INFORMATION SECURITY CHALLENGES IN BIG DATA: CRITICAL REVIEWING STUDY

Shahad Ahmed Abdulgaffar

Information Science Department,
King Abdulaziz University, Jeddah, ***SAUDI ARABIA***

ABSTRACT

Big data has revolutionized knowledge production as it represents a very large amount of data that exceeds the ability of traditional computing devices to process it. Indeed, the past few decades have witnessed the emergence of new methods of producing, storing and analyzing data to process their huge amount, culminating in the emergence of the field of data science, which combines computational, algorithmic, statistical and mathematical techniques to extrapolate knowledge from big data [1]. The open data movement has also encouraged the sharing of heterogeneous search data and linking it across large digital infrastructures. The paper aimed at shedding light on the challenges facing big data, especially the information security challenge, to try to identify the challenges and solutions that contribute to reducing challenges.

In fact, the lack of independent and detailed studies of these challenges that focus on practical solutions was the motive for conducting this paper. It aims at identifying and collecting the intellectual production, specifying the deficiencies, and looking out for future studies issues bearing in mind the importance of information security and its impact on the continuity and development of technology and facilitating the preservation, sharing and extraction of knowledge.

In this paper, the critical literary methodology was followed to specify the achieved studies and literature in the same issue and thus generalize the benefits and find solutions. Were analyzed to achieve the goal of identifying the challenges facing big data.

One of the most prominent findings of the paper after reviewing the literature and studies shows that the most prominent challenge of big data is the challenge of information security, especially personal data. A number of studies have suggested heading to encryption as one of the solutions and the principle of defining the purpose, restricting use, multi-layer security, and applying Block chain technology.

This goes hand in hand with the aim of the paper to define the challenges of information security in big data, thus contributing to finding appropriate solutions *by programmers and specialists, and taking advantage of the characteristics and advantages of big data in saving intellectual production and making it accessible.*

This paper recommended the necessity of completing specialized studies in dealing with big data challenges, especially information security, according to actual applicable systems. Leveraging technology such as Hadoop, machine learning, and Block chain to solve big data problems.

Keywords: *Information Security, Big Data, SMB'S, Symantec Protection Suite (SPS), Big and Open Linked Data (BOLD), An advanced persistent threat, Block chain*

INTRODUCTION

Big data has revolutionized knowledge production as it represents a very large amount of data that exceeds the ability of traditional computing devices to process it. The past few decades have witnessed the emergence of new methods of producing, storing and analyzing data to process its huge amount, culminating in the emergence of the field of data science, which combines computational, algorithmic, statistical and mathematical techniques to extrapolate knowledge from big data [1]. The open data movement has also encouraged the sharing of heterogeneous search data and linking it across large digital infrastructures. As mentioned in study [2], the technology of big data allows revealing the relationships and correlations between data. It also provides predictions for decision-makers, thus making correct decisions that achieve the required goals. Big data is "a set of large and complex data with its own unique characteristics (such as size, speed, diversity, variation, data validity) that cannot be efficiently processed using current and traditional technology to benefit from it. The challenges that accompany this type of data are its provision, processing, storage, analysis, searching, sharing, transmitting and photographing, in addition to preserving the privacy that accompanies it. "[2] Big data has some characteristics according to the following [3]: size, diversity, speed, and added value. Study [2] added reliability as one of the characteristics of big data. There are many fields of uses of big data, for example, it is possible through analyzing big data [2] to identify the audience and interest in a particular organization, and to predict the results of marketing and sales campaigns.

Added to that predicting natural disasters based on previous data and comparing them with the current situation, and thus taking precautionary measures, monitoring the beneficiaries' satisfaction with the services provided, especially in government sectors. It helps also to provide a variety of opportunities and options with the aim of improving student learning through adaptive learning or competency-based education since analyzing big data helps to know the learning needs or the hassles facing the learning process. In addition, it helps develop military intelligence capabilities by collecting data from various sources and thus promoting information exchange between the military sectors. Moreover, studying the behavior of patients by analyzing files and visits, and providing an accurate description of the state of health in the medical field in particular, can diagnose diseases, treat diseases and manufacture medicines.

Based on the above, this critical paper comes to reveal the challenges of information security in big data in light of the information and communication technology revolution. Hence, the study problem emerged in answering the following question: "What are the main information security challenges in big data?" It aims to identify the challenges that faced information security in big data, because the topic is particularly important in clarifying challenges and trying to avoid and reduce them as this expected growth of data from all types of sources raises concerns about data and information protection.

This confirms the importance of conducting studies on clarifying challenges and proposing solutions and recommendations. The topic also requires the largest possible amount of explanatory studies and research, especially to find appropriate solutions to ensure the protection of information and data. This is the reason for the preparation of this scientific paper to try to search for references and seek to collect information and spread the benefit. The paper used the method of critical research evaluation based on the use of theoretical references represented by books, references, scientific periodicals in addition to studies, research, Arab and foreign scientific theses as well as seminars and official reports related to the study subject. This forms a scientific background for the study to use it as a source for collecting data and information provided that it is in the historical period between 2015-2020 AD.

After reviewing the literature, one of the most prominent results was of the paper after reviewing the literature and studies shows that the most prominent challenge of big data is the challenge of information security, especially personal data. A number of studies have suggested heading to encryption as one of the solutions and the principle of defining the purpose, restricting use, multi-layer security, and applying Block chain technology.

This goes hand in hand with the aim of the paper to define the challenges of information security in big data, thus contributing to finding appropriate solutions by programmers and specialists, and taking advantage of the characteristics and advantages of big data in saving intellectual production and making it accessible.

This paper recommended the necessity of completing specialized studies in dealing with big data challenges, especially information security, according to actual applicable systems. Leveraging technology such as Hadoop, machine learning, and Block chain to solve big data problems. The paper consists of five parts, which review a number of previous literature and studies related to the topic, clarify the problem, review the results of the survey that have been referred to in the same regard, and finally come out with the results, conclusion and recommendations, and a proposal for future studies.

RELATED WORKS

In this section, a number of related studies will be reviewed. Such works dealt with big data and its challenges, especially the information security challenge. In fact, it is noticed the lack of studies that dealt with experimental solutions to guarantee information security in big data, during the historical period between 2015 and 2020 AD. They were reviewed in chronological order from the most recent to the oldest.

Study [4] entitled "Big Data and Decision-Making at King Saud University: An Evaluation Study of Itqan System" aimed at clarifying the challenges of big data. Among the most prominent challenges that the study covered were: the size of big data that is constantly increasing, and the massive and accelerating growth in the amount of data, in addition to random search and retrieval inside big data. The study also showed that the privacy of the data subject to analysis and the issues of intellectual property of the data is the subject of much current controversy, and attention is directed towards the practices of some companies concerned with data on the Internet such as Google and Facebook, and the extent of violating privacy, which raises great concerns. · The study showed that analyzing big data helps in detecting criminal, commercial, behavioral or functional trends of the owners of this data, which makes converting it into a very useful product within the circle of knowledge-based economy a fait accompli.

The Study [5] entitled "Sharing Big Data Using Block chain Technologies in Local Governments: Some Technical, Organizational and Policy Considerations," is characterized by the fact that it deals with modern technology, Block chain that includes encryption of the private key, the network distribution with a shared record book and an incentive to maintain network transactions and safe record-keeping. These entire combine together to make block chain technology a decentralized, transparent and immutable system, difficult to manipulate and highly trackable. The study was characterized by its detailed and clear presentation of the survey methodology. Once the data is stored on the block chain, access to the information will require multiple permissions from other points in the network to access the data. Therefore, it is impossible for a cybercriminal to seize it, and by using distributed block chain technology, it is not necessary to store data centrally in order to connect the requirements of each department so that each department acts as a node in the block chain. If there is a security breach, it will be easy with block chain technologies to discover the problem and try to find solutions.

The study also clarified the most prominent challenges with data security problems, citizen privacy concerns, increased operational costs, government agencies' resistance to cooperation and information exchange due to politics and fear of losing control over their data and even their identity as an individual organization. The study recommended studying additional cases in China and other places and examining whether the challenges they identified are similar and whether the recommendations are feasible and applicable to other regional and national contexts.

Study [6] entitled "Research on information security in the big data era" analyzed the challenges and causes of data security resulting from big data, and clarified that access control is one of the main challenges, which leads to data leakage, as well as the heterogeneity of data, privacy and ownership. It was distinct in explaining the techniques used in big data. It explained some of the solutions, including the use of the smart security model, which works on a comprehensive analysis of unstructured big data, which expands the security analysis and knowledge of abnormal behavior, knowledge of risk points, as well as the application of APT-An advanced persistent threat to improve confidentiality.

Study [7] entitled "Information Security in Big Data" focused on addressing data and security issues related to big data and describing the scope of big data in the field of business. The study recommended not to focus on device security since application security is more important, maintain the isolation of devices and servers containing sensitive information, the use of feature-based encryption to protect sensitive information shared by third parties, secure open source software like Hadoop, and maintain and monitor audit records across all aspects of the business. It was characterized by its accuracy in reviewing, its presentation of recommendations and applications for big data, and its clear achievement of the goal in accordance with the title.

Study [8] entitled "Big Data in Smart Cities: Analysis and Applications in Arab World" aimed at reviewing the impact of big data on the smart city and discussed the challenges and benefits of big data and its relationship to the Internet of Things, in addition to introducing big data applications in smart cities in the Arab world. The results show that big data analytics can provide many benefits for smart cities and improve quality of life despite challenges. The study clearly presented the challenges of big data including the characteristics of big data that make it difficult to manage, the volume of data that is generated from different sources, the speed of the data, the diversity of the different data that are being created, the quality and reliability of the data, the period of retention of big data, in addition to the challenge of the integration of data from multiple sources that generate large volumes of structured and unstructured data. The data quality is one of the difficulties in any mechanism of data integration, especially if the data is incorrect, missing or incomplete, and data processing is a challenge as traditional methods of storage and processing may not be appropriate from now onwards.

The research study [9] entitled “An Imperative Analysis of Security Issues and Challenges with Big Data in SMB'S”, which aimed at analyzing big data security issues and challenges among small and medium-sized enterprises in India, in addition to providing insight into the impacts of perceived risks, requirements and benefits in Big data security for small and medium businesses. The study also mentioned few observations based on the sample responses, including the need to define protection standards, awareness of security details, the use of multi-layered security, the installation of a wide range of security mechanisms, and the use of expert advice and suggestions.

Study [10] entitled "Big Data and Data Protection: Issues with Purpose Limitation Principle", aimed at discussing the protection of personal data in the era of big data. It clarified that the biggest challenge to big data from a security point of view is the protection of user privacy because big data contains a huge amount of personally identifiable information and thus users 'privacy is a major concern and it highlighted the reasons for violating users' privacy. The study suggested the application of the protection law based on the principle of defining the purpose and the principle of restricting use, as they are the traditional pillars of data protection regulations. With regard to the protection of consumer data, it proposed the application of the so-called “notification and consent” model, which is one of the most used legitimate data processing mechanisms and therefore, any collected personal data should comply with this principle. The study came out with clarifying the principle of protecting information, especially personal data, which is the principle of defining the purpose and restricting use, despite the fact that defining the purpose has a negative impact on the effectiveness of the "notification and consent" model, as big data requires its analysis using many different algorithms that reveal unexpected correlations that can be used for new purposes. This will restrict the organization's freedom to make these discoveries and innovations.

The study [11] entitled "Big data privacy: The datafication of personal information" clarified that the most important challenges were the information resulting from formatted predictive analyzes. This requires restricting or controlling access to personal information, and one of the goals of big data analysis is classification and sorting, and such classification and sorting bypasses control and restrictions, accessing individual pieces of personal information, and aggregating personal traits and habits, in addition to analyzing and producing new information about individuals outside of their control.

Study [12] entitled "Survey of Big Data Information Security" discusses the main challenges of heterogeneous big data technology and provision. How to protect big data from unauthorized access and corruption (maintain confidentiality and integrity) and maintain availability. A review of big data features, and proposing methods and algorithms to ensure information security in big data.

Study [13] entitled "Big Data for Development: A Review of Promises and Challenges" aimed at reviewing systematically the many policies available to enhance opportunities and reduce risks.

It also reviewed what was confirmed by the White House report for the year 2014 that big data leads to "disturbing problems, as it was mentioned that it can cause societal harm beyond that harm to privacy, and came up with solutions, including the development of regulations and legislative frameworks by controlling and managing electronic databases." This is what happens on sites with online user evaluations like Facebook posts, tweets, etc. Big data providers must obtain confirmation from the customer that the data provided is not misused. For example, Instant Checkmate Company provides information about individuals taken from criminal records, phone address registries, professional and business licenses, voter registration, marriage records, demographic surveys, and census data. The consumer has to click to agree that the information will not be used to make decisions, while one of the main opportunities that big data provides is to combine data and collect data from various sources together.

Study [14] entitled "Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy?" clarified the role of BOLD and large open data in analyzing individuals' behavior, increasing control, and reducing privacy. However, it is positive in that it helps with transparency between individuals and governments. It can be used to increase control over citizens in governments. This requires striking a balance between transparency and privacy. A form that guarantees privacy was presented, as the main danger to privacy is not the information itself, but how the information is used. It also guarantees transparency by automatically issuing data related to the work of government institutions in a way that can be used for effective control. The nature and impact of BOLD on privacy and transparency can be understood, and its levels can be balanced with security, safety, openness, and other socially desirable values. The study clarified the concept of transparency and privacy in detail and suggested following the open data methodology as an incentive for big data by providing data to the public. For example, geospatial data, weather and climate data, use of technical standards for the data provided (such as impractical PDFs versus structured Excel spreadsheets versus automatically readable "linked data"), web access, and legal questions such as copyright and rights standards.

Study [15] entitled “Research on the security technology of big data information” aimed at reviewing the information security of big data and reviewed the information security issues that represent a major challenge in the era of big data wealth. Among the solutions that have been demonstrated is to apply machine learning in information security and improve traditional information security technology in a comprehensive manner through the development of new devices to protect and improve access, system control, key management, authentication, trusted channels, and other preventive measures, and strictly regulate the use of data and communication protocols for distributed systems. Information security for big data requires a multi-level security policy model, monitoring network attacks and network protocol security holes, accurate analysis of source data and sorting to combat and prevent the emergence of infiltration, virus formation and encryption, firewall, data flow filtering, intrusion detection, privacy protection technology, system security and safety assessment. The study is distinguished in that it introduced information security measures for the big data system in addition to the legislation of privacy protection measures, along with the real name system of the network, the identity of the network and the related data against privacy theft and diversion. Moreover, it is characterized by detailing the challenges according to its title, but it lacks practical models and experiments for the success of information security in big data.

RESEARCH PROBLEM STATEMENT

Big data has revolutionized knowledge production as it represents a very large amount of data that exceeds the ability of traditional computing devices to process. The past few decades have witnessed the emergence of new methods of producing, storing and analyzing data to process this huge amount of data, culminating in the emergence of the field of data science, which combines computational, algorithmic, statistical and mathematical techniques to extrapolate knowledge from big data [1]. The open data movement has also encouraged the sharing of heterogeneous search data and linking it across large digital infrastructures.

As mentioned in [2], the technology of big data allows to reveal the relationships and correlations between data. It also provides predictions for decision-makers, thus making correct decisions that achieve the required goals. It was imperative to identify the challenges that faced that service, especially information and data security, and conduct studies to try to address those challenges and find solutions.

On this basis, this paper comes to reveal the challenges of information security in Big Data. Thus, the study issue is answering the following question:

“What the main challenges of information security are in Big Data?” and the following sub-questions are derived:

- What is the main challenge that the studies have agreed upon?
- What are the main solutions?
- What are the deficiencies that were left after the survey study according to the searcher’s point of view?

THE APPLIED RESEARCH METHODOLOGY

The method of critical review was used, where the analysis will be done through the use of theoretical references represented by Arab and foreign books, references and scientific periodicals in addition to Arab and foreign studies, research and theses, as well as seminars and related official reports. This work will cover the historical period between 2015-2020 AD. As a result, it is noticed that the studies agreed, regardless of their different references, that the biggest challenge and the most important threat to Big Data is the protection of information and data. It also noticed that:

Study [4] is considered, from the researcher's point of view, one of the most comprehensive studies that collected the most prominent strengths in terms of the clarity of the study’s review and its compatibility with the title, in terms of presenting the methodology, problem and terminology, clarifying challenges, and coming up with a system. The most prominent challenges revealed were the explosive and rapid growth in the amount of data, random search and retrieval within the big data, the diversity of data, the availability of personnel specialized in analyzing big data, and the availability of expert automated systems that fit the needs of the organization and have good capabilities and flexibility in use and development. The study also showed that the privacy of the data subject to analysis and the issues of intellectual property of the data are the subject of much current controversy. Thus, attention is directed towards the practices of some companies concerned with data on the Internet such as Google and Facebook, and the extent to which they violate privacy. This raises great concerns. The study showed also that analyzing big data helps in revealing the criminal, commercial, behavioral or functional trends of the owners of this data, which makes its transformation into a very useful product within the circle of knowledge-based economy a fait accompli.

Study [5] was distinguished as it stressed the modern technology of Block chain, which is one of the latest technologies to help maintain data security. It includes a set of private key encryption elements, a distribution network with a shared record book and an incentive to maintain network transactions and safe record keeping. These elements come together to make the block chain technology a decentralized, transparent and immutable system. They are difficult to manipulate and highly trackable. The study was characterized by its detailed and clear presentation of the survey methodology. It also described this technology by stressing that once the data is stored on block chain, access to the information will require multiple permissions from other points in the network to access the data. So, it is impossible for a cybercriminal to seize it, and by using distributed block chain technology, data is not required to be stored centrally in order to connect the requirements of each department so that each department acts as a node in the block chain. If there is a security breach, it will be easy with block chain technologies to discover the problem and try to solve it. The study also clarified the most prominent challenges, including data security problems, citizen privacy concerns, increased operational costs, government agencies' resistance to cooperation and information exchange due to politics and fear of losing control of their data and even their identity as an individual organization.

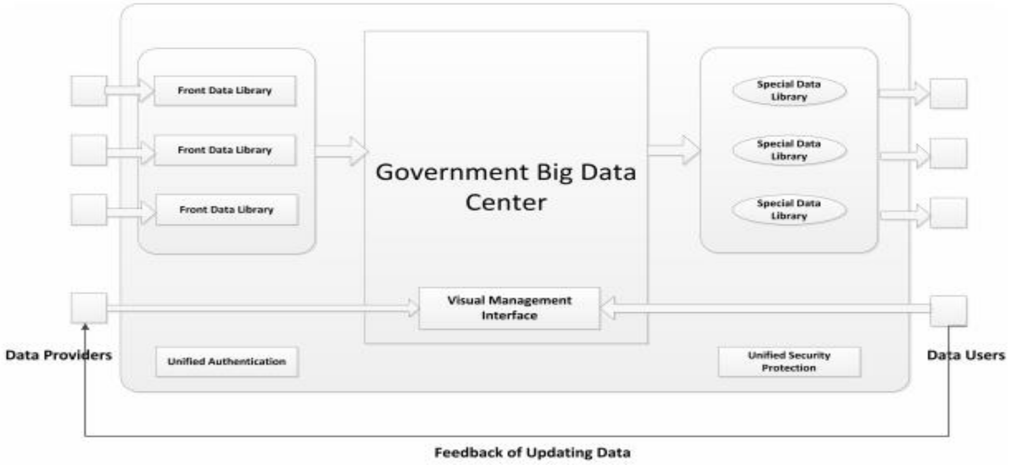

Figure 1. *Centralized sharing mechanism of government big data ([5], p. 426)*

Study [6] and study [7] are similar in analyzing the challenges and causes of data security resulting from big data, in discussing the development trend of network attacks under the background of big data, and in clarifying that access control is one of the main challenges, which leads to data leakage, as well as inconsistency of data, privacy, and ownership. It is also characterized by explaining the technologies used in big data. It explained some of the solutions, including the use of the smart security model, which works on a comprehensive analysis of unstructured big data expanding the security analysis and knowledge of abnormal behavior, knowledge of risk points, as well as the application of APT-An advanced persistent threat to improve confidentiality. The study reviewed the challenges and appropriate solutions according to technological development, but it lacks models and experiences. Study [7] also focused on addressing data and security issues of big data and describing the scope of big data in business. Securing big data platforms requires a combination of traditional methods of security tools, newly developed toolkits, and intelligent processes to monitor security throughout the life of the platform. It also clarified the need for big data environments to add another level of security because security tools must work through three data phases, namely accessing and entering data, the second phase storing data, the third phase is data results. The study clarified also a model for data flow and security.

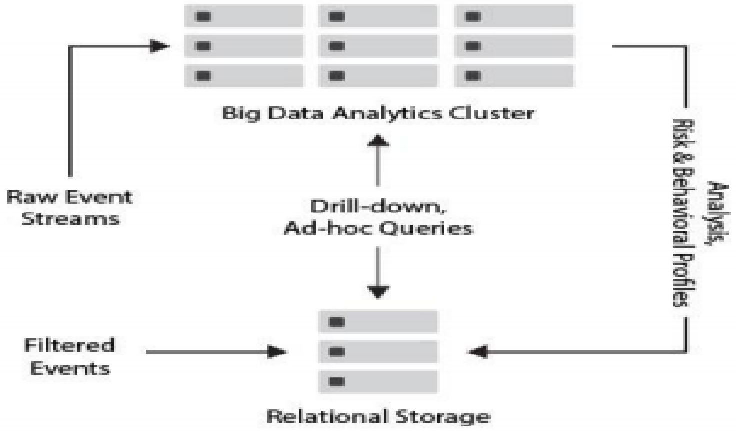
Big Data Analytics Cluster
Raw Event Streams
Drill-down, Ad-hoc Queries
Analysis, Risk & Behavioral Profiles
Filtered Events
Relational Storage

Figure 2. *Data Flow and Security in Big Data ([7], p. 25)*

What distinguished the previous study [7] is its reviewing and identifying of the challenges and solutions, which were clarified according to the following:

- Emphasizing access controls, and it is important to provide a system in which encryption, authentication, and validation take place.
- Data is stored at multiple levels, depending on business needs, and conscious strategy creation.
- precision in determining when the past attacks occured, what the consequences were, and what needs to be done to improve matters in the future.
- Most big data applications distribute huge processing functions across many systems like Hadoop which is a well-known example of open source technology.
- Encryption: Encryption also requires to work on many different types of data, both user and device generated.
- Central Key Management: A security best practice for many years. It only applies with the same strength in big data environments, especially those with a wide geographical distribution.

- User access control: User access control may be the primary tool for network security, but many companies exercise that minimal control because administrative expenses are so high.
- Intrusion detection and prevention: Intrusion detection and prevention systems are security work tools.

Among the provided recommendations were not to focus on device security since application security is more important, maintain isolation of devices and servers that contain sensitive information, introduce interactive and proactive protection, in addition to feature-based encryption to protect sensitive information shared by third parties, secure open source software such as Hadoop, and the maintenance and monitoring of audit records across all aspects of the business. What characterizes the study is its accuracy in reviewing and providing recommendations and applications for big data, and in achieving the goal clearly.

Study [8] entitled "Big Data in Smart Cities: Analysis and Applications in Arab World" was characterized by clarifying the proposed survey methodology, which relies on collecting information on smart cities, big data and the Internet of things. Then the information was used to clarify the challenges and benefits of applying big data in smart cities, and the relationship between big data and the Internet of things, in addition to the software solutions used to develop the smart city and its big data capabilities. A case study was conducted in some Arab cities and the big data applications implemented in them. The study is very distinct in terms of presentation, content, clarification of methodology and division of study parts. The challenges of big data were clearly presented, including: the characteristics of big data that make it difficult to manage the size of data that is generated from different sources, the speed of the data, the diversity of the different data that are being generated, the quality and reliability of the data, the period of retention of big data. Moreover, considering smart city applications, big data difficulties arise in collecting data on its own, which is complicated due to multiple sources with different formats, types and uses in addition to access policies. Added to that, the unstructured nature of the data makes it difficult to categorize and organize it. Security and privacy is another important challenge to using big data in the smart city profile, where smart city entities have a large size of data that includes personal and private information about individuals such as medical data and financial records. Among some of the proposed solutions are legal provisions on the use of data, as the built-in smart applications need a strong security type because they will be transmitted over different types of networks, some of which may be insecure. In addition to the challenge of integrating data from multiple sources that generate a large quantity of structured and unstructured data. Data quality is also one of the difficulties in any data integration mechanism, especially if the data is incorrect, missing or incomplete. Data processing is also a challenge as traditional methods of storage and handling may be no longer appropriate.

As for the research study [9] entitled “An Imperative Aanalysis of Security Issues and Challenges with Big Data in SMB’S”, it was similar to most studies on challenges, except that it followed the exploratory approach for the purpose of knowing the information and formulating hypotheses and analytical method in order to analyze the discovered information.

The stratified random sample method was used on a sample of 300 employees, which is a good sample to evaluate the results of the research. A number of companies were selected, namely "Dr B Lal Clinical Labrotory Private Limited, Jaipur", "Gravita India Limited, Jaipur", and "Elektrolites (Power) Pvt. Ltd., Jaipur". The primary data were collected from the results of a questionnaire. Secondary data from reports, books, magazines, researches, and other IT companies, SMEs and big data. Statistical analysis tools were used for the experimental study according to the Cronbach Alpha Tool for Reliability, and Kaiser's Rank Variability Test for variables. The statistical analysis of hypotheses was used according to the Chi Square Test, One Way ANOVA with the help of the SPSS program. The study is very distinct from the researcher's point of view as it clarifies the objectives, assumptions and the approach followed in a way that makes it easy for any reader, whether specialized in the field or not. The study also used graphs and tables for more clarifications. However, it did not mention the country of study in the sample in the title. The researcher also found it difficult to identify the term SMB “s” used, and it was concluded that it is an acronym for medium and small business technology companies. The study concluded that dealing with and obtaining big data, which is mostly unorganized and semi-structured data, requires defining a set of protection that meets the basic requirements for the security of small and medium enterprises and ensuring information security with the use of Symantec Protection Suite (SPS) which demonstrated its ability to secure business and computer data, protect them against risks of new malware and spam emails, and quickly recover computers and data in case of problems. This study identified the impact of data security on SMB and recommended the use of encryption, access control, and authentication as solutions for big data security in India. It also clarified the importance of the results for future studies of information security analyzes in big data. The study also mentioned few observations based on the sample responses, including the need to define protection standards, awareness of security details, the use of multi-layered security, the installation of a wide range of security mechanisms, and the use of expert advice and suggestions on security issues. In general, the study is very distinct and it may need to review applicable experience of information security in big data.

Study [10] is similar to other studies in that the biggest challenge for big data from a security side is the protection of user privacy because big data often contains a huge amount of personally identifiable information and therefore users 'privacy is a major concern, and also highlighted that Users' privacy may be violated under the following reasons:

- Personal information, when combined with external data sets, may lead to inferring new facts about users. These facts may be confidential and are not supposed to be disclosed to others.
- Personal information is sometimes collected and used to add value to it. For example, individual shopping habits may reveal a lot of things.
- Sensitive data is stored and processed in a properly insecure location and data leakage may occur during the storage and processing phases.

With data protection, personal data will be able to be processed if allowed by law or with the consent of the person, however, big data applications are difficult to check. The study proposed the application of the protection law which works on the principle of defining the purpose and the principle of restricting use as they are the traditional pillars of data protection regulations. Indeed, with regard to the protection of consumer data, the application of the so-called "notification and consent" model, which represents one of the most used legitimate data processing mechanisms, any personal data collected complies with this principle.

The study came out with clarifying the principle of protecting information, especially personal data, which is the principle of defining the purpose and restricting use, despite the fact that defining the purpose has a negative impact on the effectiveness of the "notification and consent" model, as big data enables data to be analyzed using many different algorithms that reveal unexpected correlations that can be used for new purposes. This will restrict the organization's freedom to make these discoveries and innovations. The study revealed two problems: The first is when data is used for new purposes. In fact, in accordance with the application of this principle, it is imperative that organizations using collected personal data as a file ensure the basis for predictive analysis and that the analysis is consistent with the original purpose of data collection. Big data analytics also repurpose data obtained for a different purpose and in some cases by another organization. This poses a challenge to the privacy principle that the collected data may not be used for the purposes collected and thus inconsistent with the original purpose of the collection. The second problem is that the use of big data helps to identify general trends, understand individuals or make decisions about them using personal data about them. The challenge is to provide an approach to obtaining approval as the analysis helps to identify a person's lifestyle in determining a credit rating. The purpose can be repurposed in ways that are most beneficial to society. The study is strong and important in content as it focused on two principles and two most important problems in big data, but it did not address an applied example or an applied experiment and followed a descriptive approach. The paper needs to clarify the methodology used according to the researcher's point of view.

Study [11] was distinguished by its focus on the privacy of personal data in big data. It is imperative to adopt a policy to ensure the consent of individuals when collecting data on the potential behavior of individuals. Analyzing the data that has already been collected can violate privacy, but this new information does not require consent. It also explained what the most prominent challenges were: the information generated by formatted predictive analyzes, which requires restricting or controlling access to personal information. In addition, one of the goals of big data analysis is classification and sorting, and such classification and sorting exceed control and restrictions, allows access to individual parts of personal information and the compilation of personal traits and habits, in addition to analyzing and producing new information about individuals beyond their control.

Study [12] was characterized by discussing the main challenges of heterogeneous big data technology and its provision; How to protect big data from unauthorized access and corruption (maintain confidentiality and integrity) and maintain availability; Issues related to providing big data features; And suggesting methods and algorithms to ensure the characteristics of information security. Among the methods discussed are privacy attacks and big data analysis. The concepts of designing a security algorithm for big data have also been proposed. It stressed the field of integrity and privacy in analyzing big data, and investigating the characteristics of big data searching. It reviewed algorithms to ensure that massive data is processed using encryption or data transformation algorithms which are divided into several types:

1. Processing the information received in response to external requests for statistical databases.
2. Processing the additional and multiplier noise generated by the probability distribution of the data values.
3. Data anonymity.

It was also characterized by a detailed and clear presented review of the survey methodology, which is consistent with the title.

Study [13] used a conceptual framework to review empirical evidence and referred to about 180 articles related to the opportunities and threats of big data. It shows that the emergence of big data has the potential for cost-effective and improved decision-making in important development areas such as healthcare, economic productivity and security. At the same time, there are challenges, the most important of which are the scarcity of human resources. The study systematically reviewed many of the available policies to enhance opportunities and reduce risks. It also reviewed the assertion of the White House's 2014 report that big data leads to "disturbing problems, as it has stated that it can cause societal harm beyond its harm to privacy, such as discrimination against individuals and groups." At the same time, the tremendous opportunities it provides are emphasized. These technologies are in terms of improving public services, developing the economy, and improving the health and safety of society.

The study showed that the big data model is currently undergoing an uneven proliferation process in terms of lack of infrastructure, human capital, availability of resources and institutional frameworks in developing countries. This creates a gap in the ability to execute data analytic processing and concerns about state and corporate control and manipulation and blind confidence in imperfect algorithms. It also came out with solutions including the development of regulations and legislative frameworks by controlling and managing electronic databases. This is what happens on sites when online user evaluations such as in Facebook posts, tweets, etc. The big data providers must obtain assurance from the customer that the data provided is not misused. For example, Instant Checkmate Company provides information on individuals taken from criminal records, telephone and address records, professional and business licenses, voter registration, marriage records, demographic surveys, census data, and the consumer must click to agree that the information will not be used to make decisions. While one of the main opportunities offered by big data is to integrate data and collect data from different sources together.

Study [14] entitled "Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy?" clarified the role of BOLD and large open data in analyzing individuals' behavior, increasing control, and reducing privacy. However, it is positive in that it helps with transparency between individuals and governments. It can be used to increase control over citizens in governments. This requires striking a balance between transparency and privacy.

It presented a model that guarantees privacy, as the main danger to privacy is not the information itself, but how the information is used. It also guarantees transparency by automatically issuing data related to the work of government institutions in a way that can be used for effective control. The nature and impact of BOLD on privacy and transparency can be understood, and its levels can be balanced with security, safety, openness, and other socially desirable values.

The study clarified the concept of transparency and privacy in detail. It is strong and important in content, as it is one of the rare studies that linked the terms transparency and privacy and the principle of balance between them as a prominent challenge in big data. However, it did not address an applied example or an empirical experience and followed a descriptive approach from the researcher's point of view. The study is very distinct from the researcher's point of view as it clarifies the objectives and approach followed in a way that makes it easy for any reader, whether specialized or in the field or not. The study also sought clarification through graphs and tables. It also explored examples of uses of big data. It also suggested following the open data methodology as an incentive for big data by providing data to the public. This approach treats data as a public good such as geospatial data, weather and climate data. In addition, it dealt with using technical standards for the data provided (such as impractical PDFs versus structured Excel spreadsheets versus automatically readable "linked data), web access, and legal questions such as copyright and rights standards.

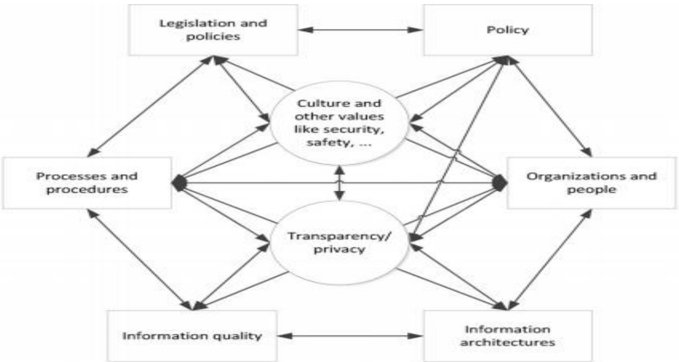

Figure 3. *Elements and Dependencies Comprising Transparency and Privacy Landscape ([14], p. 367)*

Study [15] is a survey study that reviewed the information security of big data and reviewed what information security issues represent as a major challenge in the era of the wealth of big data. Among the solutions that have been clarified is to apply machine learning in information security and improve traditional information security technology in a comprehensive manner through the development of new devices to protect and improve access, system control, key management, authentication, trusted channels, and other preventive measures, and strictly regulate the use of data and communication protocols for distributed systems. Information security for big data requires a multi-level security policy model, monitoring network attacks and network protocol security vulnerabilities, accurate analysis of source data and sorting to combat and prevent the emergence of infiltration, virus formation and encryption, firewall, data flow filtering, intrusion detection, privacy protection technology, system security and safety assessment. The study is distinct in that it introduced information security measures for the big data system in addition to the legislation of privacy protection measures, along with the real name system of the network, network identity and related data against theft and diversion of privacy. The study is also characterized by detailing the challenges according to its title, but it lacks practical models and experiments for the success of information security in big data.

RESULTS

In this section, we try to answer the main problem question and the sub-questions, and followed the survey approach resulted in the following points:

- (12) Studies were referred to according to the historical period between 2015-2020, according to the following drawing: -

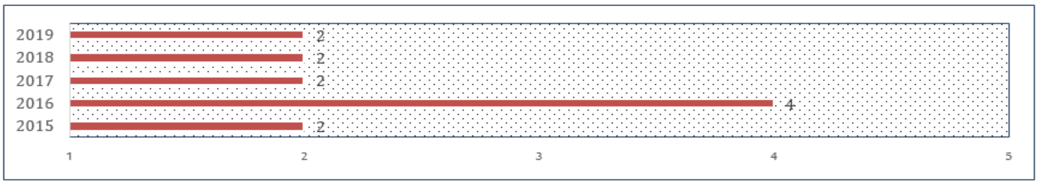

Figure 4. *The referenced studies against years*

- With regard to the question of the study issue about the most prominent challenges, it was clear that the studies agreed on the following challenges classified according to their frequency in the studies

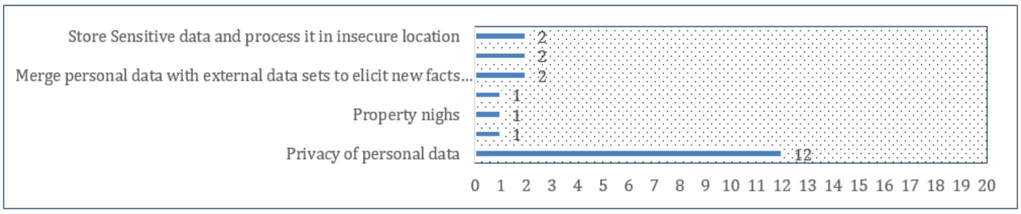

Figure 5. *Main Challenges Facing Security of Big Data*

It is noticed from Figure 5 that:

- Most studies agreed that information security is the most important and biggest challenge.
- The technical challenges are not mentioned in a large scale, and this may be due to the technical revolution in this era.
- With regard to the question of the study problem about deficiencies, it is noticed, as far as the researcher is aware of, the lack of studies that dealt with practical experiments to ensure the security of information in big data, and the lack of studies that came up with an actual application system. Most of which provided recommendations and proposals to implement mechanisms that contribute to solving challenges and threats in big data especially information security.

- With regard to the question of the study problem about the most prominent solutions, there were studies that came up with solutions according to the following figure:

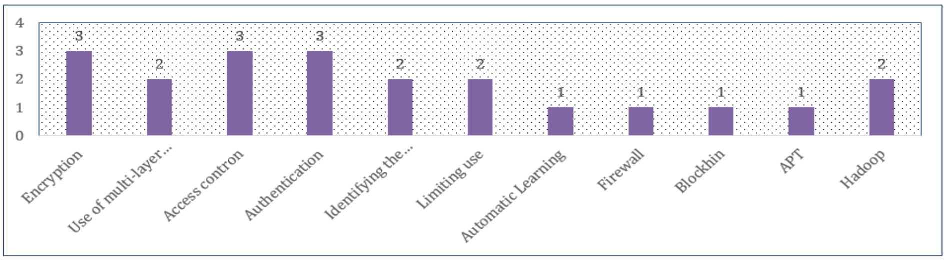

Figure 6. *Various suggested Solutions of Inf. Security*

- According to Figure 6: -
- A number of studies have agreed that encryption is one of the solutions that may prove effective in ensuring the security of information in big data, as well as access control and authentication.
- A recent study confirms the effectiveness of Black chain and big data technologies like Hadoop.
- One of the possible solutions that could be found as explained [2] is the use of tools and techniques for analyzing big data such as Hadoop, HPCC, Map Reduce. However, the most famous one is Hadoop as it is an open source program written in Java for storing and processing big data. Amazon, Apple, AVG, eBay, Electronic Arts, Facebook, Google and Yahoo are among the most famous examples of companies that use Hadoop, which sends commands to all servers at the same moment and every server provides its data, and then this data is collected and returned as one package. [3] also added the SAP HANA system, which performs the process of simultaneous analysis of big data, helps in decision-making and performs planning and implementation processes with high efficiency. [2] Clarified that the tools for dealing with big data are of three main parts, represented in Data mining tools, Analysis tools, and results display / visualization tools.
- Among the possible solutions to avoid the problem of violating data privacy:

- **Block chain technology**, which is one of the latest technologies to help maintain data security and includes a set of private key encryption elements, a distribution network with a shared logbook and an incentive to maintain network transactions and record-keeping in safety.
- **Apply-APT- An advanced persistent threat -** to improve confidentiality.

- Focusing on device security as application security is more important, maintaining isolation of devices and servers containing sensitive information, introducing interactive and proactive protection, in addition to feature-based encryption to protect sensitive information shared by third parties, securing open source software such as Hadoop, and saving Records and monitors audits across all aspects of the business. The data is stored at multiple levels, depending on business needs and the creation of a well-formed strategy.

- Carefully determine when the past attacks occured, what the consequences were, and what needs to be done to improve matters in the future.
- Most big data applications Distribute massive processing functions through many systems like Hadoop which is a well-known example of open source technology.
- Encryption: Encryption also requires to work on many different types of data, both user and device generated.
- **Central Key Management:** One of the best security practices for many years. It only applies with the same strength in big data environments, especially those with a wide geographical distribution.
- **User access control**: User access control may be the primary tool for network security, but many companies exercise that minimal control because administrative expenses are so high.
- **Intrusion detection and prevention**: Intrusion detection and prevention systems are security work tools.
- **Application of the protection law:** the protection law works with the principle of defining the purpose and the principle of restricting use as they are the traditional pillars of data protection regulations. With regard to the protection of consumer data, the application of the so-called "notification and consent" model, which is one of the most used legitimate data processing mechanisms, necessitates that any collected personal data must comply with this principle.
- The study is distinct in describing the challenges and agreed with all studies on shaping the privacy of information as a major and prominent challenge in big data.
- One of the solutions to ensure information security is the algorithms to ensure that big data is processed using encryption or data conversion algorithms, which have been divided into several types:
  - Processing information received in response to external requests for statistical databases.
  - Address the additional and multiplier noise generated by the probability distribution of the data values.
  - Anonymization of data.

CONCLUDED COMMENTS & RECOMMENDATIONS

The paper aimed at identifying the challenges of information security in big data as this contributes to finding appropriate solutions by programmers and specialists, and taking advantage of the characteristics and advantages of big data in saving intellectual production and making it accessible.

The following conclusions were made:

- The great similarity between researchers in the reviewed studies on the most prominent challenges.
- Information security is the main challenge in big data.
- Personal information, when combined with external data sets, may lead to inferring new facts about users. These facts may be confidential and are not supposed to be disclosed to others.
- Personal information is sometimes collected and used to add value to it. For example, individual shopping habits may reveal a lot about users.
- Sensitive data is stored and processed in an insecure location and data leakage may occur during the storage and processing phases.
- Big data may be an ideal solution for enterprises and organizations, but on the other hand, there are many concerns related to technology services and it may contain defects or security flaws and complex software. Therefore, it is imperative to apply security methods to preserve information on a large scale that depends on techniques to detect defects, understand their consequences, isolate their effects, and find appropriate solutions.
- However, it is worth noting that: Most of the studies lack the practical part that can be introduced and implemented.

The paper* recommends *the following:

- The necessity of completing studies specialized in dealing with big data challenges, especially information security, according to actual regulations that can be applied.
- Leveraging technology such as Hadoop, Machine Learning, and Block chain.

REFERENCES

1) Leonelli, Sabina (2020). Scientific Research and Big Data, Available at: https://plato.stanford.edu/entries/science-big-data/, Received in 14/8/2020.

2) Al-Bar, Adnan (2017). Big data and its areas of application, available at: https://www.kau.edu.sa/GetFile.aspx?Id=285260&fn=Article-of-the-Week-Adnan-Albar-01-November-2017.pdf. Accessed 10/20/2020.

3) Al-Aklabi, Ali (2017). Converting Big Data into Added Value, available at: https://kfnl.gov.sa/Ar/mediacenter/EMagazine/DocLib/23_2/82_102.pdf, accessed 10/20/2020.

4) Al-Aklabi, Ali (2019). “Big Data and Decision-Making at King Saud University: An Evaluation Study of the Itqan System”, HBKU Publishing, available at: https://tinyurl.com/y5axgg6j, retrieved on 11/2/2020.

5) Fana, Lingjun; Gil-Garciab, J. Ramon, C; Songd, Yi; Cronembergerb, Felippe; Huae, Gang; Werthmullerb, Derek; Burkeb, G. Brian; Costellob, Jim; R. Meyersb, Benjamin; Honga, Xuehai (2019)". Sharing big data using block chain technologies in local governments: Some technical, organizational and policy considerations", Information Polity 24 (2019) 419–435, DOI 10.3233/IP-190156.
6) Zhou, Linqi; Gu, Weihong; Huang, Cheng; Huang, Aijun; Bai, Yongbin (2018)." Research on information security in big data era", AIP Conference Proceedings 1967, 020020, https://doi.org/10.1063/1.5038992
7) Shaik, Naseema; Shaik, Mubeena; Mohammad, Nada Ahmed; Alomari, Fatima Ahmed (2018). "Information Security in Big Data", International Journal of Engineering Research and General Science Volume 6, Issue 5, September-October, 2018 ISSN 2091-2730.
8) Abdel Hafez, Hoda A. (2017). "Big Data in Smart Cites: Analysis and Applications in Arab World", Egyptian Computer Science Journal (ISSN-1110-2586), Volume 41– Issue 1, January 2017.
9) Singh, Reena; Goyal, Dinesh (2017). "An Imperative Aanalysis of Security Issues and Challenges with Big Data in SMB'S", Suresh Gyan Vihar University, Jaipur (Rajasthan), Volume 8, No. 9, November-December 2017, International Journal of Advanced Research in Computer Science, DOI: http://dx.doi.org/10.26483/ijarcs.v8i9.4917
10) Abdul Ghani, Norjihan; Hamid, Suraya; Izura Udzir, Nur (2016). "Big Data and Data Protection: Issues with Purpose Limitation Principle", Int. J. Advance Soft Compu. Appl, Vol. 8, No. 3, December 2016 ISSN 2074-8523.
11) Mai, Jens-Erik (2016). "Big data privacy: The datafication of personal information", THE INFORMATION SOCIETY, 2016, VOL. 32, NO. 3, 192-199, https://www.tandfonline.com/doi/full/10.1080/01972243.2016.115301
12) [12] Miloslavskaya, Natalia; Makhmudova, Aida (2016). "Survey of Big Data Information Security", 2016 4th International Conference on Future Internet of Things and Cloud Workshops, National Research Nuclear University.
13) Hilbert, Martin (2016). "Big Data for Development: A Review of Promises and Challenges", Development Policy Review, 2016, 34 (1): 135-174.
14) Janssen, Marijn (2015). "Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy?", Government Information Quarterly Journal, 32 (2015) 363-368.
15) Zhu, Hong; Xu, Zheng; Huang, Yingzhen (2015). "Research on the security technology of big data information", International Conference on Information Technology and Management Innovation (ICITMI 2015).