Big Data Security: Challenges, Risks, and Best Practices for Organizations

Hassan Mistareehi[1†] Abdulrahman Yarali[2††], and Jason Owen[3†††],

Murray State University, Murray, KY, USA

Abstract
Rapid growth in big data revolutionizes industries through the facilitation of advanced analytics and insights, while at the same time, it brings considerable challenges in data management and security. Big data is huge amounts of structured and unstructured data that originate from different sources, including social media, financial systems, and IoT devices. While its benefits are undeniable, the complexity and size of big data expose organizations to a number of risks, which include data breaches, unauthorized access, and other security-related problems. This paper discusses the definition of big data, its significance for different sectors, and the challenges in keeping such extensive datasets secure. The paper addresses several pressing issues: data origin, data encryption, data access, and vulnerabilities from distributed frameworks in Hadoop/NoSQL. Periodic audit processes, intelligent encryption systems, and strong mechanisms for access control are some strategies that may mitigate these risks. By comprehending these challenges and using best practices, an organization can leverage big data while taking care of data integrity and security. This study by implication overemphasizes the call for a proactive approach in the big data security area to allow both large and small-scale organizations to operate in a safer digital environment.

Keywords:
Big Data, Data Security, Advanced Analytics, Data Management Challenges, Hadoop/NoSQL Vulnerabilities.

1. Introduction

It is important to understand what big data means. When computers process and send information, the data is represented as a set of symbols, quantities, and characters as electrical impulses. These impulses are saved on optical or magnetic mediums and are called data or computer data. When data becomes so massive and complex that traditional storage tools and systems can no longer manage them effectively, it is categorized as big data [1, 2]. Managing and transmitting such data requires specialized storage capabilities. In today's world, the complexity of transactions and the vast amount of data exchanged—especially in large industries—have increased the importance of big data. Big data also influences research and the sharing of information across multiple platforms. For instance, the New York Stock Exchange creates nearly a terabyte of data each day through new trades. Social media sites such as Snapchat and Facebook collectively create hundreds of terabytes daily. This data includes chats, blogs, videos, photos, and comments about specific topics, as stated by [3, 4]. Besides, jet engines are very data-intensive, generating roughly one terabyte of data a second during flight. Thousands of flights every day may reach several petabytes.

2. Forms of Big Data

Big Data can be classified into three forms: structured, unstructured, and semi-structured.

1. Structured Data: When processing is easy, and the data is processed, stored, and available in a fixed format, it is called structured data. People are used to deriving and deciphering meaning out of such data in a quick and agile manner. These are the purview of the structured data, and it is easy to work with those data quickly and access them conveniently. However, the size of such data may become very big to manage, and such data then come under the big data.
2. Unstructured Data: When the form of the data is unknown and is unstructured in manner, and the expanse of such data is vast and huge almost being at the point of being non-manageable, the data comes under the head of unstructured big data in computer science. Randomly selecting text files, videos, and images can make big, unstructured data [5]. Many organizations require such big, unstructured data for their daily business and records. The search options and output generated by Yahoo or any search engine are examples of unstructured big data.

Manuscript received April 5, 2026
Manuscript revised April 20, 2026
https://doi.org/10.22937/IJCSNS.2026.26.4.15

3. Semi-Structured Data: Semi-structured data may include both structured and non-¬¬structured forms. It can be seen in a structured form but not in a defined manner. Data produced in an XML file is an example of semi-structured data. Note that unstructured data consists of transaction history files, log files, etc. Online Transaction Processing (OLTP) systems are designed to work with structured data.

3. Features of Big Data

There has been a tremendous change in technologies and businesses. These changes force companies to develop new models and compelling new ways to attract customers and reach new markets. Regarding cloud and big data, several features need to be discussed.

1. Volume: From the name itself, we realize the size of big data. Size is an important factor in determining how important the particular data in question is. It is even required to find out if we can call the data big data after checking its volume. Hence, Volume is a crucial feature of Big Data.
2. Variety: Variety is another important characteristic of Big data. It refers to the different sources from which the data (both structured and unstructured) is received. In the past, the only data sources for most applications were databases and spreadsheets. But at present, photos, videos, audio, emails, PDFs, etc., are also important sources in the application. The variety of unstructured data brings issues in analyzing, storing, and mining the data.
3. Velocity: Velocity is the speed. The speed of data generation is called the velocity of the data. The actual potential of data can be determined by its velocity, i.e., the speed of generating and processing the data. Data flows from different sources, such as business processes, networks, application logs, social media sites, and mobiles [1]. The velocity of big data deals with this massive and continuous data flow.
4. Variability: Sometimes, because of the data's inconsistency, it becomes difficult to handle and manage it. This is what the Variability of Big Data refers to.

4. Benefits of Big Data Processing

1. Decision Making: Important decisions taken by the companies can be supported by outside intelligence. Nowadays, we have access to different social sites, i.e., Facebook, Instagram, and Twitter. The data obtained from them enables the companies to strategize their businesses. The information received from various social platforms becomes a valid source in decision-making for organizations. Better Customer Service: Through big data technologies, the feedback system used in the past is replaced by new systems. Big data and natural language processing technologies are now used to evaluate customer responses. Customer responses are evaluated and read by organizations, leading to better customer service.
2. Efficacy in Operation: Using big data technologies, a landing zone for new data is created. It is a staging area for new data before moving it to the data warehouse. Big data requires huge parallel software running on thousands of servers. It is not possible for big data to work with relational database management systems and static packages. Big data's advantage is that it provides several additional information [4]. Correlation is found, business trends are identified, diseases are prevented, crimes are controlled, and so on.

Even governments are using Big data in decision-making. Different concepts of database engineering are now used for decision-making from big data. Data warehousing, Online Analytical Processing (OlAP), Distributed, parallel databases, etc., are the concepts employed [2]. Only big organizations and governments can provide the infrastructure to analyze such large data. Big Data promises a great future and will contribute to health, crime, market analysis, etc.

5. Security Problems of Big Data

Security of the data is crucial, and hence it is important to know the problems. These problems are data generation may be false, data mappers may be from non-trusted sources, protection problem of cryptography, sensible data gets mined for information, granular access control difficulties, troubles of data provenance, and the security audits are absent. Following, we discussed the major security problems that big data faces.

1. The problem of false or fake data generation: Sometimes, false or fake data is generated and

Manuscript received April 5, 2026
Manuscript revised April 20, 2026
https://doi.org/10.22937/IJCSNS.2026.26.4.15

becomes the cause of major concerns and a big threat to the maintenance of big data. To minimize the usable extent of an organization's big data and show it low, cyber hackers deliberately duplicate it and dump it into the data lake. In manufacturing houses, cyber-hackers get access to data control systems and show wrong results for certain conditions and functions [4]. The result is that in case of malfunctioning of a system, it goes unnoticed as the data does not show the change, or even in case of right functioning, it sometimes gives a false alarm and prevents the right functioning of the system. One needs a robust fraud detection system to check the penetration of such cyber-hackers and cyber-criminals.

2. Possibility of the presence of mappers who are not trusted: After data collection, the big data goes for a simultaneous processing system. In such a method, the paradigm of Map Reduce is used. It is the process of splitting the bulk data into smaller bits and storing it in particular storage settings. If access to the mapper's code is given to a malicious outsider, they can alter the setup of the prevalent mappers, delete some, or add some foreign mappers alien to the system. This way, the data can be destroyed by making an inappropriate series of key/value data pairs. The data will be shown by the Reduce mapping process as faulty. It can also give outsiders access to valuable and key information, and the data can be hacked or stolen. As already discussed, making the data faulty and hacking it is not out of reach because there is normally no added security given to big data. The dependence is generally on the perimeter security system; to date, big data has no adequate security protection system.

3. Cryptographic protection: Encryption is required to protect big data and its sensitive information. However, this security measure is often ignored. Delicate information stored in the cloud without encryption becomes vulnerable to security threats. Constant encryption and decryption of big data slow down the system's speed, making it disadvantageous.

4. Mining of sensitive information: Big data is protected by perimeter-based security, which ensures that every point of entry and exit is secured. However, some unethical IT specialists can mine unprotected data entering the system. This poses a huge threat to the company because sensitive information gets leaked. The solution to protecting the data is to add extra perimeter. The security system can also benefit from anonymization. There will be no harm if personal data contains absent names, addresses, and phone numbers.

5. Granular access control struggles: Security is a process that needs different permissions and authorizations at various levels. Granularity allows one to restrict certain specific actions for someone while permitting the same to others. For example, while using some applications, we can click and read only according to the instructions given. We cannot change their default settings or get into the technical side of the application [6]. This is because all those are low-granularity systems where users are only allowed to read or write but not to access. In a high-granularity system, things will be vice versa, meaning users are given more access.

   Granular access controls limit the information people are allowed to see and access, setting the degree of security needed. This requires a strong authentication process. So, the decision regarding granular access control is a critical one, as we have to decide the limit of access for various users at various levels for different purposes without affecting the performance of the whole system. Regular granule auditing will help to recognize any malicious activity or cyber-attack, which is more advantageous.

6. Difficulties in data provenance: For effective access control, data should be classified accurately. For this, it is necessary to know its origin and order. Data provenance is the process of analyzing the origin and history, which is undergone to derive a given data instance. Generally speaking, data provenance means tracking the influence on a document rather than any single piece of data, or more specifically, it is the point of origin of the whole thing [7, 8].

   Data provenance requires a huge collection of data, which challenges security. The changes made in metadata were a difficult problem to manage. The main challenges include authentication, validation, and access control. Traceability of origin is another problem. Unauthorized changes in metadata will create a different wrong set of data which will not meet the particular information requirements [6]. Another problem is regarding information ownership, where reliability is questioned.

   Some other issues relating to data provenance include: problem with the software used for data storage, legal problems like Copywrite, real-time security monitoring problems, confidentiality of communication, problems related to encryption, and issues with cloud providers.

Manuscript received April 5, 2026
Manuscript revised April 20, 2026
https://doi.org/10.22937/IJCSNS.2026.26.4.15

7. Absence of security audit: As the data we are handling becomes big, the security risk we face will also be high.Regular security audits will help to manage this to a great extent. Here the security of the company's information system is evaluated as per the established criteria. Before this, ensure a good solution creator software that enables us to make the process easier. Security audits often determine regulatory compliance, specifying how organizations must deal with information [9]. A security audit will help the companies to ensure the reliability of the data they are handling for doing better business. A security audit is a part of the ongoing process of defining and maintaining strong security policies.

In general, a security audit will help to: safeguard the privacy of information, improve business, the balance between efficiency and data privacy, prevention of unauthorized access, uninterrupted workflows, and protection of computer application programs.
The cost of a security audit is a crucial factor to be considered while conducting the same. It will vary according to the nature of the business, the purpose of the audit, the type of the business, the number of employees involved in the audit process, the amount of information and applications to be analyzed, etc. The five simple, inexpensive steps by which we can conduct an internal security audit include: define audit, define threats, assess current security performance, prioritize (Risk Scoring), and formulate security solutions.

Other challenges of big data: Data analysis, Data capture, Data curation, Data Search, Data sharing, Data storage, Data transfer, Data visualization, Data querying, and Information Privacy. Table 1 shows security problem in big data.

Table 1. Security Problems in Big Data

Security Problem	Description
False or Fake Data Generation	Cyber hackers deliberately generate false data, duplicate it, and dump it into the data lake, causing malfunctions or false alarms. Robust fraud detection systems are required to mitigate such threats.
Non-Trusted Mappers	Malicious actors can manipulate mappers during the MapReduce process, introducing foreign mappers to corrupt data or steal information. Lack of adequate security measures exacerbates the issue.
Cryptographic Protection	Lack of consistent encryption leaves sensitive data vulnerable. While encryption improves security, frequent encryption and decryption can reduce system performance.
Mining of Sensitive Information	Perimeter-based security can be bypassed by unethical insiders, leading to sensitive data leaks. Enhanced perimeter security and data anonymization are proposed solutions.
Granular Access Control	Balancing high and low-granularity access is essential for restricting user permissions at various levels. Strong authentication and regular granule auditing are critical for preventing malicious activity.
Data Provenance	Understanding data origin and history is crucial for security. Issues include unauthorized metadata changes, challenges in authentication, and access control, which can lead to unreliable datasets.
Absence of Security Audits	Regular security audits help identify vulnerabilities, ensure regulatory compliance, and safeguard data privacy. However, audits require careful planning and resource allocation.
Other Challenges	Additional issues include data analysis, capture, curation, search, sharing, storage, transfer, visualization, querying, and ensuring information privacy.

6. Solutions for Big Data Security

Big Data environments are personally identifiable information (PII) and are therefore governed by corporate security and regulatory compliance policies. So security, security-conscious organizations need to control access and conduct an audit regularly for a reliable and sustainable future [10, 11]. Big data security is the collective term for all the measures and tools used to guard data and analytics processes against attacks, theft, or other malicious activities that could harm or negatively affect them. Big data security challenges seem to be multi-faceted for companies those who are operating on the cloud.

When working with sensitive data, it's vital to do more than just be careful. One must always be aware of the potential threats ahead as the attacks become more sophisticated and the information is more valuable. Some of the best practices for big data solutions include:

Manuscript received April 5, 2026
Manuscript revised April 20, 2026
https://doi.org/10.22937/IJCSNS.2026.26.4.15

- Security protocols should always be opt-out: When developing an algorithm or program, features tend to be opt-in—that is, they're optional until you decide they are necessary. Ensure the databases have enough security protocols [12]. When we want to remove a security feature, ensure there is no other way to solve the current problem. It will help to work in a situation where safety is not compromised.
- Formulation of good and fit control policy: The security policy must be apt for our type of business. The security policy of another business cannot be blindly copied into ours, and it will not work. While creating policies, it should be kept in mind that access will only be given to authorized users
- Monitor and Audit the Database Consistently: Keeping track of the interactions a database receives seems very hectic. This will occasionally lead to a considerable amount of damage. Try to find out if the database is breached; we have to identify the exact location
- Making the data anonymous: It should be necessary to eliminate all the delicate information from the huge data collection. Encryption methods can be used to protect the data. Real-time monitoring will also help safeguard the information.
- Add Multiple Layers of Security for the Database: Even though database servers are essential to business intelligence, they should be separated by as many degrees of security as possible. Commonly, database administrators are content to host both their web servers and database servers on the same platform. This may be convenient, but it significantly exposes user and business data to danger needlessly.
- Preservation of communication: Data shared must be confidential and upright. Separate servers and create secure bridges for communication. Store the information in encrypted form, which can handle traffic better while filtering out malicious access attempts. Always remove non-vital data from your database and store it offline, especially law-abiding records that are not necessary for the business's daily operations.
- Should have a strong audit philosophy: Auditing helps to collect much deeper-level data that can be vital to understanding breaches and blocking future infringements. Always perform root-cause analyses and comply with a variety of protocols and standards. Constant audits help prevent attacks by noticing irregularities and potential cracks in the system. The frequency of conducting an audit is an important factor to consider.
- Setting Root Access: Business intelligence relies on having free and fast access to data sets. On the other, databases must be able to protect the data they contain from being exposed to accidental or deliberate threats. The standard solution is to layer access privileges depending on users' requirements. Some may not need the ability to modify or manipulate data, while others might require more in-depth access to sets.
- Cloud service providers: If we opt to store the information with the cloud service providers, one should ensure they have sufficient security measures. Also, it is to be ensured that they are conducting periodic security audits and accept penalties if they are not conducted the same.
- Risk evaluation: Businesses should always consider the risk associated with the data they gather. If any customer or client data is collected, the firm must implement necessary privacy policies to preserve the data and, thereby, the client's privacy.
- Storage of big data: Another problem companies face is the storage of big data. The major problems include volume, velocity, and complexity of the data. 94% of the businesses claim that they will find an improvement in security after switching to the cloud.91% said that the cloud makes it easier to meet government compliance requirements [13, 14]. If this issue is managed efficiently, then the life cycle management of the business can be managed even better.
- Encryption opportunities: Normally, organizations find it very difficult to encrypt the data. Cloud generally acts so the encrypted data can be decrypted as per the requirements. Solutions for big data security is shown in Table 2.

Table 2. Solutions for Big Data Security

Solution	Description
Security Protocols as Optout	Ensure that security protocols are mandatory and cannot be removed unless absolutely necessary to solve a specific issue, safeguarding against compromised safety.
Good and Fit Control Policy	Tailor security policies to the specific business needs, ensuring that access is granted only to authorized users. Avoid copying policies blindly from other organizations.

Manuscript received April 5, 2026
Manuscript revised April 20, 2026
https://doi.org/10.22937/IJCSNS.2026.26.4.15

Monitor and Audit Database	Consistently monitor database interactions to identify breaches and pinpoint vulnerabilities. Regular auditing helps in damage control and breach prevention.
Anonymization of Data	Remove sensitive information from datasets using encryption and real time monitoring to safeguard data integrity and privacy.
Multiple Layers of Security	Separate database and web servers, adding multiple security layers to minimize exposure and risk.
Preservation of Communi-cation	Use encrypted communication channels and segregated servers to ensure confidentiality and integrity of data transmission. Store nonessential data offline.
Strong Audit Philosophy	Conduct regular, in-depth audits to identify potential system vulnerabilities, perform root-cause analyses, and prevent future attacks.
Layered Access Privileges	Implement layered access privileges to restrict user capabilities based on their roles and requirements, ensuring better control over data access.
Cloud Service Providers	Ensure cloud providers implement robust security measures, conduct regular audits, and agree to penalties for non-compliance.
Risk Evaluation	Evaluate risks associated with collected data and implement privacy policies to protect client and customer information.
Big Data Storage Management	Address storage challenges related to volume, velocity, and complexity by adopting cloud solutions that also meet compliance requirements.
Encryption Opportunities	Incorporate encryption mechanisms to secure data, even in cloud environments, with provisions for decryption as needed.

6.1 The Loss to Organizations on Big Data Piracy

- The data becomes unorganized and unable to be used: Big data, as the term goes, is huge in the make, and hence, if the data is not able to be managed and stored sequentially and added to it, if pilferage occurs, the data becomes unmanageable and unable to be used. It requires practical analysis, a meticulous system, and good storage capacity to handle the big data. Data appears both offline and online [15]. Hence, organizations must maintain, classify, and store big data properly with security measures to make those usable for the organizations.
- Storage of data is a concern: When the accumulation piles up, the data storage has a problem in the storing ofthe data, which multiplies at an exponential rate and enlarges in volume [4]. The traditional methods of storing data fail and the enterprise must shift to a cloud-based data storage system. If the organizations cannot store and manage the data well, they may lose a wealth of valuable information in a second, which may cause huge harm to the organization in terms of its data loss.
- The cost of losing revenue: With data, there is an attached valuable asset of the organization, which has a huge monetary value. This happens either because the face value of the data is great and it has some monetary weight attached to it, for example, the data of a bank, or it can also happen that acquiring and possessing the data will enable the company to generate revenue, for example, the data of certain encrypted information sites. Hence, losing the big data poses a direct threat to the monetary status.
- The loss of privacy in the information: Companies store the data of their clients and their stakeholders in big data storage. Such systems may also be used by government organizations to commute services or business. However, the data stored may be sensitive, and data theft may put the privacy of the clients or the stakeholders at a loss. It will be considered a serious breach of trust by the people whose privacy is tossed off, and nowadays, there are legal suits that follow for such infringements. The parent company from where the data has been lost cannot shirk its responsibility for the negligence of data keeping, and faith in such a system will be lost forever. Once data piracy occurs, customers and stakeholders will lose faith in the system, and trust, once lost, will take a million efforts to regain in vain.

6.2 Fundamental steps in strong big data

1. Get security measures from the start: The security team is responsible for providing their expertise at each step in the development process. The security team must be approachable, open with their knowledge, and committed to finding a custom solution for securing big data technologies. The data analytics team must recognize its obligation to incorporate robust security measures into its big data innovations [16].

Manuscript received April 5, 2026
Manuscript revised April 20, 2026
https://doi.org/10.22937/IJCSNS.2026.26.4.15

2. Start with objectives rather than solutions: Too often, security teams prescribe simplistic solutions as requirements for any system an organization uses despite the fact those solutions won't successfully protect complex big data systems. Adequate big data security differs from normal operations and cannot be subjected to the same standard. A big data environment can't be secured with just one solution [15]. Organizations must understand that only a customized blend of tactics can completely manage the risk. Security teams should adjust their thinking regarding big data security efforts. They can begin by asking what specific security objectives your team tries to achieve. From there, they can work backward to find custom and alternative solutions to secure the environment.

3. Customize the solution: There is no one-stop shop for big data security. Big data technologies are a bunch of opensource frameworks stitched together to fill a specific need. That makes creating a security solution complicated. As big data platforms are treated more like custom applications and less like databases, you have a greater chance of using the appropriate security approach. The data analytics and security teams need to understand the low level architecture to ensure they're considering all possible threats [17, 18]. Big data platforms are too complex to be secured with a one-size-fits-all solution. To address the complex security requirements of big data platforms, organizations need to customize a stack of tactics that address the security objectives identified at the beginning of the process.

6.3 Safeguarding analytics and ensuring integrity

Big data security analytics is a collection of large and complex security data sets that become difficult (or impossible) to process using on-hand database management tools or traditional security data processing applications. The test with big data is that the unstructured nature of the knowledge makes it even more difficult to classify, model, and map the data when it is taken and deposited. The obstacle is made least by the fact that the data usually comes from outside sources, often making it difficult to confirm its correctness. If they apprehend all the knowledge available, they risk wasting time and support processing data that will add little or no value to the industry. Another hurdle in the case of big data is that you can have a big category of users, each needing access to a distinct subset of information. This means the encryption solution you wanted to preserve the data must exhibit this new reality. Access control to the data will also require being more granular to ensure people can only obtain the information they are approved to see [16].

All these technological leaps across many industries have solid support in the form of big data. Progress will remain to help build a larger society through smarter processes. But to eventually benefit from these trends, organizations must fully understand how they are practiced and how they can help accomplish business intentions.

The reason for the concern over the big data security challenge is its increasing necessity. For convenience of usage and storage, almost every organization uses big data. Gone are the days when large organizations only handled big data.

Nowadays, even small or medium-sized companies need big data because it saves costs and has more flexibility in usage and storage systems than general data. However, handling big data is a major issue, and its security needs to be, especially as it contains information about the nitty-grit of an organization's functioning and public information. The information stored may have valuable data, for example, the data in public banks or, recently, the scam of Tesco in data pilferage. The stored information may be sensitive or crucial for the maintenance of a system or the organization [19]. These should and are secured by transaction logs, but there can be causes of theft or sabotage. For example, someone among the employees has to be trusted with the particulars of the transaction log, and leakage may arise from there purposely or through errors. In either case, the party that has stored the sensitive data is at a loss.

The main factors maintaining big data are the End-Point devices. The works' functioning, storage, and processing are carried out with the help of data input, which is provided by the end-points. Organizations have to be particularly cautious in securing these and maintaining the secrecy of such data.

However, this security is lacking, particularly in a distributed framework like the Map Reduce functions of systems like Hadoop, and from there, the security issues creep in. It is advisable to maintain security audit checks, but most organizations skip it on the plea that they have a dearth of time and space due to the large handling of data, which becomes the bone of the mishap [1, 2]. In real time, the security checks for data are done by very few organizations. An intelligent option to protect the data is to make the data device secure. Yet there happen to be very vulnerable cases with the security of the data device so the best option is to encrypt the access control systems with the data storage devices security as well.

The data provenance has to be maintained diligently. In classifying the data, the origin has to be listed. It is pertinent to be aware of the origin of the data, authenticate it before using it and validate it, and have an

Manuscript received April 5, 2026
Manuscript revised April 20, 2026
https://doi.org/10.22937/IJCSNS.2026.26.4.15

access control system with a periodically checked mechanism [19]. Regular auditing is also beneficial for the analysis of different kinds of logs. Hence, periodic audits can keep in check and recognize cyber-attacks easily.

The granular access control system of big data stored by the Hadoop distributed file system of NoSQL should require strong validation and a check of authenticity, making the access control system a must and foremost requirement [20]. Data stores like NoSQL are highly vulnerable to security threats, which can lead to privacy breaches. The main drawback is that data encryption during tagging or the log and distribution to various groups is not possible. Security threats also arise from this.

Considering several consequences, encryption, data logs, security audits at regular intervals, and analysis of the source of data used are the current options to check big data vulnerability in today's world. Big data protection has yet to come from the risk of hackers and cybercriminals. So, the users and companies utilizing big data must be constantly aware and cautious while handling it.

7 Conclusion

In this paper, we present big data security, which involves managing large and complex datasets that are difficult to process with traditional tools, especially when the data is unstructured. One challenge is the uncertainty around the accuracy of external data sources, which can lead to wasted resources if not properly filtered. Furthermore, organizations must implement granular encryption and access control systems to ensure that users only access the data they are authorized to view. As the need for big data grows, even smaller organizations are adopting it for its cost savings and flexibility, making security even more critical to protect sensitive information. We also highlight that to safeguard big data, particularly in systems like Hadoop, organizations must focus on securing endpoints, conducting regular security audits, and maintaining the provenance of data to track its origin and ensure its authenticity. Data encryption, access control, and periodic checks are essential practices to prevent breaches. NoSQL data stores are particularly vulnerable due to their difficulty in encrypting data during processing and distribution, creating security risks. As cyberattacks and privacy breaches remain major concerns, organizations must remain vigilant, continuously monitoring and securing their big data infrastructure. Future work will involve exploring advanced encryption methods, automating security audits, and improving the efficiency of data provenance tracking to further enhance big data security.

Manuscript received April 5, 2026
Manuscript revised April 20, 2026
https://doi.org/10.22937/IJCSNS.2026.26.4.15

References

[1] Kamalinder Kaur and Vishal Bharti. A survey on big data—its challenges and solution from vendors. Big Data Processing Using Spark in Cloud, pages 1–22, 2019.

[2] Muhammad Naeem, Tauseef Jamal, Jorge Diaz-Martinez, Shariq Aziz Butt, Nicolo Montesano, Muhammad Imran Tariq, Emiro De-la Hoz-Franco, and Ethel De-La-Hoz-Valdiris. Trends and future perspective challenges in big data. In Advances in Intelligent Data Analysis and Applications: Proceeding of the Sixth Euro-China Conference on Intelligent Data Analysis and Applications, 15–18 October 2019, Arad, Romania, pages 309–325. Springer, 2022.

[3] Paul Quinn and Liam Quinn. Big genetic data and its big data protection challenges. Computer law & security review, 34(5):1000–1018, 2018.

[4] Damilola Oluwaseun Ogundipe. The impact of big data on healthcare product development: A theoretical and analytical review. International Medical Science Research Journal, 4(3):341–360, 2024.

[5] Big Data. What it is and why it matters. URL: https://www. sas. com/en us/insights/big-data/what-is-bigdata. html, 2017.

[6] Oluwatosin Reis, Nkechi Emmanuella Eneh, Benedicta Ehimuan, Anthony Anyanwu, Temidayo Olorunsogo, and Temitayo Oluwaseun Abrahams. Privacy law challenges in the digital age: a global review of legislation and enforcement. International Journal of Applied Research in Social Sciences, 6(1):73–88, 2024.

[7] Pratibha Gautam and YP Singh. Impact of data mining on big data analytics: Challenges and opportunities. International Journal of Computer Trends and Technology, 57(1):19–26, 2018.

[8] Alberico Travassos Ros´ ario and Joana Carmo Dias. How has data-driven marketing evolved: Challenges and oppor- ´ tunities with emerging technologies. International Journal of Information Management Data Insights, 3(2):100203, 2023.

[9] Philip Olaseni Shoetan, Adedoyin Tolulope Oyewole, Chinwe Chinazo Okoye, and Onyeka Chrisanctus Ofodile. Reviewing the role of big data analytics in financial fraud detection. Finance & Accounting Research Journal, 6(3):384–394, 2024.

[10] Arnab Laha. Statistical challenges with big data in management science. Big Data Analytics: Methods and Applications, pages 41–55, 2016.

[11] Iqbal H Sarker. Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective. SN Computer Science, 2(5):377, 2021.

[12] Ali Emrouznejad and Vincent Charles. Big data for the greater good. Springer, 2019.

[13] Alex Bekker. Buried under big data: Security issues, challenges, concerns. Science Soft, 2018.

[14] PDNK Kommisetty et al. Leading the future: Big data solutions, cloud migration, and ai-driven decision-making in modern enterprises. Educational Administration: Theory and Practice, 28(03):352–364, 2022.

[15] Shuguang Cui, Alfred O Hero III, Zhi-Quan Luo, and Jose MF Moura.´ Big data over networks. Cambridge University Press, 2016.
[16] Geetha Kurikala et al. A survey on security intelligence on bigdata. The International journal of analytical and experimental modal analysis, 13:1405–1408, 2021.
[17] Zal Azmi. Opportunities and security challenges of big data. In Current and Emerging Trends in Cyber Operations: Policy, Strategy and Practice, pages 181–197. Springer, 2015.
[18] Madhusanka Liyanage, An Braeken, Shahriar Shahabuddin, and Pasika Ranaweera. Open ran security: Challenges and opportunities. Journal of Network and Computer Applications, 214:103621, 2023.
[19] Peter Ghavami. Big data management: Data governance principles for big data analytics. Walter de Gruyter GmbH & Co KG, 2020.
[20] Mikhail Yurievich Ryabchikov and Elena Sergeevna Ryabchikova. Big data-driven assessment of proposals to improve enterprise flexibility through control options untested in practice. Global Journal of Flexible Systems Management, pages 1–32, 2022.

Hassan Mistareehi, is an assistant professor in the Department of Computer Science and Information Systems at Murray State University. He has taught courses such as Computer Security, Database Management Systems, and Programming in Python. He received his Ph.D. degree in Computer Science from the University of Kentucky.
He received his B.S. and M.S. degrees in Computer Science from Jordan University of Science and Technology, Jordan. His research focuses on the security and privacy of vehicular ad hoc networks (VANET), Vehicular Cloud (VC), Internet of Things (IoT), Cloud Computing, Edge Computing, Ad Hoc Networks, and Sensor Networks.
He has published numerous research papers in journals and conferences such as the Internet of Things Journal, MDPI network, and IEEE Global Communications Conference. Dr. Mistareehi also serves as an ad-hoc reviewer for various journals.

Abdulrahman Yarali, Professor of Cybersecurity & Network Management at school of Engineering, Murray State University (MSU), KY, USA. During tenure in the wireless industry, Dr. Yarali led research teams working on location, completing numerous contracts in wireless mobile communications systems design, implementation, and optimization for leading telecommunications companies such as AT&T, NORTEL, and Sprint PCS in the U.S. His interest continues to be focused on the field of higher generations of wireless mobile communications systems, cybersecurity, small satellites, IoT, Cloud, AI, ML, and smart grid infrastructures. He has presented articles, lectures, and keynote presentations in mobile communications networking and security in North America, the Middle East, and Europe. Dr. Yarali is the author of nine books on 4G, 5G, IoT, cloud and big data, public safety networks, intelligent connectivity with AI and IoT, and Cybersecurity and Forensics.

Jason Owen, Assistant Professor of Computer Science at the Beunerfiend School of Business at Murray State University in Murray, KY, USA. Dr. Owen completed a DBA in Information Systems from Liberty University in Lynchburg VA along with two masters degrees. While working on his doctorate degree Dr Owen worked in the Computer Science Industry in various roles including software application developer, chief business development officer, senior software engineer, delivery lead, and scrum master. Along with his extensive field experience, Dr Owen focuses his research in practical applications of computer science in business, generative AI, optimization in agile team construction, and computer science pedagogy. He has authored, lectured and given conference presentations on various topics relating to the computer science and information systems fields. Dr. Owen is the author of “Biblical Technology: Building a Decision Support System for Optimal Team Construction”. Dr. Owen also published multiple research papers on generative AI, optimal agile team development, unconventional uses of AI and ML, and advancing computer science pedagogy through research on optimal times to incorporate various topics in the computer science degree journey

Manuscript received April 5, 2026
Manuscript revised April 20, 2026
https://doi.org/10.22937/IJCSNS.2026.26.4.15