ISSN No:-2456-2165

Security and Privacy in Big Data

Hasina Masud Khadas, Heeba Aslam Mujawar, Srushti Sureandra Shetye
Student, MSc(IT), D.B.J College, Chiplun, Ratnagiri.

Abstract:- Data which is large in size, volume and processing extreme complexity and generated mainly by business is collectively known as Big Data. Due to their complexity, this type of data is not possible to be stored in the traditional database systems. They require specifically designed tools for them such as Hadoop. Big Data might involve sensitive data which if left unsecured can be exploited by cybercriminals for misusing it. As with time, the Big Data just keeps on growing, it is more difficult to secure/protect the data. Every organization has to secure and maintain privacy of their Big Data to avoid any negative impact from the data collected through the hands of cybercriminals. Big Data security involves guarding the data and analytics processes performed on it from cyber-attacks such as theft or any other kind of malicious attacks. This research paper intends to explore the security and privacy concerns/challenges occurring while securing the Big Data.

Keywords:- *Big Data, Security, Privacy, Encryption, Cybercriminals.*

I. INTRODUCTION

Big Data is useful for a business to make better decisions for their businesses. Data is always exponentially growing, so as they get continuous streams of data, it is essential for an organization to properly manage this data in order to prevent any form of cyber attacks on it. Research Data Alliance [1] states that – " Big Data security is the processing of guarding data and analytics processes, both in the cloud and on- premise, from any number of factors that could compromise their confidentiality". In a nut shell, it can be said as the term ' Big Data Security' comprises of the security measures undertaken for data collected and the security tools used to perform data analytics on it. It is important to secure the Big Data because the organization collecting the big data might end up facing legal actions even if it unintentionally becomes the source of leaking sensitive and confidential data like personal customer's information or say credit card numbers to outside harmful sources. Such organizations involved with big data need to comply with and follow the rules and regulations of General Data Protection Regulation(GDPR) that deals with basic data security measures. Hence, it is extremely important to overcome the challenges faced while securing the big data. Whereas Big Data privacy is concerned with – [2] " The more data you collect, the more important it is to be transparent with your customers about what you are doing with their data, how you are storing it, and what steps you are taking to comply with regulations that govern privacy and data protection". By this, it can be clearly understood that the data that is collected by the businesses have to be collected by the data owner's consent. But not only consent is enough to go ahead with their data, but how their data was collected and in what way the organization will use their data, the way of storing it should be communicated with the data owner to keep the transparency. Proper management of data is required to prevent privacy breaches. However, while securing and preserving the privacy of big data, the organization has to go through may obstacles in their way.

II. CHALLENGES IN SECURING BIG DATA

A. Distributed data

A distributed data can lead to security issues as it requires more attention in securing each distributed data. Today's commonly used tool is Hadoop for big data processing and it's storage which was designed without considering the security aspect in it. As the design is free from security controls on it, the big data is more vulnerable to any malicious attack on it. A distributed data technique helps in sharing the workloads of data but at the end of the day it is more tedious to secure it and look after the security issues arising from all the distributed data.

B. Securing NoSQL Databases

Normally, to store data we prefer our traditional database systems that allows us to store data as set of records. Due to it's scalability and diversity, it is not the solution for storing Big Data. Hers is when the NoSQL database design comes to Big Data's rescue. In NoSQL database, data is stored as information in JSON documents. Alex Bekker in his blog mentions- [3] " NoSQL databases are continuously being honed with new features. And just like we said in the beginning of this article, security is being mistreated and left in the background. It is universally hoped that the security of big data solutions will be provided externally. But rather often it is ignored even at that level". So it is safe to assume that in NoSQL database, security aspect should not be overlooked while updating the features of NoSQL.

C. Endpoint Vulnerabilities

[4] Security problems with big data often start at the point of entry. The source of the data flowing into a company's big data system could be compromised. Suppose a cybercriminal hacks into the data source, he can easily manipulate that data which can then end up going to it's destination for further data analytics. In this way, false and malicious data will enter the system and moreover the results that would be derived from this data will also be incorrect. Therefore, it is important to validate the authenticity of data at the endpoints to overcome it's vulnerability for precise and productive business decisions.

D. Data Mining

Data mining involves finding the common and similar patterns from the raw data to make it turn into useful information. However, the data collected can be sensitive to deal with e.g. credit card numbers or any personal information or financial information. To avoid leaking of sensitive information, it is necessary to provide an extra layer of security hereby preserving the anonymity of data owner. Otherwise, if no anonymity is maintained then the organization can mostly face severe consequences out of it.

E. Access Controls

Data is divided into items for granular access by the people. In this mechanism, people can access the data only that they are allowed to see, all other items in that data set are kept secret. However , the big data tools are not fit for providing granular access. So to implement the granular access – [5] " The parts of needed data sets, that users have right to see, are copied to a separate big data warehouse and provided to particular user groups as a new 'whole'. For a medical research, for instance, only the medical information (without the names, addresses and so on) gets copied ".

III. CHALLENGES IN BIG DATA PRIVACY

A. Data Breaches

When a confidential information is leaked out in hands of some untrusted party, it is said to be data breaching. The data is accessed without authorization in such case. Data breach can happen due to weak passwords, out of data software, poor authorization mechanism which invites and is vulnerable to attacks. When sensitive information of users is leaked and eventually their privacy is invaded through hands of the organization, the organization will have to face legal actions due to it's poor authorization system. So, strong security measures needs to be practiced upon Big Data in order to avoid data breaches.

B. Data Brokerage

Big Data is sometimes bought from a third party (that collects the data) by the organization to fulfill its business decisions. And using such data, the further operations on data are carried out to get the results out of it needed to make an organizations business successful. But it is important to check if the data bought is correct, accurate and not some false data. Otherwise false/incorrect data will lead to incorrect outcomes. Therefore, the organization while purchasing the Big Data from a third party has to check if it is a trustable source that provides accurate data.

C. Data Discrimination

Data can be unfairly used for analytics purposes that supports data discrimination. Suppose a data consisting of user's information is discriminated by age, gender, religion or caste by developing an algorithm to support data discrimination. With the use of such algorithm the outcome will also be unfair. Organizations should always choose fairness while using the data.

IV. CONCLUSION

In conclusion, though Big Data is the fuel of a business's growth and success, the challenges that comes with it for securing and protecting privacy is foremost aspect too. The greater the data , the greater it's risk for privacy invasion and more efforts in securing it. Suppose, a data breach incident like an airlines's database is cyber attacked and the data of users are sought by the attacker, thus all the credit card information is leaked and goes in the hands of the attacker now can prove cost an arm and leg for the airline. At the same time, the credibility of such airline is questionable. This attack will lead to damage on reputation of airline and hence, their business might see a collapse. Such incident clearly manifest that, no matter how small or big or critical or uncritical the business is, the importance of securing the big data by overcoming the hurdles cannot be overlooked.

REFERENCES

[1]. Research Data Alliance . "Big Data Security - Issues, Challenges, Tech & Concerns." Research Data Alliance , https://www.rd-alliance.org/group/big-data-ig-data-security-and-trust-wg/wiki/big-data-security-issues-challenges-tech-concerns
[2]. Informatica . " Big Data and Privacy: What It Is and What You Need to Know." Informatica,https://www.informatica.com/in/resources/articles/what-is-big-data-privacy.html.
[3]. Bekker Alex . "Buried under big data: security issues, challenges, concerns". ScienceSoft ,04 April 2018 ,https://www.scnsoft.com/blog/big-data-security-challenges.
[4]. Kowalke Peter . "Six Security Vulnerabilities with Big Data" . ToolBox , 27 June 2017 ,https://www.toolbox.com/tech/it-strategy/blogs/six-security-vulnerabilities-with-big-data-062717/
[5]. Bekker Alex . "Buried under big data: security issues, challenges, concerns". ScienceSoft , 04 April 2018 ,https://www.scnsoft.com/blog/big-data-security-challenges