Wasit Journal for Pure Science

Journal Homepage: https://wjps.uowasit.edu.iq/index.php/wjps/index
e-ISSN: 2790-5241 p-ISSN: 2790-5233

Security and Information Technology for Big Data

Haifaa Jassim Muhasin[1]*, Ali Yahya Gheni [2], Hiba Adil Yousif [3]

[1,2,3]Computer Science Department, College of Education for Pure Science, Ibn Al-Haitham University of Baghdad, IRAQ

*Corresponding Author: Haifaa Jassim Muhasin

DOI: https://doi.org/ 10.31185/wjps.418
Received 01 Mat 2024; Accepted 12 Jun 2024; Available online 30 Jun 2024

ABSTRACT: Big data has grown in importance in a variety of corporate and academic fields as a result of substantial changes in human behavior and lifestyles. While this focus simplifies data management, big data is regarded as important in many industrial areas because it can be merged with other types of data. Big data can be managed alongside other types of data, particularly in large organizations that handle a range of data types. This type of management contributes to the improvement and development of the organizational structure, as well as raising the skill level of employees in these organizations. Despite these benefits, there are numerous concerns, particularly throughout the stages of data gathering and storage, when there is a risk of hacking or leakage, as well as the difficulty of sorting and identifying data. For these reasons, big data protection is regarded as one of the most important issues discussed in this paper in order to develop an appropriate framework for protecting big data related to companies, as well as methods for dealing with this data in order to develop an appropriate strategy to protect big data security and privacy.

Keywords: big data, big data security, privacy, information technology, data management

1. INTRODUCTION

Big data is one of the most prominent subjects in information systems today, but it is a relatively new term in information technology. Big data remains an innovative technology that serves as the basis for a wide range of applications [1], and it is described using four key elements: volume, diversity, velocity, and veracity [2]. The four Vs represent the most significant elements and characteristics of big data. Whereas volume refers to the amount of big data that grows over time, variety represents the many forms of data, velocity indicates the speed at which big data is generated, and veracity refers to the correctness and reliability required to maintain data accuracy [3].

This innovative approach to data management enables companies to acquire and manage a greater volume of data. Many firms use big data to evaluate and find trends in order to better understand various topics and fields. For example, consider the industrial and medical sectors. Medical researchers seek to investigate illness cases quickly and efficiently. In industrial sectors, research is underway to discover the best strategies for processing enormous amounts of data stored across several production branches [4].

Other characteristics that can be used to describe big data include valence, validity, variability, and volatility, with valence referring to the degree of limitation in massive data sets and the presence of relationships between them. While validity refers to the validity and dependability of a huge amount of data used to make appropriate and proper decisions, Variability refers to unanticipated changes that arise in data, resulting in an inability to regulate and manage it correctly. While volatility refers to storage periods and validity periods, it also represents the time when the data is effective [5]. Figure 1 shows these qualities and features for big data security.

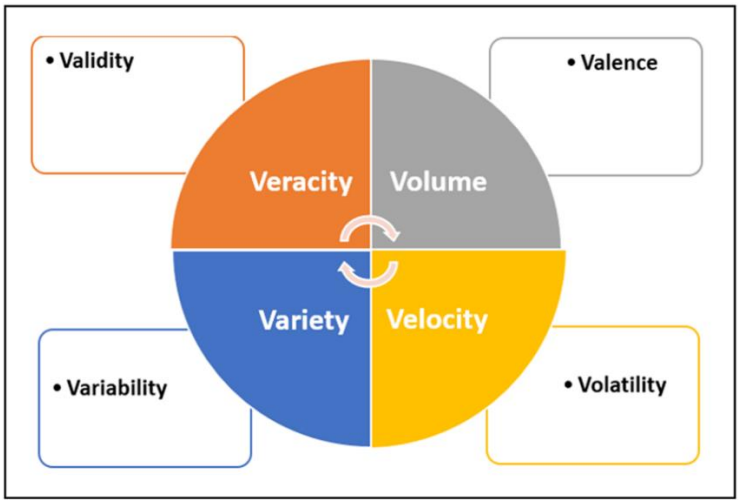

Figure 1. - Security big data features

Big data is new to novel and developing enterprises, but many major organizations have been waiting for it for a long time. Many firms recognize that big data is an important aspect of data development, and new forms of data are continually introduced to the systems utilized by these companies [6].This means that many companies dealt with big data prior to its current form. Big data has been one of the most important factors in production since its introduction into numerous industries, and with technological advancements, valuable information may be recognized, serving as a crucial reference for decision-making [7].

The current use of big data has effectively led to advancements in how to deal with data sources, but it has also produced numerous dangers to privacy and data security [8]. To mitigate these concerns, efforts are being made to raise awareness of the necessity of data protection before using new technology to assure data protection [9].

Big data is a wonderful motivator for IT staff. Big data can provide the necessary data to enable access to the best ways to develop data-handling methods, such as product marketing, planning, financial management, and crucial efficiency. The Speed in big data refers to both the rate at which data is connected and the rate at which data is effective [10]. Traditional frameworks cannot assess data that is constantly moving and unstable.

2. THE SECURITY ISSUES FOR BIG DATA

- The problem of connecting numerous remote data regions poses a significant challenge in how to collect data from various areas and enable the institution to manage it.
- Organizational cooperation is an issue that firms have been working to overcome by collaboration among diverse departments, such as information technology, engineering, marketing, and finance.
- Protection and security in big data
  Big data security is a major concern due to the potential for personal data to reveal information about the data. This has a negative impact on data and increases the risk of hacking [10].
- Access and share data
  One of the most essential issues for big data is access speed and sharing capabilities, with the most critical being that the data is usable and correct [10].
- Choosing relevant data is one of the most essential issues that companies face, because with such vast amounts of data, distinguishing between important and relevant data becomes a vital activity [11].

As the volume of information expands, so does the prospect of targeting it for different objectives. The same sensitive information that was compromised can now be utilized to conduct an attack [12]. Large amounts of data include all activities and events that could expose it to a digital threat or attack. This information at risk includes information on the user, the method of access, runtime information, device information, clients, record information, file contents, and access

to multiple applications. If more log information is maintained in companies, it enhances the probability of discriminating while also safeguarding the organization from digital attacks by discovering faults in the information and unexpected circumstances, resulting in increased protection against threats.

Big data is used to collect a wealth of information about businesses, including information about clients and prospective customers. However, there are numerous questions about data ownership rights and who has access to this data. Most companies want to ensure that their and their customers' data is secure and under their control. Customers want to know how much personal information is stored in large databases.

3. METHOD

3.1 BIG DATA MODEL

Big data refers to the vast volumes of data collected by organizations and governments concerning various locations, as well as the data generated by a variety of clients and tools, which must be kept and processed on large servers [13]. The major purpose of saving time with big data is to be able to connect with consumers and data locations in real time using customer-derived procedures and data. This entails receiving, gathering, processing, and evaluating data from multiple locations quickly.

By taking everything into account, it is feasible to create a system framework for collecting data from various geographic areas. The model demonstrates how to handle large amounts of data in one of the Ministry of Industry and Minerals' firms. This corporation has branches and factories in Baghdad and other Iraqi cities, as illustrated in Figure 2, which depicts the types of production factories. Data is sent between the company's headquarters and the ministry, as well as among the company's branches and factories in Baghdad and the governorates. Figure 3 depicts the framework for big data transfer between the Ministry of Industry, the corporate center, and governorate production factories.

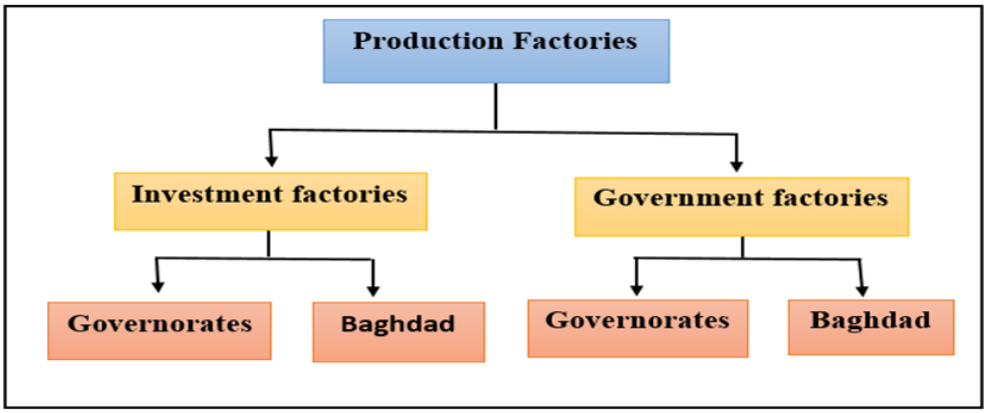

Figure 2. - Types of production factories

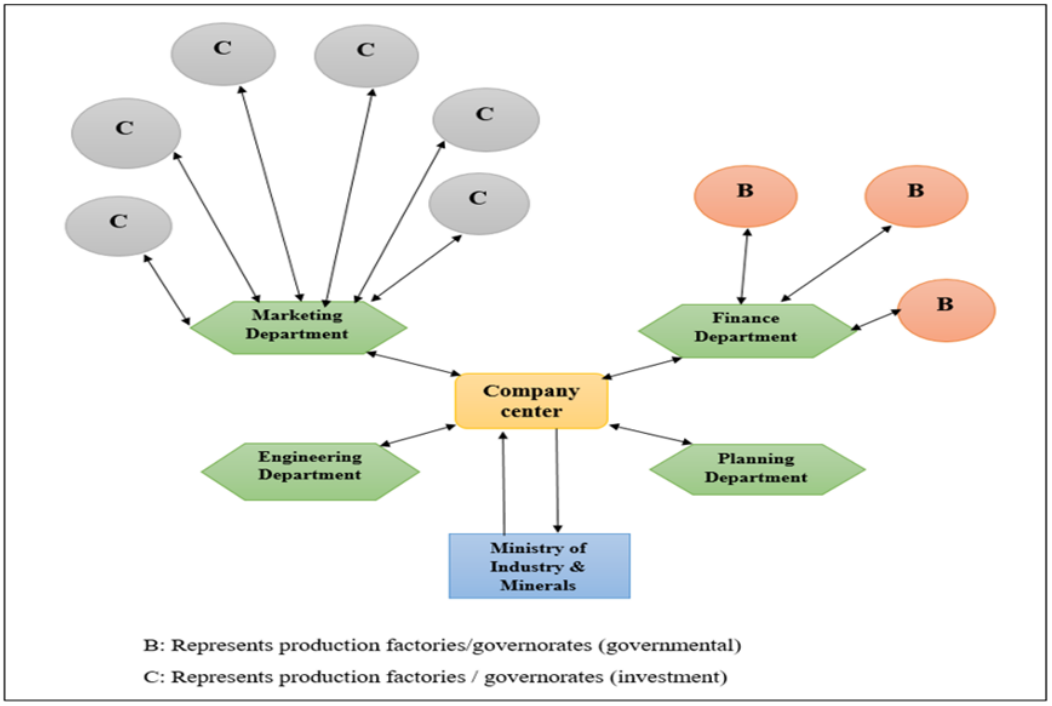

Figure 3. - The framework between the Ministry of Industry, the company center, and the production factories in the governorates

Figure 4 depicts the framework for large data transfers between Baghdad's production factories and the company's center. The data is transported from the production factories to the main company's marketing department in order to coordinate the marketing process and complete the electronic marketing forms. Following the completion of the procedures, financial data is forwarded to the finance department.

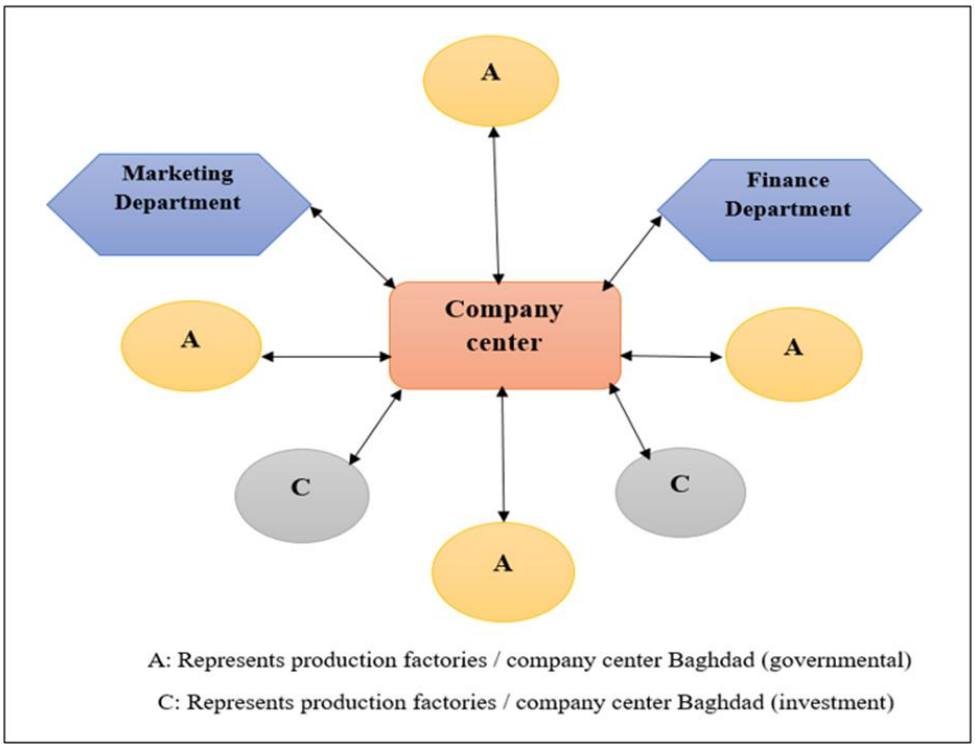

Figure 4. - The internal framework between the company's center and the production factories inside Baghdad

Figure 5 depicts the interaction between government factories in Baghdad and the governorates with the company's headquarters' financial and marketing departments. Figure 6 illustrates the interaction between investment factories in Baghdad and the governorates, as well as the company's financial and marketing departments.

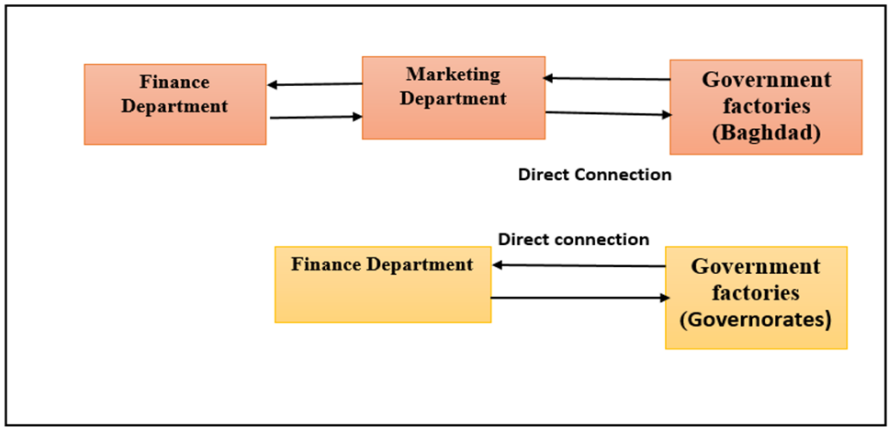

Figure 5. - shows the connection between government factories in Baghdad, the governorates, and the financial and marketing departments at the company's headquarters.

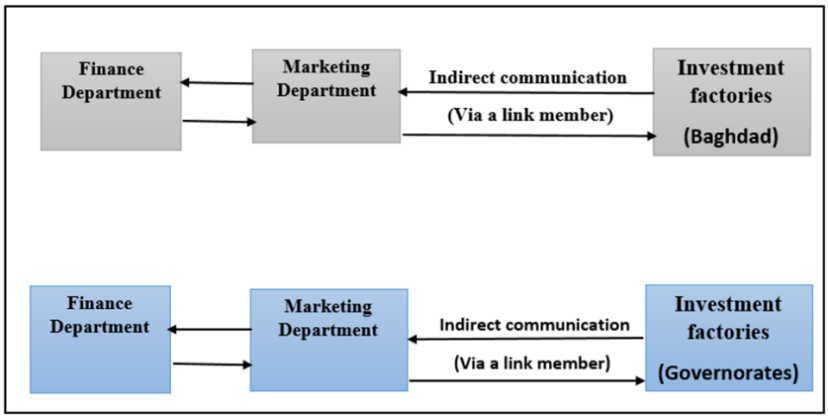

Figure 6. - shows the relationship between investment factories in Baghdad, the governorates, and the financial and marketing departments at the company's headquarters.

The data transmitted between the company's departments and production factories is divided into the following categories:

1. The information transmitted between factories and the marketing department includes:
   a) Production quantity.
   b) Product prices.
   c) Raw materials (types, specifications, prices).
   d) The type and quantity of marketed materials.
   e) Information about the transportation of marketed materials, authorized persons, and information about the transport vehicle.

Table 1 depicts a record of data transferred between factories and the marketing department.

Table 1. - **Record of data between factories and the marketing department**

The factory	production quantity	Product price	Raw materials			Marketing materials		Transportation information and authorized persons
			Type	Specifications	The price	Type	Quantity	

2. The marketing and finance departments share information, such as
   a) Production quantities.
   b) Product kind.
   c) Expenditures (raw materials and other costs).
   d) Net profit.
   e) Beneficiary party (per contract).

   Table 2 illustrates a record of data transferred between the marketing and finance departments.

Table 2. - **Record data between marketing and finance departments**

The factory	Product Type	production quantity	Expenditures		Net profit	The beneficiary party according to the contract
			raw materials	Other expenses		

3. Data exchanged between the Finance Department and the Ministry.

   Periodic trial balance

   a) Sales quantity
   b) Operating budget expenditures
   c) Classification of exchange axes
   d) The revenue generated

   Table 3 shows a record of the contents of the periodic trial balance.

Table 3. - **Record the contents of the periodic trial balance**

The factory	Sales quantity	Operating budget expenditures	Classification of exchange axes	The revenue generated

4. Data shared between the Engineering Department and the Ministry of Industry.
   a) The quantity of electrical power consumed by all factories (excluding investment ones).
   b) The amount of fuel consumed.

   Information is collected from factories and forwarded to the firm center, followed by a full report to the Ministry of Industry.

   Table 4 depicts the data interchange between the Engineering Department and the Ministry of Industry.

Table 4. - **shows the record data exchanged between the Engineering Department and the Ministry.**

Government factory	The amount of electrical power consumption	The amount of fuel consumption

5. Data sent between the Marketing Department and the Security Authorities.
   a) The quantity and type of products marketed.
   b) The intended recipient (to whom the production is marketed).
   c) Security approval (itinerary).
   d) Details regarding the transportation vehicles and drivers.

   This information is critical to ensuring that products are not smuggled across borders.

   Table 5 depicts a record of data transferred between the Marketing department and security agencies.

Table 5. - **shows a record of data transferred between the Marketing department and security agencies.**

The factory	The marketed product		The beneficiary party	Security approvals	The transport vehicle	The driver name
	Quantity	Type				

6. Data were transferred between the Planning Department and the Ministry.

Create tables of fuel material amounts and annual demands after gathering information from the company's Baghdad branches.

Using the preceding information, a framework for an electronic work system will be developed to process the massive data exchanged between the company's center and its affiliated divisions, with numerous authorities ensuring that data is delivered and received by authorized personnel. The system center is located within the company, and system information might be stored in the cloud. The system is accessed using the user name and secret access code. Accessing data necessitates the usage of numerous resources as well as user authentication, and using this strategy in access control is a critical tool for preventing and detecting internal system intrusions [14]. All data is saved in a database and classified based on the type of factory and department that has access to it. The security of the big data system is the responsibility of all system participants, including managers, security officials, users, and consumers. Indeed, many security threats originate from system staff and users. As a result, the major goal is to increase everyone's security awareness and improve security procedures across all system entities. As a result, everyone must follow and comply with security laws and standards [15]**.**

On this basis, a record will be created that contains all common information about materials and factories, such as the type of product, the producing party, the quantity of raw materials, the beneficiary party, the amount of fuel consumed, and the consumption of electrical power. This record is displayed in Table 6.

Table 6. - **The main record in which information is collected from sub -databases to the main database**

Factory type		Location		Marketed product (sales)		Raw materials			The beneficiary party	Consumption	
Investment	Government	Governorates	Baghdad	Quantity	Type	Type	Specifications	The price		electrical power	fuel

The primary goal of companies is to provide services using big data while ensuring data security and privacy. After achieving this goal, companies continue to look for ways to provide these services at a cheaper cost and in less time [6]. The new framework model proposed in this research, shown in Figure 6, uses a centralized approach to provide big data services from the company's headquarters by taking the necessary data and information from databases in the company's branches and linking them to the central database at the company's headquarters.

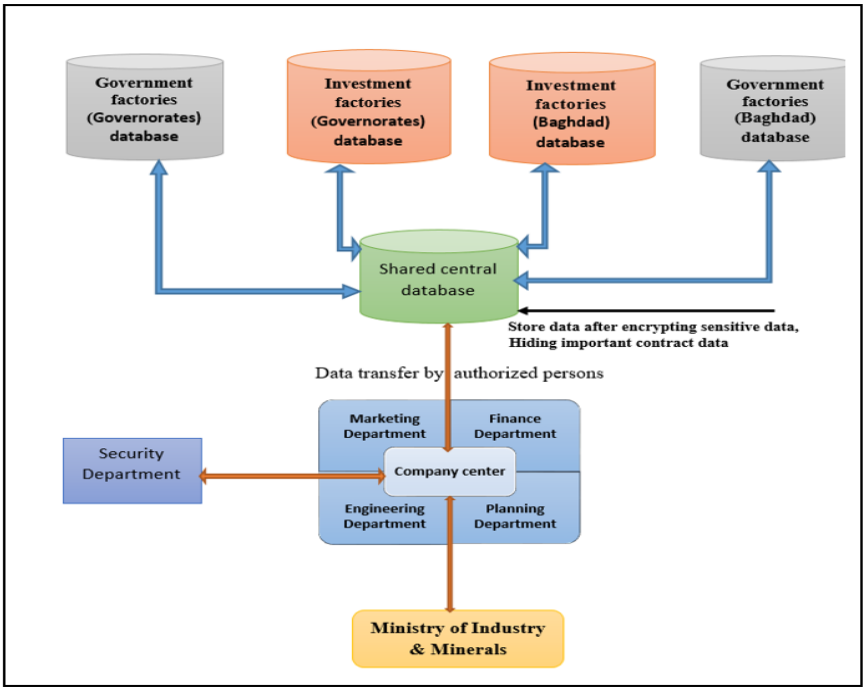

Figure 6. - A big data security framework for a company inside the Ministry of Industry and Minerals.

4. COMPARISON OF RESULTS AND DISCUSSION:

Companies' key goal is to provide services based on big data while maintaining data security and privacy standards. After accomplishing this goal, the organization continues to look for new ways to deliver these services at a lower cost and in less time. Previous research has produced a variety of solutions, but they have not met all needs. As a result, the proposed new framework uses measures to ensure security, save costs, and speed up the data collection process while maintaining the goals of the organizations using it.

In terms of the framework model now in use in businesses, the company's work has been implemented in the company's numerous branches in Baghdad and the governorates using a decentralized method (see figures 3 and 4). While the new secure big data framework model in figure 6 takes a centralized approach to providing big data services from the company's main center, it does so by extracting data and information from the company's branch databases and linking it to the central database at the company's headquarters. This approach ensures data security within the organization, and it is possible to utilize encryption for sensitive data, as well as the method of hiding important contract data so that only authorized personnel can access it.

4.1 STAGES IN DEVELOPING THE PROPOSED FRAMEWORK:

According to the new framework presented in Figure 6, the following are the actions, stages, and procedures that data goes through in the proposed framework to assure data security and privacy:

1. Data collection: Data is gathered from sub-databases of factories both inside and outside Baghdad.
2. Data storage sorting and classification: Received data is sorted and classified based on its source.
3. Create the main record from data obtained from sub-databases: The data is received after being sorted in the previous step, and the necessary information is saved in the main record.
4. Security methods for incoming data.
   a. Procedure for hiding essential investment contract information.
   b. Encrypting sensitive data such as material quantity and kind, as well as beneficiary information.
   c. Authorization Procedures
5. Exchange information with the company center.

Based on the stages of development, the suggested framework contributes to big data security as well as rapid access to the essential data at the lowest cost. The fundamental purpose of the proposed framework in this paper is to assure big data security by specifying permissions and allowing only authorized personnel to enter the system via authorization and entry procedures, as well as methods for hiding crucial data related to investment contracts. And, to reduce the costs of data transfer and the time it takes for data to arrive, the data in the proposed framework is now collected in a single database, with authorized persons controlling the procedures for transferring and adding data.
This technique will help to lower the cost and time required to process this data while maintaining its security. Using this framework, the company has made an appropriate decision to ensure the security of big data related to the company's various branches through the information technology unit at the company's center, as well as to adopt security, economic, and technical standards that help improve resources and production while maintaining confidentiality.

5. CONCLUSION

Big data is an essential part of our daily lives, particularly in the modern era. The study of big data and its relationships has helped to strengthen medical and industrial data, and big data has been integrated into industrial systems to take advantage of its benefits and aspects. Despite these traits, there are hazards associated with big data security that users should be aware of and handle, one of which is data privacy and confidentiality. This study presents a framework model that is based on gathering large amounts of data for a firm from many locations and storing it in a single database to ensure its security and regulate its use by authorized personnel. This framework is regarded as a model that may be utilized for any form of large data.

REFERENCES

[1] A.M., Ghimeș, V., Patriciu. "A survey of Big Data Analytics Techniques in Cyber Security," *International Journal of Latest Research in Engineering and Technology (IJLRET) ISSN: 2454-5031*, Volume 03 - Issue 02, pp. *21-25, February 2017.*

[2] A., Attaallah, H., Alsuhabi, S., Shukla, R., Kumar, B., K., Gupta, and R., A., Khan. "Analyzing the Big Data Security Through a Unified Decision-Making Approach," *Intelligent Automation & Soft Computing* 32, no. 2, 2022.

[3] S., Koley. "6 Big data security issues with challenges and solutions," Big Data Security 3: 95, 2019.

[4] D., Cole, J., Nelson, & B., McDaniel. "Benefits and risks of big data," 2015.

[5] Y., Perwej. "An experiential study of the big data," *for published in the International Transaction of Electrical and Computer Engineers System (ITECES), USA* 4, no. 1, pp.14-25, 2017 .

[6] T. H., Davenport, & J., Dyché. "Big data in big companies," *International Institute for Analytics* 3, no. 1-31, 2013.

[7] D.N.K., Gupta. "Addressing big data security issues and challenges," *International Journal of Computer Engineering & Technology* 9, no. 4, pp. 229-237, 2018.

[8] D., Zhang. "Big data security and privacy protection," In *8th international conference on management and computer science (ICMCS 2018)*, pp. 275-278. Atlantis Press, 2018.

[9] H. J., Muhasin, A. Y.,Gheni, & H. A. Yousif. "Proposed model for data protection in information systems of government institutions," *Bulletin of Electrical Engineering and Informatics* 11, no. 3, pp. 1715-1722, 2022.

[10] D., Trivedi, R., Tushar, Karishma "TITLE: A study on Security of Big data," *IJARIIE-ISSN (O)*-2395-439, Vol-3 Issue-5, 2017.

[11] R., Toshniwal, K. G., Dastidar, and A., Nath. "Big data security issues and challenges," *International Journal of Innovative Research in Advanced Engineering (IJIRAE)* 2, no. 2, 2015.

[12] H. J., Muhasin, R., Atan, M.A., Jabar, & S., binti Abdullah. "Factors and model for sensitive data management and protection in information systems' decision of cloud environment," *Journal of Theoretical and Applied Information Technology*, Volume 96, Issue 24, pp. 8097 – 8108, 2018.

[13] H., J., Muhasin, R., Atan, & S.,binti Abdullah. "Cloud computing sensitive data protection using multi layered approach," In *2016 2nd International Conference on Science in Information Technology (ICSITech)*, pp. 69-73. IEEE, 2016.

[14] H. J., Muhasin, R., Atan, M. A., Jabar, S., Abdullah, & S. Kasim. "Multilayered Framework to enhance management information systems decision on sensitive data in cloud computing environment," *JOIV: International Journal on Informatics Visualization* 1, no. 4-2, pp. 179-183, 2017.

[15] F.Z., Benjelloun, and A.A., Lahcen. "Big data security: challenges, recommendations and solutions," In *Web Services: Concepts, Methodologies, Tools, and Applications*, pp. 25-38. IGI Global, 2019.