International Journal of Scientific Research & Engineering Trends
Volume 7, Issue 1, Jan-Feb-2021, ISSN (Online): 2395-566X

Big Data Security Challenges and Prevention Mechanisms in Business

Anusha Dissanayake
Department of Information Technology
Sri Lanka Institute of Advanced Technological Education
Kandy, Sri Lanka
asandarenu@gmail.com

***Abstract*- Sensitive data analytics have reached gradually a smart area in the business world over the past few years. Various research has emphasized the importance of this field in augmenting the business performance in the industry. Excessive collection of data is making harmful effects on human beings. These data are extremely vulnerable for outsiders thanks to their hidden value. Big data gives us more advantages to make progress in many fields including business. It improves the competitive advantage of companies and to add value for many social and economic sectors. Sensitive data sharing brings new information security and privacy challenges. Earlier technologies and methods are no longer appropriate and lack performance when applied in a big data context. This research focuses on investigating the challenges and providing viable solutions to minimize the risk of the threats of user data.**

***Keywords*- Big data, security, privacy, challenges, threats.**

I. INTRODUCTION

Over the last few decades, the security of sensitive data will become the basic question among business societies. Data is produced at an increasing rate. Internet-connected devices and systems promote this trend all over the world. When the advantages of big data increases, a huge debate has been created about the security and privacy of the data [1].

Meeting the challenges of big data would be a difficult task with the rising of the Internet and related systems. So, this research is focused on identifying the security challenges of big data. Big data is becoming an emerging area when it comes to business. Since big data is being operated via communication technologies, and various protocols, security issues regarding availability, data integrity, data confidentiality, and authentication can be existed [2].

These issues hamper operational inefficiency, robustness, and throughput. For a sustainable and robust user experience, security and privacy issues need to be adequately addressed.It is expected to explore the current security procedures and examine the loopholes of those procedures. In addition to that study is aimed to introduce new security safeguards to protect the value of big data.

In terms of the people who have already proposed and analyzed this topic, I have made some recommendations and suggestions to overcome the security challenges.

II. BIG DATA: DEFINITION

Big data is a collection of data that is massive in size while growing exponentially with time. It is more complex data sets, especially from new data sources. These data sets are so voluminous that traditional data processing software can't manage them. Big data can be used to address business problems easier and convenient way. It can be both structured and unstructured– that in undates a business on a day-to-day basis. But it's not the amount of data that's important, what organizations do with the data that matters.

Big data can be analyzed for insights that lead to better decisions and strategic business moves.The use of big data is becoming common these days for companies to outperform their peers. These sensitive data would help the organizations to create new growth opportunities as they have a lot of information about the products and services, buyers and suppliers, consumer preferences that can be captured and analysed [3].

III. CHALLENGES

Some of the challenges have been identified as the critical factors that can be hampered by the security of the

sensitive data of the users. It is identified that big data storing issues, big data transferring issues, and privacy and security issues are the major challenges [4]. The Public's sensitive information is connected with social media sites would be a major threat. It has descriptively analyzed the big data life cycle phases, associated risks and defensive mechanisms. As per the findings, phases are data collection, storage, analytics and knowledge creation [4].

Transformational issues of big data are linked with the nature of the transmission of the data. Mostly, data are transferred in a networking environment. Data in transit has to be fully secured [5]. These valuable data can be used by outsiders to predict customer behavior and ultimately to gain competitive advantages in their business world. In big data architecture, the data is usually stored on multiple tiers, depending on business needs for performance vs. cost. For instance, high prioritized data will usually be stored on flash media. So that locking down the data and providing access control is somewhat critical in that environment. Furthermore, a lot of data are stored in non-relational databases which leads to a lack of security.Cloud computing platforms are found as another challenge for big data security.

Data storage contains sensitive data of users that can be leaked to outsiders as there is no control over them [6]. Compliance is another issue when moving to the cloud. There is no guarantee of where the data is stored and who will be accessing the data. Even though the volume and the speed of processing are overwhelming, the security of the data has to be taken care of seriously in the cloud environment.

IV. PREVENTION MECHANISMS

It is also studied the possible technologies to compensate for these identified challenges. The main issue found here is the security of sensitive data. There is an increasing trend of analyzing patterns of customer data to utilize customer preferences for the benefit of their businesses. Some recommendations have been made to compensate for the security challenges. These recommendations include the implementation of SSL, NoSQL, OLAP, RAID and Hadoop like technologies to minimize security attacks [7]. It is proven that these technologies are cost-effective, affordable mechanisms. In addition to that results have shown that big data audits are a better way to track illicit activities. Furthermore, the use of strong algorithms RSA and AES are another best effort to safeguard the security of the system.

There are some other mechanisms to safeguard the customer's sensitive data. Among them digital certificates and digital signatures play important role in preserving the security of big data. They ensure from where the data is coming and to whom it is delivered. No intruder can get access to the data. MD5, SHA are some highly reliable encryption mechanisms that operate inside this framework [8]. Their primary aim is to preserve privacy and the confidentiality. Implementing cryptographic techniques to preserve confidentiality and privacy of the data is also important to protect data. Senders and receivers are sharing keys in this technique. To open the message receiver should know the key. This is how the protection is done. There are two types of mechanisms.

One is Symmetric Key Encryption where a shared key is used. Another mechanism is use of two keys one as private and another as public. Public Key is shared and the receiver should have the Private Key to decrypt the message. It also suggests that proper access control is a better solution to mitigate those threats by providing strong authentication and authorization mechanisms. It is also found that even though Hadoop is a better technology for managing a large volume of data, there are some weaknesses associate with that such as poor endpoint protection and no real-time protection. However, when Hadoop is connected with Kerberos which provide protection against eavesdropping, unauthorized access and replay attack [9].

It is found that real-time monitoring would provide a better solution for maintaining data security. Big data audits can be used to safeguard the security of big data. It guarantees endpoint security while providing cost-effective methodology. It is accepted by the current market with its affordable nature. Cloud Computing platforms are found as a basic issue for big data security. To protect data at rest, encrypting database content is an efficient way. Meanwhile, to protect transit data it is identified another technology known as Secure Socket Layer (SSL) [10].

According to further findings of the earlier study, granular access control is also well-known technique to provide better security and usability. It is also compatible with NoSQL. To prevent data storage issues, object-based storage systems can be used that are securely managed data as objects rather than traditional files. The Advantage of this technology is this is coupled with RAID technology. In addition to that study recommends a backup strategy as a better way to secure data in transit like cloud environment. FTP and SSL are also recommended to protect data in a networked environment. Furthermore, it is found that data visualization, social network analysis and data mining are proven techniques to handle data securely and effectively. Dryad is a better programming tool that is capable of processing clustered data. Another one is Apache Mahout which is also scalable machine learning technique that provides secure data analysis. It is emphasized that the need of addressing these issues to protect the privacy of customer data. It is introduced new technique known as K-anonymity technique which prevents the leakage of sensitive information. Hence, preserving the privacy of data.

V. CONCLUSION

It is found that big data privacy issues are critical in today's business world. Companies still use traditional methods to protect their customers' sensitive data. But it doesn't give proper protection. There are some security loopholes in current practices. Researches are still evolving to find some superior measures to get rid of the mentioned challenges.

This research's primary goal was to study the security challenges of big data in the business field. The major achievement of this study was to understand the basic security threats in the business area and how to provide solutions to them. Managing risk is a critical issue. Major threats identified are loss of privacy, spoofing, spamming, unauthorized access and replay attacks. Due to those attacks, customer's sensitive data is not secured. By adapting to the latest emerging technologies, it is proven that there is a possibility to minimize those risks. However, achieving zero risks is not practicable with the cloud-based infrastructure. It is the primary responsibility of owners and managers to ensure their security level in managing big data on their own.

VI. RECOMMENDATION

Big data has formulated extensive problems for business companies due to the difficulties of managing them. One solution as this research is focusing on treating the data set with sophisticated tools and converting the data to a less readable format like semi-structured format. Structured data are easily leak able to outsiders by applying simple discovery methods. ETL (Extract, Transform and Load) systems in data warehouses are ideal for this process were extract, transform and load data from multiple databases. In other words, this is the process of converting structured data into a semi-structured format. As an example, NoSQL is one of the developed new tools for doing this job. That can handle large databases from different websites. This strategy provides so many advantages over the traditional DBMS. Due to its simplicity, users are overwhelmed to accept this product.

Fast access capability, higher scalability, the capability of integrating a large number of servers and lower cost has led them to stay in the front line. Not only No SQL provides these functions but also comes up with advanced security features. Nowadays data breaches are a special concern in the big data world. The latest versions of No SQL have come up with advanced security features. As an example, MongoDB has become the leading database technology for the last few years.

It provides authentication, authorization, encryption and audit facilities. It uses a higher authentication mechanism such as Kerberos, LDAP and windows active directory. It permits granular permissions that enhance the authorization of users. Practicing a better cryptographic technology is another acceptable solution to big data security. Unlike traditional techniques,Homomorphic cryptography provides additional computing even after the encryption has been taken place.This technique guarantees 100% confidentiality of data. MapReduce framework is well compatible with Homomorphic technology.To ensure the further performance it is recommended to use key exchange known as CBHKE (Cloud Background Hierarchical Key Exchange). Centralized Security Management is also recommended to enhance the data security.

Another recommended solution for big data protection is to enhance controls by integrating controls. Public Key Infrastructure (PKI) is ideal for facilitating access control with the help of multi-factor authentication and a very reliable mechanism. That could be integrated with secured communication based on TLS-SSL protocols and VPN tunnels.

Finally, implementing these technologies would not be enough to cover all threats. It is recommended to conduct regular audits and adapt to security policies for best practices to get rid of security challenges.

REFERENCES

[1] Hussain, N., Choudhury, B., & Rakshit, S. (2014). A Novel Methods for Preserving Privacy in Big-Data Mining. International Journal of Computer Applications, vol.103, issue. 16. Retrieved from https://pdfs.semanticscholar.org/cbf8/982a2421f7a3ed4d328eff84ba7f73da121b.pdf .

[2] Hamami, O.(2014). Big Data Security: Understanding the Risks. Business Intelligence Journal, vol.19, issue 2, pp.20-26. Retrieved from https://search-proquest-com.ezproxy.csu.edu.au/docview/1539323005/fulltextPDF/96CBB5EDB4D743F8PQ/1?accountid=10344.

[3] Taylor, A., Angie, M., Chen, Y., Rachel, L., & Zane, M. (2017). Big data analytics: Megatrends to business success. Internal auditing, vol.32, issue. 11. Retrieved fromhttps://searchproquestcom.ezproxy.csu.edu.au/docview/1939751370/fulltextPDF/7EB439992CDB4517PQ/1?accountid=10344.

[4] Sankaram Alladi, B., & Srinivas Prasad, D. (2017). Big data life cycle: security issues, challenges, threat and security model. International Journal Of Engineering & Technology, 7(1.3), 100. Doi: 10.14419/ijet.v7i1.3.9666.

[5] Baesens, B., Bapna, R.., Marsden, J., & Vanthienen, J. (2016). Transformational issues of big data and analytics in networked business. MIS Quarterly, Vol. 40, issue.4, pp.807-818. Retrieved from http: //we b .a.ebscohost.com.ezproxy.csu.edu.au/ehost/pdfviewer/pdfviewer?vid=1&sid=908bc0c9-3c80-4c1f-933a-56add6807585%40sessionmgr4007.

[6] Payolotsky, J. (2013). Privacy in the Age of Big Data. American Bar Association, Chicago, vol.69, issue.1, pp.217-225. Retrieved from https://search-proquest-com.ezproxy.csu.edu.au/docview/1490901635/abstract/2FCD16EBA4F947FEPQ/1?accountid=10344.

[7] Murray, M. (2016). Big Data and Intelligence: Applications, Human Capital, and Education, Journal of Strategic Security. vol.9, issue.2, pp.92-122. Retrieved from https://search-proquest-com.ezproxy.csu.edu.au/docview/1800146612?accountid=10344&rfr_id=info%3Axri%2Fsid%3Aprimo.

[8] Kude, T., Hoehle, H., & Sykes, T. (2017). Big Data Breaches and customer compensation strategies. International Journal of Operations & Production Management vol.37, issue.1, pp.56-74. Retrieved from https://search-proquest com.ez proxy.c su.edu .au/docview/1986724170/fulltextPDF/744CD316CF804DDAPQ/1?accountid=10344.

[9] Vivekanand, B., Vidyavathi, M. (2015). Security Challenges in Big Data: Review. International Journal of Advanced Research in Computer Science, vol.6, issue.6. Retrieved from https://search-proquest-com.ezproxy.csu.edu.au/docview/1725326370/fulltextPDF/8A15351D57B4449CPQ/1?accountid=10344

[10] Ferguson, Boucher, R. (2012). The storage and Transfer Challenges of Big Data. MIT Sloan Management Review, vol.53, issue.4, pp.1-4 Retrieved from https://search-proquest-com.ez proxy.csu.edu.au/docview/1023761999?accountid=10344&rfr_id=info%3Axri%2Fsid%3Aprimo