

CHALLENGES AND APPROACHES OF SECURITY IN BIG DATA

Aala Ibrahim Al-Handhali^{#1}, Halya Jaber Al-Salmami^{#2}, Dr. M.Senthilkumar^{#3},

*Department of Information Technology,
University of Technology and Applied Sciences, Ibra,
Sultanate of Oman.*

¹36J1643@ict.edu.om

²36J1656@ict.edu.om

³senthilm@ict.edu.om

Abstract - Large data has become a serious subject of research in all areas of government, academia and institutions as a result of the rapid development of information technology. Large data brings opportunities that are not available through small data in many areas, such as business, education and health care. On the other hand, massive data development still faces many security and privacy issues over the life of large data due to its huge quantity. Security and privacy issues may affect users and businesses. It also impairs the expected progress of data and huge opportunities. In this search paper, we first reviewed BIG DATA and its characteristics. The new security challenge posed by the big data, the first five security risks, and as we discussed the privacy and security challenges of big data problems and finally, we discussed the discovery technology of threats based on big data and Approaches for security in big data. Security analysis techniques and their characteristics using big data. Thus, the purpose of this paper is to clarify the challenges of big data security and privacy.

Keywords: Big data, security analysis

I. INTRODUCTION

Data is becoming one of the most significant assets of companies in all fields. It is important for many things not only for companies associated with computer science, but for institutions also for example state government, health care, education, or the engineering sector. Data are necessary in terms of carrying out their daily activities, as well as helping companies manage their goals and making the best results on the base of information mined from them. We live in the era of large data. These data are often unorganized, which indicates that traditional systems are not able to analyze them. Organizations are willing to extract more useful information from this large volume and a variety of data. Therefore, a new analysis model emerged to analyze and better understand this data, not only to obtain special advantages, but also general and these were big data. All new technology brings problems. In the case of big data, these problems relate to the size, diversity, quality, privacy and security of the data. We need more regulations to address the concerns of data storage and analysis. Big Data will not achieve the required level of confidence without sufficient security. Because large data brings great responsibility [1]

II. RELATED SURVEY

There are many different definitions of "big data" that have caused noise over the last few years. To develop consensus definition, classifications, safe reference structures, and technology roadmap NIST has formed a large data working group as a community with industry, academia and government members. The characteristics of large data are defined as various comprehensive data sets, such as structured, semi-structured and unorganized data from different areas [6].

The solution that is expected in large enterprises is data security and privacy. It can create the maximum amount of data centre-centric protection through which large data exchange between the data centre and users is minimized and key searches that may be part of security problems Home Data Centre. Some protocols are easy to design [7].

III. CHARACTERISTICS OF BIG DATA

Big data has certain properties that we can separate from normal data. There are six main characteristics that determine large data. These characteristics are also known as the six large data [5].

Size: Indicates the huge amount of data that is generated every second. Many factors contribute to increasing volume, such as storing transaction data, live broadcast data, and data collected from sensors and human interaction on systems. The amount of data that is generated is not in terabytes but in bytes or byte bytes [5].

Diversity: Indicates the type of data stored. Today's data has different types of formats

from multiple sources. Data is generated in many formats, such as text, web data, sensors, Twitter, audio, video, etc. There are three different types of data, structured (relational) data, semi structured data (XML data) and unstructured data (text and multimedia content). 80% of the world's data is unstructured (texts, images, video, audio, etc.) [5].

Speed: This means the speed of data production and processing. It also indicates the speed at which data are moving and where new data is generated. Such as social media messages. A new technique allows us to analyze data as it is created (memory analyzes), without putting them into databases [5].

Veracity: It is the level of reliability associated with different types of data. Searching for data quality has become an important demand and challenge for large data, and the best data cleansing methods can not eliminate the unpredictability of data such as weather, economics, or customer purchasing decisions [5].

Value: All data is important and has value. Good information may be hidden in unorthodox, unorganized data. The challenge is to determine the value and then convert and extract the data for analysis [5].

Variable: is chaos or data merit. With large-scale data quality and control accuracy, where technology allows us to work with this type of data. With increasing speeds and types of data, as data flows may be very inconsistent [5].

IV. BIG DATA FRAMEWORK

The large data architecture is distributed and can reach thousands of data and contract processing as the data is divided, repeated and distributed among thousands of nodes. Data is divided into two categories:

Hot data and cold data. The automobile is called through disks rather than data as it transforms a large model from traditional to modern architecture. Supports modern architecture (real-time analytics) and data collection from different combinations of input sources. And transfer them to large data solutions on a regular basis. Custom queries are also provided along with powerful parallel programming and robust layout. Many frames such as map reduce where the program is divided into several maps are executed on the relevant data points and then are reduced to one set, Storm Topology (Spouts & Bolts), where Spouts are sources of data and Bolts are data processing The following nodes are used to calculate real time [11].

V. NEW SECURITY CHALLENGES

Unusable data in some organizations is now of high value, subject to privacy laws and protection has become important. Many mobile operators are collecting data from cell towers. Oil and gas companies collect data from seismic sensors. As for power plants, they collect data from power plants and distribution systems. Companies also collect large amounts of user-generated data from potential customers, such as credit card and social security numbers, purchasing habits and usage patterns. All these things have led to the large data flow and the need to transfer data across the enterprise in order to create

a huge new target for hackers and Internet criminals Others [8].

big data presented challenges in terms of data security. There is a growing need for research in technologies capable of handling large volumes of data and making them efficiently secure. When these technologies are applied to secure data on huge amounts of data, they may be slow [9].

Challenges in the large data system are divided into four aspects:

- Security Infrastructure
- Integrity and interactive security
- Data privacy
- Data management [10]

VI. SECURITY RISKS

Authors must Security mechanisms are generally weak in large data technology. Where strong security mechanisms were found for the purpose of using features such as automotive, parallelism and others [11]. Problems such as invasion of privacy, complexity of drive storage, gaseous marketing, etc. were difficult problems. They led to challenges in implementing Big Data Analytics tools for large data solutions and applications. Such as:

1. Account is not safe
2. Validate input and filter
3. Granular Access Controls
4. Secure storage data
5. Privacy concerns in data extraction and analysis [11].

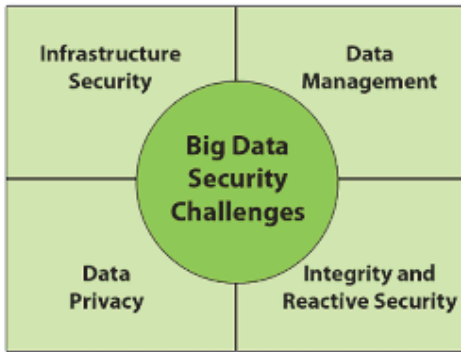


Figure 1 – Big data Challenges

VII. THE PRIVACY OF BIG DATA SECURITY ISSUES AND CHALLENGES

big data security is the process of protecting and analyzing data in places where it may be threatened. As the security of large data has become a constant concern, because of potential hackers who aim to disseminate big data. When analyzing big data often, personal information of people must be combined with large sets of external data. If This information is from the database or social networking sites. To become aware of any person may be confidential. It also leads to insight into the lives of people who do not realize it. A learner often benefits more than a less educated person. From learning better concepts about analyzing big data from predictive analysis [3].

1) Problems associated with the SQL injection type cannot be lost. Where they are transported with Hadoop components such as Hive and Impala. At the moment SQL functionality is not available, but you may be able to enable query and data separation [2].

2) Sensitive data Do not find any native encryption controls to protect it frequently, this security is provided only outside the data stack or application [2].

3) When connecting between Data Node and Data Node, clear text data is sent. Data locations cannot be strictly enforced. Also, the scheduler will not be able to find resources next to the data to force it to read over the network [2].

VIII. APPROACHES FOR SECURITY IN BIG DATA

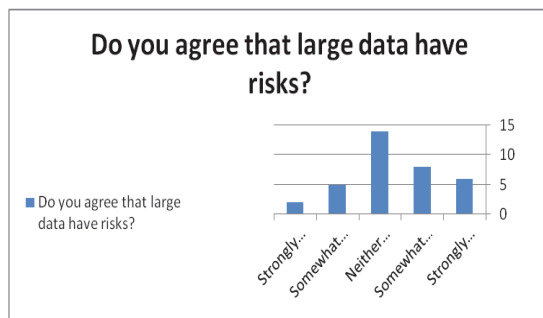
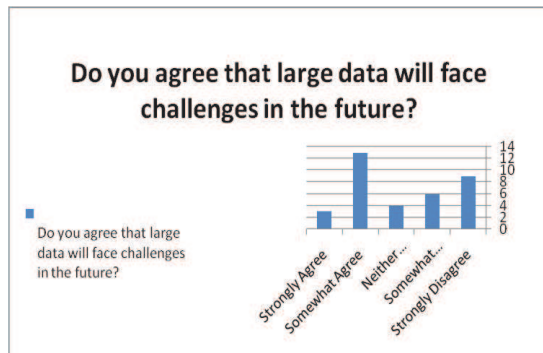
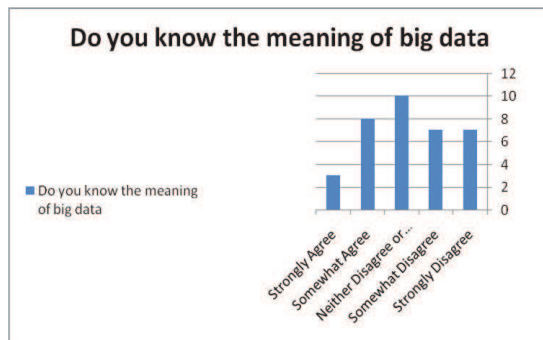
Massive data technologies were not designed with security mode. Safety Data security means unauthorized change, destruction, or exposure (intentional, unintentional or malicious). It can protect a database of critical forces and unwanted actions from unauthorized users. Security must be the first priority of large data in enterprises. To protect data, you must understand its risks, know common attacks and maintain security. Organizations that maintain the data necessary for their success. May be important to other organizations or individuals. Be sure they are safe from unauthorized access. It requires more accurate efforts to steal data and destroy the reputation of organizations. These problems are constantly evolving and we cannot ignore them [5].

New confirmation on encryption. Security as a service, Real-time data collection, Privacy by design, Data protection, Manage Log. Authentication, Hide data, Use a secure connection [5].

IX. RESULTS AND DISCUSSION

In this research, we suggested that some questionnaires be conducted for university students from different disciplines to know their knowledge of the big data, its importance, the importance of its security, the challenges it face, and the risks involved. We then distributed 35 question sheets

containing only 4 basic questions. Finally, we calculated the percentage of student information and reached these results.



X. CONCLUSION

In this study, we first reviewed BIG DATA and its characteristics. The new security challenge posed by the big data and the five security risks first, as we discussed the privacy and security challenges of big data problems. Finally, we discussed the technology of detecting threats based on large data and Approaches for security in big data. Security analysis techniques and their characteristics using large data. There is no doubt that the current large data technologies and tools to solve the security problems of large, practical data and privacy are very limited. Our future research will focus on security and privacy reviews broken down by application domains, such as health care, Internet (IOT) and social media. Through the proper analysis of both big and fixed data sets, we can make better progress in many scientific, medical and profitability disciplines for many companies. Where the application cannot be imagined without the consumption of data and the creation of new forms of data, we also highlighted in this paper the large data and its characteristics. As well as the challenges they represent and the most important security and privacy issues that need to be addressed if we are to make the infrastructure for large data processing safer. We believe that this paper will stimulate research and development work in the community to focus collaboratively on the video that transforms increased security and privacy into large data platforms.

REFERENCES

[1] Julio Moreno *, Manuel A. Serrano and Eduardo Fernández-Molina Volume,"Main Issues in Big Data Security".

[2] Ajit Gaddam," Securing Your Big Data Environment"

[3] Getaneh Berie Tarekegn, Yirga Yayeh Munaye," BIG DATA: SECURITY ISSUES, CHALLENGES AND FUTURE SCOPE",International Journal of Computer Engineering & Technology (IJCET)Volume 7, Issue 4, July–Aug 2016, pp. 12–24, Article ID: IJCET_07_04_002.

[4] Pooja Bisht¹, Kulvinder Singh²,"BIG DATA SECURITY: A REVIEW OF BIG DATA, SECURITY ISSUES AND SOLUTIONS", Pooja Bisht et al, International Journal of Computer Science and Mobile Computing, Vol.5 Issue.7, July- 2016, pg. 142-147.

[5] Ali Gholami and Erwin Laure,"BIG DATA SECURITY AND PRIVACY ISSUES IN THE CLOUD",International Journal of Network Security & Its Applications (IJNSA) Vol.8, No.1, January 2016.

[6] Vijey Thayanathan & Aiiad Albeshri, "Big data security issues based on quantum cryptography and privacywith authentication for mobile data center", Procedia Computer Science 50 (2015) 149 – 156.

[7] "The Big Data Security Gap:Protecting the Hadoop Cluster"

[8] Raghav Toshniwal & Kanishka Ghosh Dastidar & Asoke Nath," Big Data Security Issues and Challenges", International Journal of Innovative Research in Advanced Engineering (IJRAE) ISSN: 2349-2163 Issue 2, Volume 2 (February 2015).

[9] Khantil Choksi¹, Niriksha Dalal², Mr. Kshitij Gupte³ and Dr. Anjali Jivani⁴," Security And Privacy Challenges In Big Data",International Journal of Latest Trends in Engineering and Technology Vol.(7)Issue(3), pp. 313-318.

[10] Renu Bhandari¹, Vaibhav Hans^{2*} and Neelu Jyothi Ahuja³," Big Data Security – Challenges and Recommendations" ,International Journal of Computer Sciences and Engineering Open AccessReview Paper Volume-4 , Issue-1 E-ISSN: 2347-2693.

[11] Rongxin Bao¹ Zhikui Chen¹ Mohammad S. Obaidat²,"Challenges and techniques in Big data security and privacy:A review"

[12] Gang Zeng , "Big Data and Information Security",ISSN (e): 2250 – 3005 || Volume, 05 || Issue, 06 || June – 2015 ||International Journal of Computational Engineering Research (IJCER)