Security and Privacy in Big Data: A Challenge

Dr. Geetika
Associate Professor
Government P.G. College
Sector 9, Gurugram
Gurugram University
Email: geetika1608@gmail.com

Abstract

Nowadays as the technology is growing and spreading in the entire world at a very fast pace as a result of the use of networking services, smart and portable devices, and other online services on the Internet is also increasing. So it is observed that there is a noticeable and considerable increase in the amount of data to be used by various users in different ways. So the main challenge faced here is how to handle this very large amount of data in secure and efficient ways. All over the world, almost every organization is trying in various ways to deal with and handle this huge amount of data. We can say that, the challenge is not only to storing and analyzing this big data with old traditional systems, but also it has been challenging maintaining privacy issues and security issues.This is a very important concern as security and privacy are a very important concerns. What makes data big? Think about all the billions of different types of devices that are now using the internet in different ways like smartphones,tablets and other electronic devices which use(IOT) sensors. So it is very obvious that big data security issues are challenging. For this reason, I have tried to discuss the concerns of big data and data privacy and security issues. "Big data" is just the term used for all the data that a business collects in the area of business interest keeping in mind for finding hidden patterns and trends within the data in different ways. These, can be helpful in achieving faster service delivery, higher customer satisfaction, more revenue, etc. On the other hand, the architecture used to store big data should be handled very carefully to overcome security issues for unwanted criminal activities and some malware. But the main problem is that many of the tools avail for handling big data issues are open source and lead to security issues. Usually, these tools are not designed with keeping the security and privacy of data in mind as a main key function, which leads to more big data security and privacy issues. Some of the commonly faced big data security issues are Distributed frameworks, N o n - relational data stores, Storage ,Endpoints, Data mining solutions, regular auditing ,Access

Reference to this paper should be made as follows:

Received: 06.02.2023
Approved: 20.03.2023

Dr. Geetika

Security and Privacy in Big Data: A Challenge

RJPP Oct.22-Mar.23, Vol. XXI, No. I,

pp.001-006
Article No. 1

Online available at *:*
https://anubooks.com/rjpp-2023-vol-xxi-no-1

controls, Data provenance, etc. Big data needs some more extra requirements and techniques for security and privacy in all operations like: data gathering, storing, analyzing, and transferring. In this paper, I have discussed various security and privacy faced while using it.

Keywords

Centralized, data mining, encrypted authentication,

Introduction

The main point is What is big data? This term refers to data that is very large in amount, and needs fast or complex operations that it's very difficult and sometimes impossible to handle using available traditional data handling methods. The popularity of big data has been increasing because with big data we achieve Customer satisfaction , Focused and Targeted Promotions, Risks Identification, Innovation,More Supplier Networks, Cost optimization and Improved Efficiency. As we all know that benefits bring challenges also. There are challenges in big data like security issues (main concern), ethical issues, sometimes the intentional abuse of big data by mischievous malevolent players available on the internet, and unintentional misuse. In this paper we are discussing about big data security issues and how can we handle these. As we all know that nothing can be 100% secure but if we follow the right methods then we can secure data .

Security Issues in Big Data

As we all know that security of data is the most important aspect of an organization so here are some big data security issues are listed below which should be taken care of.

a. Distributed frameworks of big data . We know that big data is distributed on many systems for faster analysis and processing.So Mostly implementations require retrieving this from distributed frameworks and in these types of processing fewer data is processed by one system. The system may require more systems as a result of which we may face security issues with data. This is a very alarming situation. Sometimes the organization has to face serious issues.

b. Non-relational data stores. As the security of data is taken care of in relation to the database by implementing checks at various levels and No SQL databases in big data, as a result, we may face a lack of security.

c. Storage pattern of big data. Usually, big data architecture, uses multiple tiers for data storage , as per business needs for better performance and optimum cost. For example, high-priority(the data required frequently and quickly) we may call it HOT DATA will be stored on a high-speed storage device like: flash media so that it can be retrieved quickly.

d. Real-time security/compliance tools. Nowadays with the advancement of technology data is required quickly and processing in real-time and this type of data processing on large large amounts of data is undoubtedly a big big challenge for big data handlers, some times there is no proper of support system for this massive real-time processing of a large amount of data frame as compared to conventional static data. A huge amount of information is generated from data stores from time to time ; so it becomes important to find a way to ignore the likely false positives, so that researchers, data handlers and designers can be focused on the true results.

e. Data mining techniques . Data mining techniques are majorly used in almost every big data environment as these techniques draw various useful patterns which are very important in suggesting business strategies effectively. That's why , it is not only very important to be sure that data is secured against external threats, but to protect from insiders (may be an employee or other person of competitors) who can misuse network privileges given to them. They can steal and reveal sensitive information and can be very dangerous for businesses. So try to avoid these risks by adding layers of security.

f. Endpoints. As data breaches can happen if unauthorized access is done so , it is very important to keep tracking the logging system of the organization. Always try to install and use such types of security solutions that can draw logs from endpoints whenever required . This will help validate the authenticity of endpoints, Otherwise, one may face compromised security

g. Encrypted authentication/validation. Nowadays, As in the IT industry, almost everyone is an expert and can steal or corrupt the sensitive informations for some reason so, so it is very important to provide a system that uses encrypted authentication/validation of the credibility of the user and access permission is given to one or not.

h. Data provenance. We know that metadata is data about data and it primarily concerns with it. As this can be very helpful in determining from where data came from, who has accessed and handled it, or what activity was done with it. On the safer side , this type of data should be analyzed quickly to minimize the risk . If action is taken timely then the loss can be minimized. Privileged users of organizations who are engaged in this type of activity must be trustworthy and we should have some mechanism to closely monitor them so to ensure they cannot create serious problems for big data security issues.

With time we are observing that the percentage of a data breaches in different industries is increasing day by day. The data theft is happening everywhere if one ignores security even for some time. This is really alarming and proper security measures are required. This can have such an impact on industry that the shape of the whole organization can be changed in every aspect.

Tips for Securing Big Data

Here are some basic steps you can take to secure your big data:

Think about security/privacy issues before starting a big data project. We pay all attention to designing but here we have to consider that security is equally important as designing and other activities. We shouldn't wait for a some undesired data breach incident to happen and then secure data. There should be a technically alert and expert security team and this team and others stakeholders should be involved in your big data project so that they may have a serious data security discussion and prepare strategies accordingly before installing and feeding data .

Accountability should be Centralized. It is very important to keep uniformity in policies in decentralized data. So there should be centralized accountability so that consistent policy enforcement and access control can be enforced on data to maintain security .

Encrypt data at every stage. We always to keep data secure but still something undesirable penetration can happen. So it will be very dangerous to keep data in understandable shape. This may seem very helpful in data security .So always use some data encryption techniques at the file layer. SSL encryption can be helpful here because it moves between nodes and various applications. Encryption protects data from malicious users. This is a cost-effective way to handle various data security threats in big data.

Separate your keys and your encrypted data. If we store encryption keys and encrypted data on the same server it may be very risky . It can be compared to locking the front door of house and then leaving the keys in the lock. So the key management system should be secure so that encryption keys could be kept safely and separately from the data system you're trying to protect.

Use secure automation. While working in a multi-node environment, It becomes difficult to deploy consistency . Automation tools should be used which can be helpful.

Have a watch on logging into the system and data. If one feels that a breach has been happened then by tracking logged activities one can find the breach point and find solutions. It gives you a place to look where the fault happened and why?.

Without an event trace, one feels blind and may face big problems. By observing logging MR requests and some other cluster activities one can easily track problems and threats whenever required."

Eliminate unnecessary information.One of the smartest ways to prevent a data breach and protect your data is not to keep very sensitive data in the first place. Always analyze the data and remove unnecessary and duplicate data from time to time.By conducting these regular audits, organizations can keep the data necessary for their business operations, while removing the remains. It gives us the benefit of focusing the analytics tasks where they're most important and needed.

Conclusion

In this paper, we have found and discussed that the security and privacy issue of big data is a challenge for an organization. Big data needs extra requirements for security and privacy in data gathering, storing, analyzing, and transferring. As big data analytics is a new discipline. Naturally, mistakes will be made. The key thing is to learn from these mistakes and improve safety. By implementing security measures and ethical guidelines, we can reap big data's benefits while mitigating its risks. In this paper, we examined studies on big data security and privacy, comparatively. It is hoped that this study would help us understand the big data and its ecosystem better and develop better systems, tools, structures, and solutions not only for today but also for the future.

References

1. Vijey, T., Aiiad, A. (2015). "Big Data Security Issues Based on Quantum Cryptography and Privacy with Authentication for Mobile Data Center". *Procedia Computer Science*. vol.50. Pg. **149-156.**
2. Matturdi, B., Zhou, X., Li, S., Lin, F. (2014). "Big Data security and privacy: A review". *Big Data, Cloud & Mobile Computing, China Communications* vol.11. issue.14. Pg. **135-145.**
3. Chen, C.L.P., Zhang, C.Y. (2014). "Data-Intensive applications, challenges, techniques and technologies: A survey on Big Data". *Information Sciences*. vol.275. Pg. **314-347.**
4. Miloslavskaya, N., Senatorov, M., Tolstoy, A., Zapechnikov, S. (2014). "Information Security Maintenance Issues for Big Security-Related Data". *Future Internet of Things and Cloud (FiCloud)*. Barcelona. Pg. **361-366.**
5. Cloud Security Alliance Big Data Working Group. (2013). "Expanded Top Ten Big Data Security and Privacy Challenges". April.

6. Ibrahim, A.T.H., Ibrar, Y., Nor, B.A., Salimah, M., Abdullah, G., Samee, U.K. (2015). "The rise of "big data" on cloud computing: Review and open research issues". *Information Systems*. vol.47. Pg. **98-115.**
7. Adluru, P., Datla, S.S., Xiaowen, Z. (2015). "Hadoop ecosystem for big data security and privacy". *Systems, Applications and Technology Conference (LISAT), Long Island, Farmingdale*. NY. Pg. **1-6.**
8. Divakar, M., Shrikant, K., Shweta, J. (2015). IBM, "Big data architecture and patterns. Part 1: Introduction to big data classification and architecture". http://www.ibm.com/developerworks/library/bd-archpatterns1/. Accessed Date: 1 August.
9. https://www.cio.com/article/2390490/big-data/how-to-secure-big-data-in-hadoop.html
10. Saraladevi, B., Pazhaniraja, N., Victer Paul, P., Saleem Basha, M.S., Dhavachelvan, P. (2015). "Big Data and Hadoop-A Study in Security Perspective". *Procedia Computer Science*. vol.50. Pg. **596-601.**
11. Kumar, A., HoonJae, L., Singh, R.P. "Efficient and secure Cloud storage for handling big data". *Information Science and Service Science and Data Mining (ISSDM)*. Pg. **162-166.**
12. A survey on security and privacy issues in big data (PDF Download Available). Available from: https://www.researchgate.net/publication/300413833_A_survey_on_security_and_privacy_issues_in_big_data [accessed Mar 12, 2018].
13 https://www.alienvault.com/blogs/security-essentials/9-key-big-data-security-issues.
14. http://bigdata-madesimple.com/data-security-issues-main-challenges-in-2017.
15. https://careerfoundry.com/en/blog/data-analytics/is-big-data-dangerous/#:~:text=Broadly%20speaking%2C%20the%20risks%20of,crime)%2C%20and%20unintentional%20misuse.