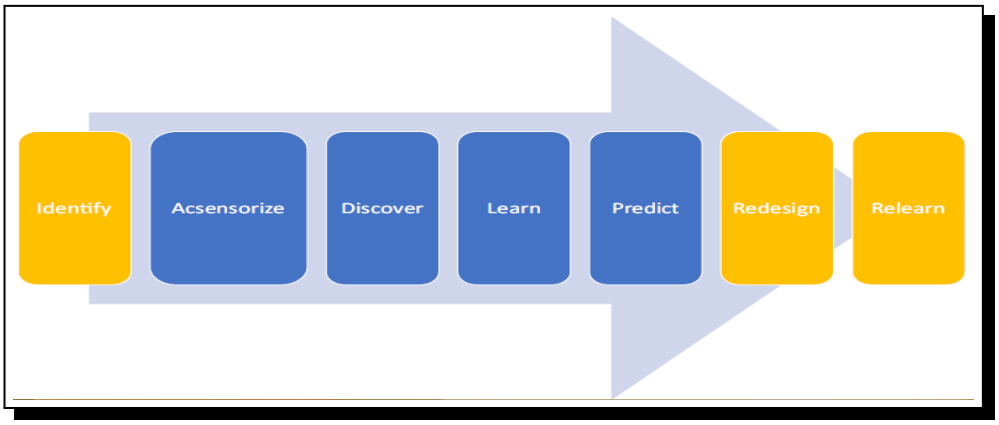

Figure 6. Comprehensive Quality 4.0 problem-solving strategy (new steps in yellow) [7]

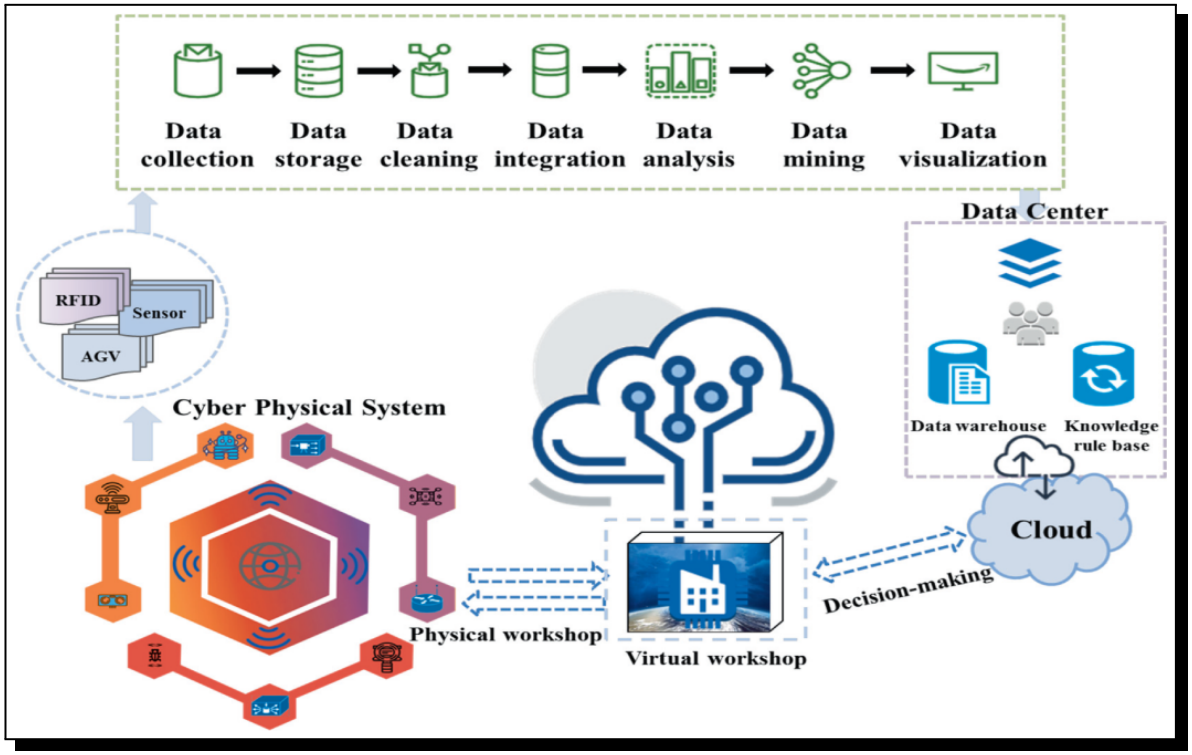

Figure 7. Intelligent decision-making analysis conceptual framework based on industrial big data-driven technology [11]

- With regards to smart hospitals, the [9] paper discussed some of the opportunities and challenges associated with big data and its analytical tools in the healthcare informatics area, along with a full description of big data and its properties. Additionally, this chapter covered cloud computing and big data analytics technologies such as Hadoop and Spark. Additionally, [9] study included an overview of the literature on data mining and big data in biomedical informatics. It also presented a case study of monitoring patients with hypertension to demonstrate the efficacy of using AALs such as IHCAF-PUSH and big data analytics such as Spark in detecting the patients' health status. The experimental

results indicated that when Spark is used to distribute storage and processing across multiple clusters, J48 and Nave Bayer perform well with monitoring datasets.

- It was the goal of the work by [10], [9] to identify and analyze the interrelationships between the issues associated with effectively implementing Big Data in smart city development. To identify and validate the 14 Big Data problems, [10] conducted a literature review. Fuzzy Interpretive Structural Modeling (fuzzy-ISM) (Figure 8) and fuzzy Decision-Making Trial and Evaluation Laboratory (fuzzy-DMTL) are used to develop the inter-relationships between the highlighted issues (fuzzy-DEMATEL). Big Data integration in smart cities is hindered by a diverse population and a lack of infrastructure, according to the [10] study's conclusions.

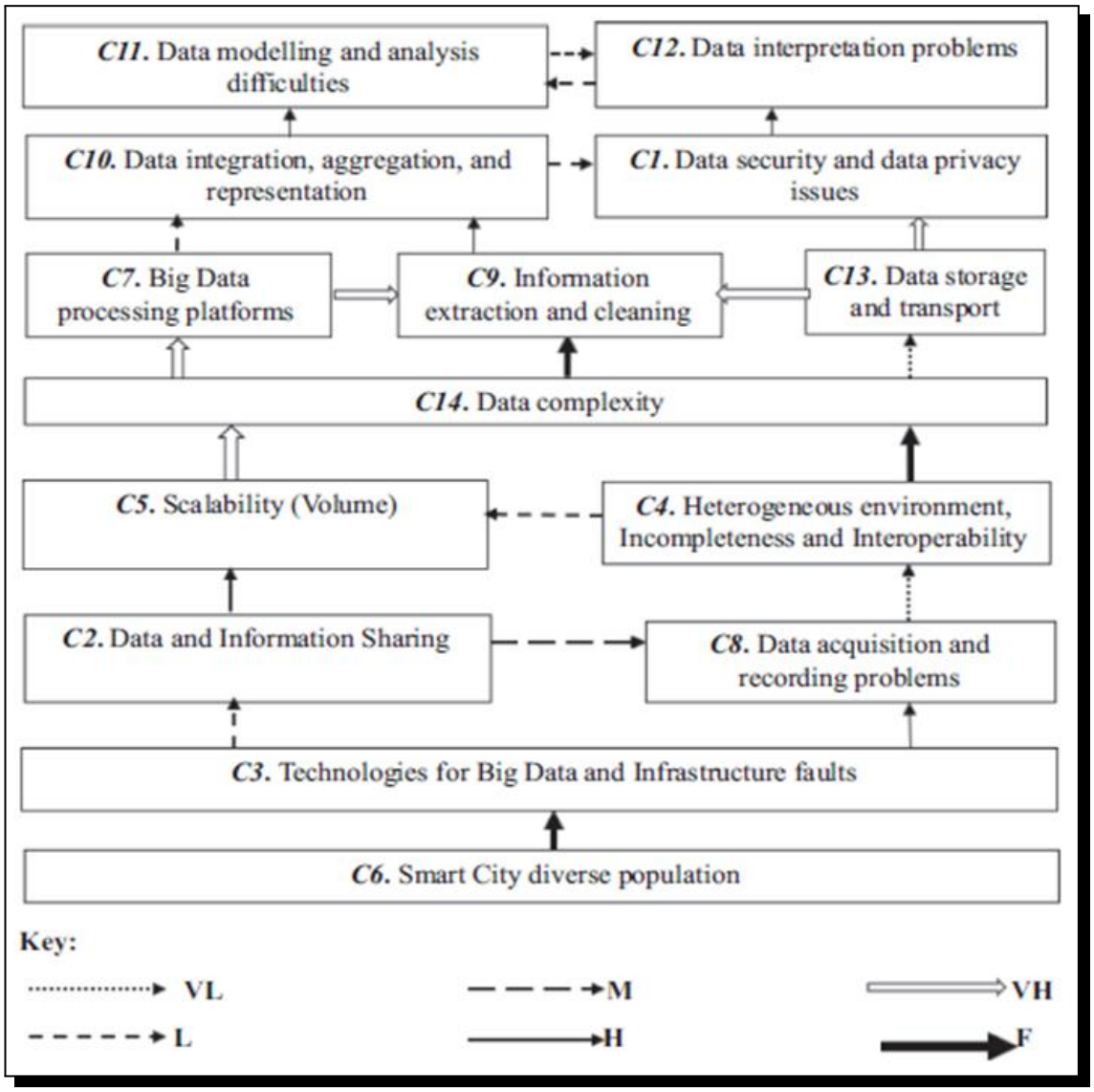

Figure 8. The Fuzzy ISM based model for Big Data challenges in development of smart cities [10]

According to [10] study, a hierarchical and structural model of the digraph is shown in Fig 8. According to the level partitioning, the fuzzy ISM model demonstrates relative importance. The model's foundational challenges, namely "Smart City diverse population" and "Technologies for Big Data and infrastructure failures," drive all subsequent challenges associated with the application of Big Data in the development of smart cities. The challenge "Data complexity" reaches the model's middle hierarchy. Level 1 challenges, namely "Difficulties with data modeling and analysis" and "Data interpretation problems," have no effect on the model because they have a lower influence. All other factors in the model act as a driving force for these factors. It indicates that it is inextricably linked to other issues, as it possesses both driving force and dependency. On the other hand, [9] study claimed that a variety of data techniques, such as Big Data, OLAP, large data, large data transmission, and large data privacy, are used in research and development in the field of large research. Excellent data is a necessary and efficient component of next-generation information systems.

- Both [14], [19] studies concentrated on the developing countries. The aim of [14] study was to show the knowledge, attitude, and issues associated with big data analytics from the perspective of IT employees in a developing country. The study enrolled 120 IT professionals out of a total of 250. 35.0 percent, 33.3 percent, and 31.7 percent of participants, respectively, had low, moderate, or high levels of knowledge. The two most significant variables determining participants' knowledge levels were their age and sex. IT professionals have a favorable view toward big data analytics. The majority of IT personnel agreed that big data management is vital for the country and that big data analyses can give several benefits to business leaders. Additionally, 35 big data analytics difficulties were discovered. The findings indicated that big data analytics faces numerous challenges in the following areas: awareness and education, recruiting skilled specialists, communicating the benefits of big data analytics to IT managers and policymakers, conducting research projects, and developing a national and local strategic plan. On the other hand, [5] study examined the obstacles that businesses may face during the implementation stage of big data. The [5] research connected these problems to an organization's internal and external resources, based on the *resource-based view* (RBV) of enterprises. It offered three use cases from Bangladesh's contemporary fashion wear, modern footwear, and ethnic clothing industries through case study analysis. The findings indicated that lacks of financial resources, followed by a lack of human, complementary organizational and technological resources, are important problems for resource-constrained enterprises, particularly those in developing countries.
- Regarding big data mining, the [16] article explored numerous issues encountered during data mining and proposed adaptable ways to address such scenarios. These insights into the obstacles assist in a more complete understanding of the issues at hand. The [16] study articulated the different concerns and problems associated with Big Data Mining, as well as approaches (techniques) for monitoring Big Data linked to map-reduce over Hadoop (Figure 9). HDFS enables enterprises to reach a broader audience and the marketplace to make more informed decisions. Additionally, some big data mining technologies may be adjusted, which aids in describing how to extract useful Big Data information from it. This enables researchers to select the most appropriate instrument for the job.

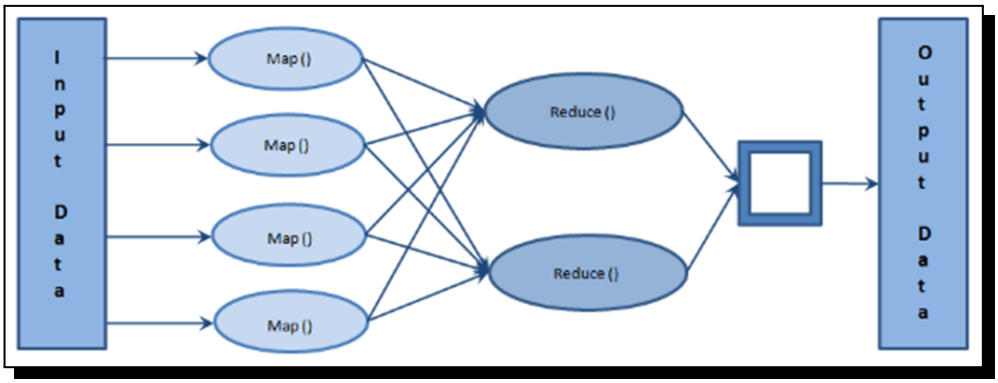

Figure 9. Map-Reduce Architecture [16]

According to The [16] study, it is possible to see an example of a map-reduce architecture in Figure 9, where input data are processed by a variety of mappers, and the summation is handled by reducers. Reducers receive data from mappers and output the desired result. Mapping is a must when dealing with large amounts of data Sets because it allows you to create intermediate key/value pairs for each logical record in your input data set. In order to properly extract the data collected, another step called "reduce" is now performed on the data containing the same element.

- With a focus on privacy issues, the [17] offered an overview of big data, the challenges associated with big data mining, and the privacy concerns associated with big data. It is impossible to conduct operations in big data mining without jeopardizing the privacy. Business firms own sensitive information on their clients, which they regard as a valuable asset. To protect this data from unwanted access, number solutions have been presented in the literature, however they all have limitations. Thus, the authors feel that additional strategies and mechanisms that aid in safeguarding privacy during the data analysis process should be developed, as violating an individual's privacy can have devastating consequences for that individual's life.
- The [18] study used a systematic review of the literature to observe and comprehend past trends and extant patterns/themes in the BDA research area, evaluating contributions and synthesizing knowledge, thereby identifying limitations, implications, and potential future research avenues to assist the academic community in exploring research themes/patterns. Thus, to track the deployment of BD methods, a profiling method is used to evaluate Scopus-extracted papers (published in English-language peer-reviewed journals between 1996 and 2015). The analysis reported in this work identifies pertinent BD research projects that have contributed conceptually and empirically to the BDA's expansion and accumulation of intellectual riches in the technology and organizational resource management disciplines.

Table 2. Present challenges and solutions in each study under review

Ref.	Challenges	Author recommendations/solutions
[7]	The paradigm problem	To develop a customized solution, a thorough understanding of the process is required. Iteratively developing such a solution requires collaboration between model developers and subject matter experts (engineers with domain knowledge).
	The project selection problem	Keep in mind that data science projects are not self-contained. Ideally, data science teams should be working on multiple projects at once. A successful project will stimulate new projects and ideas to further the mid/long term visions. Companies with a clear project selection strategy will thrive in Quality 4.0.
	The redesign problem	Data-driven procedures (e.g., feature ranking/selection) should be utilized to extract information (e.g., identify hidden patterns and associations). This data, together with a physics analysis, can be utilized to create valuable hypotheses regarding probable links between product attributes and quality.
	The relearning problem	To create a sustainable solution, a relearning scheme should be created prior to deployment that takes into account the type of drift experienced by the manufacturing system, e.g., temporal, gradual, repeating, or abrupt.
[10]	Data security and data privacy - Data and information sharing -Technologies for Big Data and infrastructure faults - Heterogeneous environment - incompleteness and interoperability - Efficiency, availability and scalability (volume) - Smart city diverse population - Big Data processing platforms issues - Data acquisition and recording - Information extraction and cleaning. - Data integration, aggregation, and representation - Data modeling and analysis - Data interpretation - Data storage and transport issues - Data complexity	Big Data applications must evolve to address the concerns threatening cities' economic, environmental, and social viability. The next hurdles will be developing tools for Big Data analytics and identifying infrastructure failures, which will require major expenditures in software and hardware to support real-time analysis of hundreds of millions of records.
[11]	The problem of data quality management The problem of data security and privacy protection The problem of the generality of the conceptual framework in actual production The problem of data integration processing in industrial manufacturing systems The problem of accessing primary manufacturing data	This paper presents a shared data-driven model for industrial production big data. Data collection, storage, cleansing, integration, analysis, mining, and visualization are covered. To establish a new intelligent manufacturing paradigm, this technology uses big data analytics to develop real-time dynamic perception and correct decision-making. It emphasizes data interaction and standardization, and encourages data interchange between data suppliers and users quickly.

Contd. Table

Ref.	Challenges	Author recommendations/solutions
[14]	Lack of knowledge and attitude of managers and policymakers Lack government policies and plans Lack of knowledge and attitude of IT managers Lack of educational resources and courses Low data quality Weakness in data management Dispersion of information and lack of an aggregation standardized Lack of expert staff Inadequate equipment Lack of successful implemented big data analytics projects Cost Shortage inadequate research	The study's findings indicate that the most critical areas requiring intervention are awareness and education, recruiting skilled specialists, presenting the benefits of big data analytics to IT managers and policymakers, conducting research projects, and developing a national and local strategic plan.
[16]	Divisiveness and Incompleteness Scale and Complexity Timeliness Privacy and Security	Various concerns and problems linked to Big Data Mining and approaches (techniques) for viewing Big Data related to map-reduce over Hadoop are articulated in this article. Also, some big data mining technologies may be configured, which helps specify how to extract useful Big Data information. They can then pick the best tool for the job.
[17]	Heterogeneity – Scalability - Infrastructure faults – Timeliness – Privacy - The extreme volume, velocity, and variety of data	Some privacy-preserving techniques include: Slicing technique - Cryptographic technique - Differential privacy - K-anonymization - Anonymization through generalization - Homomorphic encryption - Data mashup technique – Bucketization - Hybrid approach
[19]	Due to a lack of human experience, BD implementation is hard and time consuming. Senior management's supportive attitude Business development is a critical (complementary) organizational resource. When lower-level managers are not involved in BD planning and implementation choices, users may be more resistant to change.	Education and preparation Programs are needed to help users better understand business development and lessen their reluctance to change. Long-term planning is required for a successful adoption of behavioral economics.
[9]	The volume of data, the rate at which data is generated, the diversity of data kinds, the authenticity of data, and the privacy of patient medical information. The necessity to establish parallel data processing facilities. Providing safe storage for massive unstructured data collections. The requirement for a fault-tolerant method with a high degree of availability.	Hadoop technology has been successful in addressing the majority of the above-mentioned difficulties facing the healthcare business. The HDFS and MapReduce engines are capable of innovatively processing gigabytes of data on commodity systems.

Contd. Table

Ref.	Challenges	Author recommendations/solutions
[13]	To begin, there is the transportation issue, as large amounts of data are stored in a variety of private and personal resources (legacy systems, web, storage of scientific data, sensor databases, publications, social networks, etc.). Second, graphs, plans, and pianos are utilized to aid with decision-making.	A number of significant data techniques, including Big Data, OLAP, large data, large data transfer, and large data privacy, are used in research and development in the field of huge research. Excellent data is a necessary and efficient component of next-generation information systems.
[4]	Big Data Nature The Need to Share Information Multiple Security Requirements Inadequate Traditional Solutions New Security Tools Lack of Maturity Data Anonymization Compatibility with Big Data Technologies Information Reliability and Quality Compliance to Security Laws Regulations and Policies Need of Big Data Experts Big Data Security on Social Networks	Security Foundations for Big Data Projects Choosing Adequate Security Solutions Risk Analysis Related to Multiple Technologies Anonymization of Confidential or Personal Data Data Cryptography Centralized Security Management Data Confidentiality and Data Access Monitoring Security Surveillance and Monitoring
[5]	Data energy efficient clustering Data gathering Data analysis Energy efficiency Data storage Data aggregation	The authors suggest optimal cluster derivations to save energy. The suggested clustering techniques are based on the modified Expectation-Maximization (EM) methodology, which leverages the Gaussian mixture distribution of the network nodes. To solve continuous big data collection and energy efficiency of densely dispersed WSNs, developed a strategy based on sink node mobility for cluster selection. Based on an energy efficient approach for spanning tree node creation, suggested an itinerary planning scheme for multi-agent systems.
[2]	Storage and management challenges Transmit and sharing challenges Analytical challenges	Some techniques for big data security and privacy include: Homomorphic encryption Secure multiparty computation Attribute-based encryption Anonymous protection in social network
[3]	Real Time Data Collection Real Time Data Processing Real Time Data Visualization	Big data analysis requires the best technologies at every stage: collection, integration, cleansing, organization, visualization, and conclusion. The best technology improves the total system's processing power for real-time analysis. So each level requires efficient and inexpensive implementation. Open stack technology with a parallel and distributed approach is one technique to improve efficiency and economy.

Contd. Table

Ref.	Challenges	Author recommendations/solutions
[18]	Data challenges include data amount, variety, velocity, truthfulness, volatility, quality, discovery, and dogmatism. Process challenges include how to acquire data, integrate data, transform data, select the correct model for analysis, and present outcomes. Management challenges encompass a range of issues, including privacy, security, governance, and ethical considerations.	Customizing BDA solutions to an organization's specific needs may need integrating many data sources and installing software on the organization's hardware. Actions are determined and assessed in relation to corporate objectives, requirements, and restrictions via prescriptive solutions.
[1]	Among the Big Data security and privacy concerns are: Users' (people's) data is collected and used to improve the organization's business. This is done by revealing hidden insights in their lives. A literate person would benefit from Big Data predictive analysis, whereas the less privileged would be easily detected and treated poorly. In the absence of a means of retaliation or even awareness, the exploitation of Big Data by law enforcement increases the danger of undesired consequences. Re-identification attacks, which include correlation attacks, arbitrary identification attacks, and targeted identification attacks, involve scanning a large dataset for correlations that lead to a single individual's unique fingerprint.	If the obtained data is personal or sensitive, we must protect it physically and electronically before storing it. We must protect data both physically and electronically. Big data sharing and processing techniques should be efficient and privacy-preserving.

3. Conclusion

Massive opportunities have opened up as a result of the dangers and problems associated with big data. Big Data is a term that refers to the digital traces of human activity that are extremely large in size. While the properties of big data are well established, their constraints remain complicated and must be addressed effectively if the industry is to more efficiently deploy big data analytics tools. This article examines various aspects and issues associated with big data security, including data privacy, data and information sharing, big data technologies and infrastructure failures, heterogeneous environments, and incompleteness and interoperability. Additionally, it encompasses data quality management, managers' and policymakers' lack of knowledge and attitude, the absence of government policies and plans, the absence of successfully implemented big data analytics projects, the absence of human experience, real-time data collection, real-time data processing, and real-time data visualization. The most frequently seen threats and attacks are directed at data privacy. Additionally, it covers ways to avoid future issues. These contributions to research are theoretical as well as practical. Theoretically, this

study focuses on the most widespread concerns in a variety of fields, supporting scholars in building a comprehensive knowledge of these issues and confirming the approaches employed to address them. To solve these challenges, this study integrates previously established remedies. These solutions will assist businesses and individuals who interface with big data analytics systems on a daily basis. While significant progress has been made, much more effort remains to be done to avoid the dangers of big data.To handle the security risks associated with big data efficiently, technology solutions must be matched by suitable legislation and regulation. Additionally, the authors believe that additional techniques and mechanisms that aid in preserving security and privacy during the data analysis process should be developed, as violating an individual's privacy and security breach can have catastrophic consequences for that individual's life, organizations, and the whole society.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

All the authors contributed significantly in writing this article. The authors read and approved the final manuscript.

References

[1] M. Arora and D.H. Bahuguna, Big data security – the big challenge, *International Journal of Science & Engineering Research* **7**(12) (2016), 399 – 402.

[2] R. Bao, Z. Chen and M.S. Obaidat, Challenges and techniques in big data security and privacy: a review, *Security and Privacy* **1**(4) (2018), p. e13, DOI: 10.1002/spy2.13.

[3] R.K. Behera, A.K. Sahoo and C. Pradhan, Big data analytics in real time - technical challenges and its solutions, in *2017 International Conference on Information Technology* (ICIT), (2017), pp. 30 – 35, DOI: 10.1109/ICIT.2017.39.

[4] F.-Z. Benjelloun and A.A. Lahcen, Big data security: Challenges, recommendations and solutions, in: *Handbook of Research on Security Considerations in Cloud Computing*, K. Munir, M.S. Al-Mutairi and L.A. Mohammed (eds.), IGI Global, Hershey, PA (2015), pp. 301 – 313, DOI: 10.4018/978-1-4666-8387-7.CH014.

[5] S. Boubiche, D.E. Boubiche, A. Bilami and H. Toral-Cruz, Big data challenges and data aggregation strategies in wireless sensor networks, *IEEE Access* **6** (2018), 20558 – 20571, DOI: 10.1109/ACCESS.2018.2821445.

[6] M. Elhoseny, G. Ramírez-González, O.M. Abu-Elnasr, S.A. Shawkat, N. Arunkumar and A. Farouk, Secure medical data transmission model for IOT-based healthcare systems, *IEEE Access* **6** (2018), 20596 – 20608, DOI: 10.1109/ACCESS.2018.2817615.

[7] C.A. Escobar, M.E. McGovern and R. Morales-Menendez, Quality 4.0: A review of big data challenges in manufacturing, *Journal of Intelligent Manufacturing* **32**(8) (2021), 2319 – 2334, DOI: 10.1007/s10845-021-01765-4.

[8] M. Hammer, K. Somers, H. Karre and C. Ramsauer, Profit per hour as a target process control parameter for manufacturing systems enabled by big data analytics and Industry 4.0 infrastructure, *Procedia CIRP* **63** (2017), 715 – 720, DOI: 10.1016/j.procir.2017.03.094.

[9] M.K. Hassan, A.I. El Desouky, S.M. Elghamrawy and A.M. Sarhan, Big data challenges and opportunities in healthcare informatics and smart hospitals BT, in *Security in Smart Cities: Models, Applications, and Challenges*, Lecture Notes in Intelligent Transportation and Infrastructure, A.E. Hassanien, M. Elhoseny, S.H. Ahmed and A.K. Singh (eds.), Springer, Cambridge, (2018), pp. 3 – 26, DOI: 10.1007/978-3-030-01560-2_1.

[10] M.I. Khan, S. Khan, U. Khan and A. Haleem, Modeling the big data challenges in context of smart cities – an integrated fuzzy ISM-DEMATEL approach, *International Journal of Building Pathology and Adaptation*, (2021), DOI: 10.1108/IJBPA-02-2021-0027.

[11] C. Li, Y. Chen and Y. Shang, A review of industrial big data for decision making in intelligent manufacturing, *Engineering Science and Technology, An International Journal* **29** (2022), 101021, DOI: 10.1016/j.jestch.2021.06.001.

[12] R. Lu, H. Zhu, X. Liu, J.K. Liu and J. Shao, Toward efficient and privacy-preserving computing in big data era, *IEEE Network* **28**(4) (2014), 46 – 50, DOI: 10.1109/MNET.2014.6863131.

[13] T.A. Mohammed, A. Ghareeb, H. Al-bayaty and S. Aljawarneh, Big data challenges and achievements: applications on smart cities and energy sector, in *DATA'19: Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems*, (2019), pp. 1 – 5, DOI: 10.1145/3368691.3368717.

[14] E. Nazari, Z. Ebnehoseini, Z. Agharezaei and H. Tabesh, Knowledge, attitude, challenges of big data analytics based on IT staffs point of view in a developing country, *Frontiers in Health Informatics* **9**(1) (2020), Article Id 36, DOI: 10.30699/fhi.v9i1.225.

[15] B. Nelson and T. Olovsson, Security and privacy for big data: A systematic literature review, in *2016 IEEE International Conference on Big Data (Big Data)*, (2016), 3693 – 3702, DOI: 10.1109/BigData.2016.7841037.

[16] P.K. Sadineni, Mining in big data: Challenges, solutions, in: *Proceedings of the 2nd International Conference on IoT, Social, Mobile, Analytics & Cloud in Computational Vision & Bio-Engineering* (ISMAC-CVB 2020), 9 pages, (2020), DOI: 10.2139/ssrn.3734166.

[17] A. Sharma, G. Singh and S. Rehman, A review of big data challenges and preserving privacy in big data BT, in: *Advances in Data and Information Sciences*, M. Kolhe, S. Tiwari, M. Trivedi and K. Mishra (eds.), Lecture Notes in Networks and Systems book series (LNNS, Vol. 94), (2020), 57 – 65, DOI: 10.1007/978-981-15-0694-9_7.

[18] U. Sivarajah, M.M. Kamal, Z. Irani and V. Weerakkody, Critical analysis of Big Data challenges and analytical methods, *Journal of Business Research* **70** (2017), 263 – 286, DOI: 10.1016/j.jbusres.2016.08.001.

[19] H. Sultana, R.B. Zaman and M. Zahan, Big data challenges for resource-constrained organizations in a developing economy, *Journal of Information Technology Case and Application Research* **22**(2) (2020), 111 – 130, DOI: 10.1080/15228053.2020.1807173.