ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20220124135330

Preliminary communication
Received: 2022-01-24, Accepted: 2022-02-01

Big Data Major Security Issues: Challenges and Defense Strategies

Adel Al-Zahrani*, Mohammed Al-Hebbi

Abstract: Big data has unlocked the door to significant advances in a wide range of scientific fields, and it has emerged as a highly attractive subject both in the world of academia and in business as a result. It has also made significant contributions to innovation, productivity gains, and competitiveness enhancements. However, there are many difficulties associated with data collecting, storage, usage, analysis, privacy, and trust that must be addressed at this time. In addition, inaccurate or misleading big data may lead to an incorrect or invalid interpretation of findings, which can negatively impact the consumers' experiences. This article examines the challenges related to implementing big data security and some important solutions for addressing these problems. So, a total of 12 papers have been extracted and analyzed to add to the corpus of literature by concentrating on several critical issues in the big data analytics sector as well as shedding light on how these challenges influence many domains such as healthcare, education, and business intelligence, among others. While studies have proven that big data poses issues, their approaches to overcoming these obstacles vary. The most frequently mentioned challenges were data, process, privacy, and management. To address these issues, this paper included previously discovered solutions.

Keywords: Big Data; Challenges; Defense; Security; Strategies

1 INTRODUCTION AND LITERATURE REVIEW

In today's society, many individuals use the internet to communicate their social information and behaviors, which has resulted in an explosion of data produced [7]. The constant advancement of technology has resulted in an "explosive" increase in the amount of data generated by various sources, including social networks, mobile devices, sensors, X-ray machines, telescopes, space probes, application logs, climate predictions, and geo-positioning systems, to name a few [21]. Big Data is a term that refers to a huge volume of unstructured, semi-structured, and structured data that is generated at an alarming rate and is often utilized by businesses to gather information for timely actionable decision making [4]. In today's digital world, data creation and gathering are outpacing the ability of the system to keep up [17].

According to various surveyed studies, the quantity of data produced worldwide over the last three years has surpassed the preceding 400 years. This data includes documents, images, videos, web pages, e-mail, and microblogging, among other kinds. Unstructured data is more prevalent than structured data [11]. Additionally, using big data may enable the development of a variety of new kinds of sophisticated services that have been launched lately or will be released in the near future that can be used to improve people's quality of life and assist mitigate dangers and hazards [14]. However, the sheer scale of such data sets creates significant technical difficulties in terms of storage capacity and administration, as well as organization, processing, and analysis, among other things [14]. As with all contemporary information systems, big data includes security concerns throughout the storage, processing, and transmission phases, and similarly requires data and privacy protection [7].

Despite these challenges, if we can manage big data well, it may help you generate money, improve executive efficiency, make strategic choices, offer better services, define requirements, discover new trends, and create new products, all of which are covered by data science [5]. Aside from parallel and distributed processing, data science researches methods for similarity search and graph analysis and clustering and stream processing [12]. Big data problems come in a variety of forms and sizes across all businesses and areas in life. Large-scale data (big data) sets provide a variety of challenges, such as distributed and non-relational computing, cryptography, data provenance; validation; and filtering; safe data storage; and real-time monitoring [6]. Big data will be used more efficiently if the causes of challenges can be identified and addressed [17]. There is no consensus among studies discussing the issues associated with big data security. The majority of them focused on the data itself, its processing, and analysis [4, 9, 15], while ignoring additional issues that could affect big data analytics and decision-makers when dealing with big data. These additional issues may include leadership, company culture, and specialization, among others. To fill this gap, this article, as a consequence, highlights the most significant big data problems encountered by nearly all disciplines while also presenting some defense strategies that have been suggested in other academic research papers. Unlike previous studies that focused on a single issue and a single solution, this research focuses on the most significant challenges and how they may be resolved most effectively.

Based on the preceding, and given the importance of big data and the benefits it possesses, the idea of preparing this scientific study arose, the importance of which lies in attempting to enrich intellectual production in the field of big data and its applications as well as assisting decision-makers in the organization based on right and correct information that rely on extensive data analysis. The study's goal is to identify the most significant security challenges that may impede the use of this data and then determine how these threats may be mitigated.

So, the research problem of this article can be stated as follows question: What are the significant data security challenges and their corresponding solutions? The documentary analytical descriptive strategy will be used in

this study, which refers to documents and literature such as research, articles, books, and the like, and addressing them in the study with description and analysis to extract the results and indications. To address the study's topic, this study will evaluate and criticize existing papers on security problems in the big data analytics environment using the following research tools: databases available through the Saudi Digital Library and worldwide search engines. A total of 12 papers have been extracted and analyzed.

This study relies on a qualitative method because it is primarily concerned with describing the phenomenon under investigation. According to Saunders et al. (2003) [16], a quantitative approach discusses what occurs during a phenomenon, whereas a qualitative approach explains why this occurs. This study relies on a descriptive analytical method to conduct the research. Descriptive assessments rely on firsthand observation of behavior and events occurring in an individual's natural surroundings and place a greater emphasis on environmental variables [3]. Descriptive analyses may help us better understand how reinforcement works in the natural world [3].

2 ARTICLE BODY

The researcher will discuss extracted papers related to big data challenges in this section. A total of twelve papers will be presented, ranging from the oldest to the most recent.

Al-Abassi et al. (2020) [1] study titled "Industrial Big Data Analytics: Challenges and Opportunities": This review study focused on industrial big data analytics as well as the evolution of the sector over the years. Big data has been discussed in terms of its characteristics, sources, applications, concerns, and challenges, among other things. The following are some of the challenges associated with large data:

- Lack of Large-scale Spatiotemporal Database Representation: Each device's time stamp is acquired and examined statistically. Manufacturing generates huge amounts of data but lacks the infrastructure needed to analyze, integrate, and fuse it. Finding a less priced approximation will be difficult.
- Lack of Effective and Efficient Online Machine Learning Algorithms: Rapid response is necessary in industrial organizations to detect machine irregularities and check production quality, and adding equipment results in ineffective and inefficient preventative measures. Boosting big data analytics with online large-scale machine learning algorithms is already happening.
- Lack of Whole Processes Lifecycle Data Management Systems: Due to restricted storage management capabilities, data quality assurance approaches should be employed to distinguish vital and irrelevant data.
- Lack of Data Visualization Systems: Large amounts of raw data can be turned into graphical displays that help make judgments and disclose intuitive knowledge rapidly. Data visualization illustrates patterns, trends, anomalies, consistent and fluctuating data that text and tables cannot.
- Lack in Data Confidentiality Mechanisms: For financial reasons, small organizations cannot study big data exchange among geographically distributed web sources. Other organizations and web tools must examine their data, potentially compromising security. Small businesses should be cautious when dealing with third parties and preserve their sensitive data.

Hamad et al. (2020) [8] study titled "Big Data Opportunities and Challenges for Analytics Strategies in Jordanian Academic Libraries": The purpose of this study was to examine the idea of big data from the perspective of information technology workers at three large Jordanian university libraries. This study sought to elucidate the use of big data, analytics, and problems in Jordan's university libraries. We conducted a study of the research on big data in libraries and offered an overview of the applications and research objectives in this subject. The status of big data in Jordanian libraries was discussed, as well as the issues related with it. It is a qualitative research study used interviews with 23 librarians working in IT department. The findings revealed the following challenges:

- Staff competency: Participants agreed that libraries must carefully plan to overcome a variety of hurdles when incorporating big data. Staff experience in new technologies like analytics, visualization, and data curation is one of these impediments.
- Infrastructure: Another issue is a lack of infrastructure in terms of hardware and software. The three universities reported having appropriate infrastructure to enable big data usage and analysis, but this infrastructure needed to be upgraded to keep up with technological advances and optimise big data consumption and analysis.
- Financial support: The lack of financial backing for building library infrastructure and educating library staff to manage large data were the main obstacles.
- Privacy and information security: Academic libraries confront additional privacy and security challenges, and staff expressed a wish to update library rules, indicating the necessity for an information security specialist. Academic libraries are most concerned about this because most big data comes from library users and their interactions with library services.

Al-Sai et al. (2019) [2] study titled "Big Data Impacts and Challenges: A Review": Big Data was reviewed for three purposes in this paper: first, to highlight the definitions and characteristics of Big Data and summarize the most common definitions of existing works; second, to identify the impacts and opportunities for the Big Data; third, to identify the main critical challenges associated with Big Data and classify these challenges as (People, Technology, Organizations, Processes, and Data Management) challenges. This paper used a descriptive analytical review. It classified the various challenges as follow:

- People Challenges: The main issue most firms will have when attempting to adopt Big Data is preparing for

implementation and finding engineers with Big Data knowledge.
- Technology Challenges: Big Data analytics requires new methodologies, skills, and capacities for gathering, storing, and analyzing data using Big Data technologies like Hadoop, Spark, and others.
- Organization Challenges: There are several organizational hurdles that must be overcome before Big Data can be successfully deployed.
- Process Challenges: Many businesses are challenged with processing ever-increasing amounts of data. Firms will have to choose between maintaining all data and storing only the most valuable data to handle this issue. Firms must develop ways for determining the most business value from Big Data.
- Data Management Challenges: Organizations faced challenges in capturing, managing, and administering Big Data Achieving real-time data management is difficult. Data warehouses record financial transactions, insurance claims, medical procedures, personal data, diagnosis codes, etc.

Wani & Jabin (2018) [20] study titled "Big Data: Issues, Challenges, and Techniques in Business Intelligence": Using the descriptive analytical method, the purpose of this study is to identify the most critical concerns and challenges associated with big data and to provide a complete evaluation of alternative strategies for dealing with big data problems. This study discussed two facets of big data: the issues surrounding big data and the challenges associated with big data. Issues include management, storage, and processing, while challenges include the following:
- Lack of big data professionals: Firms need highly skilled employees to handle and exploit these high-performance, complex technology for big data processing.
- Interactiveness (or Designing): Interactivity refers to a data mining system's ability to encourage user feedback, support, and ideas.
- Loading and Synchronization: To load data from several sources into a single repository, one must first load the data from each source, and then synchronize it with the common repository.
- Visualization: Data visualization is the practice of visualizing knowledge to aid decision-making.

Vassakis et al. (2018) [18] study titled "Big Data Analytics: Applications, Prospects and Challenges": Using the descriptive analytical method, this study focused exclusively on big data analytics. Big data analytics has great potential and the benefits to data-driven enterprises are critical determinants of competitiveness and innovative performance. However, there are many barriers to implementing a data-driven approach and obtaining important knowledge from big data. As a result, this study concentrated on big data analytics, focusing on the uses, opportunities, and challenges associated with it. What distinguishes this study from others is its emphasis on big data analytics challenges rather than on big data challenges in general. Here are big data analytics challenges:
- Leadership: Big data's power cannot be completely fulfilled without vision or human knowledge. Business leaders who can see future trends and opportunities will be able to motivate their teams to work efficiently and achieve their goals.
- Talent Management: Many data scientists lack both analytical and subject knowledge skills. A data scientist must be able to perform statistical analysis, big data mining, visualization, and machine learning.
- Decision making procedure: Decision-makers struggle with huge data sets. As a result, decision-makers must be able to solve problems using the correct facts or interact with others.
- Decision making Quality: Quality decision making is linked to data, big data analytics, staff, and decision makers.
- Data-driven culture: Enterprise culture also hinders data-driven decision-making. A data-driven culture requires quick summarization, evaluation, and delivery of essential business information to decision makers.
- Data privacy: Businesses must take efforts to protect their consumers' data. Data policies like as privacy, security, intellectual property, and responsibility should be addressed to maximize the benefits of big data.

Mishra et al. (2017) [13] study titled "A bibliographic study on big data: concepts, trends and challenges": Citation and co-citation analyses were conducted using bibliographic and network approaches. This analysis included a review of 57 publications published in ten selected journals over a five-year period (2011–2015). The findings indicate that the number of publications devoted to the study of "big data" has expanded significantly during the last several years. Additionally, the analysis highlights several of the most prominent articles in this field. Finally, the study analyses emerging trends and the corresponding issues with big data. Regarding big data challenges, this paper concluded that challenges could be one or some of the following:
- The real challenge "was to deal with a diversity of data kinds (variety), time-sensitive response needs (velocity), and data inaccuracies (veracity)".
- Applications must also deal with semi-structured and unstructured data, such as text, images, video, and speech.
- Another concern is late responses. This could be due to a lack of resources to collect, store, and analyze massive amounts of data quickly.
- Determining legitimate from invalid data is tough because even the finest data cleaning processes cannot remove inherent ambiguity.
- Even the best data cleansing procedures cannot eradicate data's intrinsic volatility, according to IBM.

Lee (2017) [11] study titled "Big data: Dimensions, evolution, impacts, and challenges": This review study underlined the necessity of data analytics in processing

various organized and unstructured data. Data analytics for merchant reviews was proven in this paper. 400 reviews were analyzed using multiple regressions. A multivariate regression model found the characteristics substantially associated with usefulness votes. The influence of big data on key business metrics is next investigated. Six technical and managerial challenges are then discussed as follow:

- Data quality: The quality of data tends to decline as it grows more unstructured and comes from diverse sources. This might cause serious harm to patients if a medical monitoring system sensor delivers inaccurate data. Quality measures must be developed, data must be assessed, errors must be corrected and the cost-benefit ratio must be evaluated.
- Data security: Implementing comprehensive security management protocols and solutions like as intrusion prevention and detection systems, encryption, and firewalls can help secure big data.
- Privacy: Sensors, including smart health devices and smart car emergency services, can collect data on a person's travels, health, and purchasing habits. Many people fear about their privacy. To improve service quality and save expenses, big data is needed. Businesses and customers must find a way to combine personal data consumption for services with privacy concerns.
- Investment justification: Despite the vaunted benefits of big data, businesses are struggling to justify their investments. Many big data projects have vague problem definitions and rely on emerging technologies, raising the risk of project failure and irreversibility.
- Data management: Edge computing and Hadoop could help firms handle data better. Hadoop is used for large-scale data processing and calculations in distributed computing.
- Shortage of qualified data scientists: The need to analyze unstructured data like text, video, and images is expanding. If the scarcity persists, organizations may have to construct data analytics training programmers to prepare internal staff to meet demand.

Khan et al. (2017) [10] study titled "Big data challenges and opportunities in the hype of Industry 4.0": In this study, the difficulties and prospects of industrial big data are discussed in the context of Industry 4.0, although from a different point of view than previously. When it comes to building big data algorithms and methodologies, the current study assisted researchers in determining the thresholds of these latest Industry 4.0 systems. Big data presents a variety of issues across a variety of systems, but the current study focuses on the challenges and potential associated with Industry 4.0. These challenges are as following:-

- Acquisition of Automation Data: It is challenging to collect data in Industry 4.0 due to the proliferation of technologies and communication networks. Sensors, actuators, and PLCs generate data in the automotive sector. The sensor detects physical activity and sends it to the PLC for processing. Big Data Digitization Modern factory automation is required to preserve price stability and increase production.
- Data Transformation: To draw conclusions or forecast machine failure using archival big data, heterogeneous data must be translated into an interoperable format. Using smart technology in Industry 4.0 involves converting data into smart device-friendly formats.
- Data Integration and Modeling: Bringing together disparate data kinds for rapid production is difficult. Controlling, automating processes, and calculating product costs require industrial big data integration and modeling.
- Real-time access: Cyber-physical systems with sensors, actuators, and other devices require real-time access. As all actuators run sequentially with predetermined time slices, any delay in remote controlling physical devices affects following physical devices. Numerous agents are installed in industrial control engines of automobiles, triggering instructions based on big data storage.
- Security and Privacy: As Industry 4.0 expands, the volume of heterogeneous data increases, as does the move to the cloud. Because everything in Industry 4.0 is administered remotely via a web interface, a hacker may potentially take control of the physical machinery.
- Data Analytics: Incomplete data hinders real-time data analytics, needing pre-processing before analysis. Scalability of analysis is also an issue when output data volume grows.
- Data presentation: Industrial big data is needed for mining and knowledge extraction. Business CEOs need several reports. Before buying something, customers want to know everything about it. As a result, Industry 4.0 has a tough time providing data in several formats to multiple users. The LexisNexis HPCC systems distributed data intensive computation platform is significant because it can encapsulate data and write code for data reading in multiple activities concurrently.

Behera et al. (2017) [4] titled "Big Data Analytics in Real Time – Technical Challenges and its Solutions": The purpose of this article was to demonstrate 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 discusses hardware topology using an open-stack solution. This article focused on three technical challenges: real-time data collecting, real-time data processing, and real-time data visualization, and presented ways for overcoming them. To address these issues, they implemented open-stack technology and a parallel and distributed strategy. Here are steps of this approach:

- Data Sources which contains structured, unstructured, and semi-structured data.
- Data Collection: Data is extracted from a number of sources, data bazaars, and data stockrooms. Hadoop is a distributed and parallel computing system for large datasets. It's an Apache-sponsored open source project.

- Data Processing and Data Storage: Spark is an Apache Software Foundation tool for speeding up Hadoop data processing. Because Spark uses in-memory computation, it can process data faster by caching data and turning it to real-time. Tolerant Distributed Dataset (RDD) saves data in memory transparently and persists to disc only when needed.
- Data Consumption: Data visualization and analysis employing approaches like business intelligence and big data analysis. Business intelligence (BI) tools let business/operations professionals visualize data to aid/improve business/operations. The tool's goal is to find patterns, trends, and other information in the data sea.
- Data Monitoring: Monitoring is a major bottleneck due to the solution's enormous machine count. Ganglia is a great tool for cluster-based monitoring and data processing.
- Data Security: The goal of big data security is to protect an organization's internal data and its customers' data in real-time or near-real-time. Techniques like strong authentication, regular audits, and mandatory access control may be used to protect personal and sensitive data on the cluster of machines.

Wang et al. (2016) [19] study titled "Towards felicitous decision making: An overview on challenges and trends of Big Data": Bibliometric analysis of 2924 articles published between 2000 and 2016 were used to conduct this study. On the basis of four issues, this paper provided an overview of Big Data, which included: (I) Big Data concepts, characteristics, and processing paradigms; (II) the most up to date techniques for decision making in Big Data; (III) successful decision making applications of Big Data in social science; and (IV) the current challenges of Big Data, as well as potential future directions. Focusing on the goal of this study, this paper stated that most big data challenges are:

- Challenges in Data Capture/Storage and Curation: Solid-state drives, phase-change memory, and data access optimization may help solve this problem. Data security is a concern during these periods. Prior to deciding on information exchange strategies and protocols like certification, access control, and anonymization, privacy should be considered. Anonymization techniques may hinder data analysis by raising data uncertainty.
- Challenges in Data Analysis and Visualization: Data analysis is tough because of data and computing complexity. A decision-making dilemma comprising diverse sources, enormous quantities, and quickly changing datasets cannot be addressed with ordinary computer technologies. Thus, new approaches should be proposed to re-examine Big Data's computability (and then computational complexity). Insufficient sample sizes, confusing data linkages, and unbalanced (or even uncertain) value density distributions must be thoroughly evaluated.
- Systematic Challenges: Building suitable system architecture is crucial to supporting decision-makers in managing complex data and executing complex computations on Big Data. One option is to use cluster computers connected to an HPC platform. However, this stresses both hardware and software system architectures. Their final answers will help create system designs.
- Non-Technical Challenges: Rather than technological challenges related with Big Data processing, this refers to management issues faced by service providers and users. Big Data should help them communicate with users better.

Nasser & Tariq (2015) [15] study titled "Big Data Challenges": This narrative and descriptive study examined Big Data problems in three categories: data, process, and management. Data difficulties are a subset of challenges linked to data quality. The process category includes all challenges encountered while processing Big Data, from data acquisition to product presentation to clients. The management group addresses data access legal and ethical issues. The "big data technologies stack" is a layered design guide for theoretical Big Data challenges. Each layer will provide technologies to address specific issues, but together they will provide the entire solution. Starting with data issues, this study discussed volume, variety, velocity, veracity, volatility, quality, discovery, and dogmatism. The second issue is (Tab. 1). Privacy, Security, and Governance are all management issues.

Table 1 Challenges according to process stages

The process	Challenges
Data Acquisition and Recording	- Smart filters - Data reduction - Automatic meta-data generation - Data fidelity
Information Extraction and Cleaning	- Transforming structure less data to analytics friendly format - Extraction right information - Adequate error models
Data Integration and Aggregation	- Heterogeneity of data - Effective DB design for big data - Automating integration and aggregation
Analysis and Modeling	- Noisy, untrustworthy and heterogeneous data - Dynamics and inter-related data - Scaling queries - Integrating DB systems and analytics tools - Analytics-on-the-fly
Interpretation	- Making assumptions available to users - Wrong modeling - Erroneous data used - Application bugs

To solve these challenges, the paper introduced a layered architecture reference known as the "big data technology stack" which consisting of seven layers. These layers are:

- Layer 1 - Redundant physical infrastructure (Data challenges): Data properties such as high-volume, high-variety, and high-velocity offer issues for the construction of new technical infrastructure.
- Layer 2 - Security infrastructure (Management challenges): Encryption is the major way for protecting data. Before saving data records, any personally identifiable information must be erased. Before saving the data records, anonymize them and remove all personal sensitive info.
- Layer 3 - Operational databases (Process challenges): The database must support the ACID transactional behavior, which stands for Atomicity, Consistency, Isolation, and Durability.
- Layer 4 - Organizing data services and tools (Data & process challenges): Simply put, the programmer can't

turn 0s and 1s into valuable insights. Big Data services and technologies collect, validate, and organise it. Apache Hadoop is an open source software system designed to handle massive volumes of data in real time.

- Layer 5 - Analytical data warehouses (Process challenges): Data warehouses and Big Data are a hybrid framework. Traditional data warehouses manage highly ordered data while Hadoop manages widely scattered and dynamic data.
- Layer 6 - Big Data analytics (Process challenges): Reporting and dashboards, visualization, and analytics and advanced analytics are the three main categories of analytics technologies that organizations can use individually or together to gain business value.
- Layer 7 - Big Data applications (Process challenges): A well-defined API interface allows developers to access the functionality exposed by each tier via those interfaces.

Jin et al. (2015) [9] titled "Significance and Challenges of Big Data Research": The concept of big data is briefly addressed in this narrative and descriptive paper. Onto the significance and promise of Big Data. The video then showed real-world big data projects from throughout the world. Finally, it identified the primary obstacles (such as data, computational, and system complexity) and recommended solutions. Like the previous study [15], this investigation confirmed that some of the difficulties stem from big data characteristics, some from current analytical models and methodologies, and yet others from current data processing technologies. Authors in this paper categorized challenges as the following:

- **Data complexity:** Identifying and quantifying significant features is tough when dealing with large data sets. To do so, we'll require data dispersion theories and models for various contacts. We'll also need to understand the relationship between computational and data complexity.
- **Computational complexity:** Stating the obvious, we should look at the weak CAP network shared-data system model and its algebraic computational theory. Then we'll need to enhance networking, storage, and processing for big data. Existing reduction-based computing technologies must also be studied.
- System complexity: Consider current workloads and resource distribution. Application development requires study in performance evaluation, distributed system architecture, streaming computing, and online data processing. Benchmarks are useful for predicting and validating system performance.

3 DISCUSSION

While studies have confirmed that big data presents challenges, they differ in their approach to these challenges. Some research has concluded that challenges can be separated into three categories: data, process, and system or management (see, for example, [4, 9, 15], whereas Mishra et al. (2017) [13] have added a fourth area, which he refers to as data quality. According to Wang et al. (2016) [19], the prior issues have been reinforced by the addition of systemic challenges, which refer to developing a functionally adequate system design.

With the addition of the following difficulties, Khan et al. (2017) [10] brought categorization to a whole new level: acquisition of automation data; data transformation; data integration and modeling; real-time access; security and privacy; data analytics; and data presentation, among others. Rather than focusing solely on data, process, and visualization, Vassakis et al. (2018) [18] added the following: Leadership, Talent Management, Decision Making Procedure, Decision Making Quality, and Data-Driven Culture to his list of priorities. Apart from the lack of professionals and visualization, Wani & Jabin (2018) [20] emphasized the importance of interactiveness (or design), which refers to the capacity of a system to facilitate user interaction, such as feedback, assistance, and ideas, and loading. They also introduced synchronization, which refers to the act of transferring data from multiple heterogeneous data sources to a single data repository, as well as the process of ensuring data consistency across time between different data sources and shared storage. Al-Sai et al. (2019) [2] categorize big data difficulties into broad categories: People Challenges, Technology Challenges, Organization Challenges, Process Challenges, and Data Management Challenges, among others. Finally, Hamad et al. (2020) [8] added infrastructure and financial support to the list of earlier obstacles that previous investigations had identified.

Regarding solutions, Nasser & Tariq (2015) [15] introduced a layered architecture framework dubbed the "big data technology stack" that consists of seven layers beginning with redundant physical infrastructure, security infrastructure, operational databases, organizing data services and tools, analytic data warehouses, and big data analytics. According to Jin et al. (2015) [9], the following solutions are recommended to overcome the computational complexity associated with big data applications:

1) Move away from traditional computing paradigms and investigate the weak CAP network shared-data system model and its algebraic computational theory.
2) Develop algorithms for distributed and streaming computing and a framework for big data computing that integrates and optimizes communication, storage, and processing.
3) Investigate the non-deterministic algorithmic theory that does not assume independent and identical distributions.
4) Investigate existing reduction-basis techniques.

Apart from deconstructing the relationship between the complexity, computability, and efficiency of big data applications, we'll also need to quantify a variety of energy efficiency factors, such as system throughput, parallel processing capability, job calculation accuracy, and energy consumption per unit [9]. While access to Big Data is constrained by the system imbalance created by CPU-intensive yet I/O-deficient systems, several related technologies such as solid-state drives, phase-change memory, and data access optimization may help [19]. A decision-making dilemma comprising diverse sources, enormous quantities, and quickly changing datasets cannot

be addressed by ordinary computer technologies. So, new approaches should be proposed to re-examine Big Data's computability (and then computational complexity). Insufficient sample sizes, confusing data linkages, and unbalanced (or even indeterminate) value density distributions must all be thoroughly investigated [19]. To address challenges relating to real-time data collecting, real-time data processing, and real-time data visualization, Behera et al. (2017) [4] implemented open-stack technology and a parallel and distributed strategy shown in Fig 1.

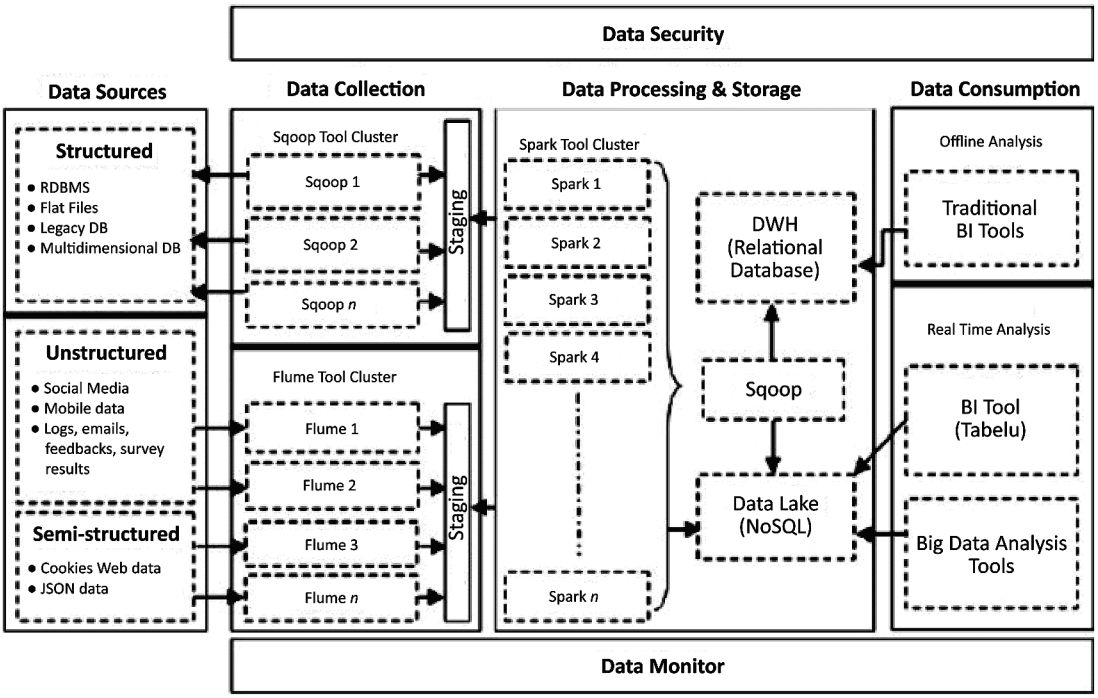

Figure 1 Open-stack technology [4]

4 CONCLUSION AND FUTURE WORKS

Enormous possibilities have arisen as a result of the advent of the Big Data age. In addition to influencing everyone's social and economic conduct, big data has also affected their way of life and thought. Big data is a valuable tool for solving various challenges; yet, it has also introduced a slew of security concerns. This paper examines multiple security features and problems in the big data environment from the perspectives of privacy protection, trust, analysis, technique, and access control. These include technical challenges, security, user privacy protection challenges, and safe storage of massive data, analysis challenges, and trust security issues. It also discusses preventive solutions for these issues. These study contributions appear theoretically and practically. Theoretically, despite the fact that studies have demonstrated that big data presents challenges, the ways taken to overcome these challenges differ. Data, procedure, privacy, and management were the four most often reported difficulties in the survey. This study includes previously identified solutions to these problems in order to solve them. Practically, these solutions would be useful to decision-makers in a wide range of industries, including healthcare, industry, education, marketing, and others. Although there has been some progress, there is still more work to be done to protect big data security and privacy. To effectively address big data security problems, it is necessary to integrate technological solutions with appropriate laws and legislation.

5 REFERENCES

[1] Al-Abassi, A., Karimipour, H., HaddadPajouh, H., Dehghantanha, A., & Parizi, R. M. (2020). *Industrial Big Data Analytics: Challenges and Opportunities*. In K.-K. R. Choo & A. Dehghantanha (Eds.), Handbook of Big Data Privacy (pp. 37–61). Springer International Publishing. https://doi.org/10.1007/978-3-030-38557-6_3

[2] Al-Sai, Z. A., Abdullah, R., & husin, M. h. (2019). Big Data Impacts and Challenges: A Review. *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, 150-155. https://doi.org/10.1109/JEEIT.2019.8717484

[3] Anderson, C. M. & Long, E. S. (2002). Use of a structured descriptive assessment methodology to identify variables affecting problem behavior. *Journal of Applied Behavior Analysis, 35*(2), 137-154. https://doi.org/10.1901/jaba.2002.35-137

[4] Behera, R. K., Sahoo, A. K., & Pradhan, C. (2017). Big Data Analytics in Real Time – Technical Challenges and its Solutions. *International Conference on Information Technology*, 30-35. https://doi.org/10.1109/ICIT.2017.39

[5] Chen, C. P. & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275, 314-347. https://doi.org/https://doi.org/10.1016/j.ins.2014.01.015

[6] Cloud Security Alliance. (2013). Expanded Top Ten Big Data Security and Privacy Challenges.

[7] Gupta, N. K. & Rohil, M. K. (2020). Big Data Security Challenges and Preventive Solutions. In Sharma, N., Chakrabarti, A., & Balas, V. E. (Eds.), *Data Management, Analytics and Innovation*, Springer Singapore, 285-299. https://doi.org/10.1007/978-981-32-9949-8_21

[8] Hamad, F., Fakhuri, H., & Abdel Jabbar, S. (2020). Big Data Opportunities and Challenges for Analytics Strategies in Jordanian Academic Libraries. *New Review of Academic Librarianship*, 1-24. https://doi.org/10.1080/13614533.2020.1764071

[9] Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and Challenges of Big Data Research. *Big Data Research, 2*(2), 59-64. https://doi.org/https://doi.org/10.1016/j.bdr.2015.01.006

[10] Khan, M., Wu, X., Xu, X., & Dou, W. (2017). Big data challenges and opportunities in the hype of Industry 4.0. *2017 IEEE International Conference on Communications (ICC)*, 1-6. https://doi.org/10.1109/ICC.2017.7996801

[11] Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons, 60*(3), 293-303. https://doi.org/https://doi.org/10.1016/j.bushor.2017.01.004

[12] Miloslavskaya, N., Senatorov, M., Tolstoy, A., & Zapechnikov, S. (2014). Information Security Maintenance Issues for Big Security-Related Data. *2014 International Conference on Future Internet of Things and Cloud*, 361-366. https://doi.org/10.1109/FiCloud.2014.64

[13] Mishra, D., Luo, Z., Jiang, S., Papadopoulos, T., & Dubey, R. (2017). A bibliographic study on big data: concepts, trends and challenges. *Business Process Management Journal, 23*(3), 555-573. https://doi.org/10.1108/BPMJ-10-2015-0149

[14] Mohamed, N. & Al-Jaroodi, J. (2014). Real-Time Big Data Analytics: Applications and Challenges Nader. *2014 International Conference on High Performance Computing & Simulation (HPCS)*, 305-310. https://doi.org/10.1109/HPCSim.2014.6903700

[15] Nasser, T. & Tariq, R. S. (2015). Big data challenges. *J Comput Eng Inf Technol, 9307*(2). https://doi.org/10.4172/2324-9307.1000133

[16] Saunders, M., Lewis, P., & Thornhill, A. (2003). *Research methods forbusiness students*. Essex: Prentice Hall: Financial Times.

[17] Terzi, D. S., Terzi, R., & Sagiroglu, S. (2015). A survey on security and privacy issues in big data. *2015 10th International Conference for Internet Technology and Secured Transactions (ICITST)*, 202-207. https://doi.org/10.1109/ICITST.2015.7412089

[18] Vassakis, K., Petrakis, E., & Kopanakis, I. (2018). *Big Data Analytics: Applications, Prospects and Challenges BT - Mobile Big Data: A Roadmap from Models to Technologies* (Skourletopoulos, G., Mastorakis, G., Mavromoustakis, C. X., Dobre, C., & Pallis, E. (eds.), Springer International Publishing, 3-20. https://doi.org/10.1007/978-3-319-67925-9_1

[19] Wang, H., Xu, Z., Fujita, H., & Liu, S. (2016). Towards felicitous decision making: An overview on challenges and trends of Big Data. *Information Sciences*, 367, 747-765. https://doi.org/https://doi.org/10.1016/j.ins.2016.07.007

[20] Wani, M. A. & Jabin, S. (2018). *Big Data: Issues, Challenges, and Techniques in Business Intelligence BT - Big Data Analytics* (Aggarwal, V. B., Bhatnagar, V., & Mishra, D. K. (eds.), Springer Singapore, 613-628. https://doi.org/10.1007/978-981-10-6620-7_59

[21] Xia, F., Yang, L. T., Wang, L., & Vinel, A. (2012). Internet of things. *International Journal of Communication Systems, 25*(9), 1101. https://doi.org/10.1002/dac.2417

Authors' contacts:

Adel Al-Zahrani
(Corresponding author)
Department of Information Science,
King Abdulaziz University,
Jeddah, Kingdom of Saudi Arabia
E-mail: muchkoirii@gmail.com
https://orcid.org/0000-0001-7823-303X

Mohammed Al-Hebbi
Department of Healthcare Technology,
King Abdulaziz Medical City, Jeddah,
Kingdom of Saudi Arabia