International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 08 | August -2020 ISSN: 2582-3930

Cyber security issues in Big Data Analytics: A survey

Thangamma K C[1], Akhila C V[2], Ayesha Taranum[3]

[1]Department of ISE, GSSS, Institute of Engineering & Technology for Women.
[2]Department of ISE, GSSS, Institute of Engineering & Technology for Women.
[3]Department of ISE, GSSS, Institute of Engineering & Technology for Women.

Abstract - Big data analysis in technology requires the ability to conduct data analysis. The collection of large quantities of digital information which has to be analyzed. To visualize and draw observations that would make it possible to predict and stop cyber threats from occurring. In this paper, we are going to concentrate on how large data will enhance the information security when best practices are inculcated.

***Key Words*:** Big data, security, block chain, security issues.

1. INTRODUCTION

The cyber-safety industry is aggressively developing new technologies to mitigate emerging threats to rising handheld and IoT system segments. Big Data provides enormous possibilities and affects all markets, market functions and fields: customer engagement, emerging technology design, and medical science. Threats to computer security are growing.

Increasingly advanced methods of attack used by cyber criminals and the growing involvement of malicious insiders in many recent large-scale breaches of security clearly suggest that conventional approaches to information security are no longer able to keep up. Technologies which are likely to have the most scope for data security include

Big Data: allows for automated risk control and predictive analysis. The implementation would be motivated mainly by the need to detect use and behavioral trends to aid in detecting anomalies in security operations.

Machine Learning: This helps security teams to prioritize corrective steps and automate multiple variables in real-time analysis. Machine learning algorithms will use the large pools of data collected by businesses to zero in on the root cause of the attack, and address observed anomalies in the network.

Blockchain: Unable to modify or delete the data stored on blockchain by default. Tractability of process performed out onblockchain is vital to the creation of a trusted network between endpoints. In comparison, blockchain's global nature greatly increases the cost of breaching blockchain-based networks which discourages hackers.

Big Data Protection: It is the collective term for all approaches and techniques used to safeguard based on analytical processes from threats, embezzlement or other malicious activities which does damage or impact them. Unlike so many other cases of cyber-security, the big data variant is concerned with either online or offline attacks. Big data "leads in the quantity of IP-equipped endpoints from this incredible escalation. It is really just the word for all the data available in a given field that a company gathers to identify hidden patterns or trends within it. Those can be leveraged until revealed by analytical methods to produce an better outcome along the way. Big data management attacks on an enterprise may have significant financial implications such as damages, legal costs and penalties or sanctions. There are three big best practices in Big Data Security or rather problems that will determine how an enterprise is setting up its BI security.

2. Main security issues related to big data

Big data is a key target for hackers. Professionals in computer protection need to take an active role as soon as possible. The fact is that the need to make fast business decisions will result in the safety professionals being left out of key decisions or being seen as business growth inhibitors**.** Big data relies heavily on the cloud but the big data protection threats aren't generated by the cloud alone. Apps, particularly unknown pedigree third-party apps, may easily introduce risks into corporate networks when their security mechanisms do not follow the same requirements as existing corporate protocols and data governance policies.

1. Many Big Data applications simply spread large computing jobs for faster analysis across many networks. Hadoop is a well-known instance of open source software involved in this, and was originally without any kind of protection. Distributed processing does mean less data being handled by any one system, but it does mean a lot more systems where security problems can occur.
2. The data is typically processed in big data architecture on various levels, depending on the output vs. cost needs of the company. For example, "soft" data of high significance would usually be stored on flash media. Thus locking down storage would mean a tier-conscious approach.

3. Security solutions that pull logs from endpoints will need to verify those endpoints' validity, or the analysis won't do any good that the real violations can be concentrated on human talent.
4. They produce a huge amount of information; the trick is to find a way to disregard the false positives, so that the real violations can be concentrated on human talent.
5. Many big data environments are at the heart; they identify trends that indicate business strategies. For this very purpose, ensuring they are protected against not only external threats, but insiders who misuse network privileges to access confidential information is especially critical – adding yet another layer to big data protection issues.
6. Just as with corporate IT as a whole, it is important to have a framework in which encrypted authentication / validation verifies that users are who they claim they are and who can see what.
7. Granular auditing will help identify where failed attacks have happened, what the effects have been and what needs to be done in the future to strengthen matters. It is a lot of data in itself, which need to be allowed which secured to be useful in solving big data protection issues.
8. The provenance of data mainly involves metadata (information), which can be immensely helpful in identifying where data came from, who accessed it, or what was done with it. This type of data can typically be processed with exemplary speed to minimize the time a breach is active. Privileged users participating in this form of operation must be carefully tested and monitored closely to ensure that they do not become their own big data security issues.

3.Enforcing protection in Big Data:

There are a variety of ways companies can enforce security measures to secure their resources for analyzing big data. The encryption is one of the most popular security methods, a fairly easy tool that can go a long way. Encrypted data will be of no use to external actors like hackers if they do not have the key to unlock it.

Another important Big Data security tool is to create a solid firewall. Firewalls are efficient at filtering traffic that enters servers and leaves them. Organizations can prevent attacks by developing strong filters that eliminate any third parties or unknown sources of data before they happen. To build an in-depth approach, data protection must be complemented by other protection measures such as endpoint security, network security, application security, physical site security and much more. Through planning ahead and being prepared to incorporate big data analytics into your enterprise, you will be able to safely help the company achieve its goals.

4. Safety issues for Big Data:

Securing big data can pose many challenges which can compromise its security. Keep in mind that such problems are by no means confined to big data systems on-premise. We contribute to the cloud, too. Take nothing for granted as you run the Big Data application in the cloud. Consult together with your company to address these same issues with clear agreements on security service standards.

5. Large Data Securing Challenges:

- Advanced computational methods for unstructured big data and non-relational databases (NoSQL) are the newest active-development technologies. Protecting these new tool sets can be hard on security tools and processes.
- Mature security instruments effectively secure the entry and storage of data. However, from multiple analytics tools to different sites, they may not have the same effect on data production.
- Big data administrators can decide to mine data without prior notice or permission. Whether the motive is interest or criminal benefit, the security tools need to track and warn suspicious access regardless of where it originates.
- The sheer scale of a big data deployment is too high for regular security audits, terabytes to petabytes huge.
- When the big data owner does not upgrade protection for the system on a regular basis, they run the risk of data loss and disclosure.
- Security tools need to track and warn device, database or web CMS like wordpress on suspected malware infection, and Big Data Security experts need to be professional in cleaning up and know how to delete malware from wordpress.

6. Big Data reliability technologies

- **Encryption:** The encryption tools need to protect in-transit and at-rest data, and they need to do so through large data volumes. Encryption often needs to be used for several different types of data, both user-generated and machine-generated. Encryption tools will need to work with various analytics toolsets and output data, as well as specific broad data storage formats, including relational database management systems (RDBMS), non-relational databases such as NoSQL, and advanced file systems such as Hadoop Distributed File System (HDFS).

- **Unified Key Management:** Unified key management has been the best security technique for many years. It applies equally and strongly to large data environments, particularly those with a broad geographical distribution. Good practices include policy-driven automation, logging, on-demand key distribution, and key management isolation from key use.

- **User Access Control:** User access control can be the most simple network security tool, but many organizations have limited control because overhead management can be too high. This is risky enough at the level of the network and can be catastrophic for the big data platform. Effective user access control includes a policy-based approach that automates user-based access and role-based settings. Policy-driven automation handles complex user control levels, such as multiple administrator settings that protect the big data platform from internal attacks.

- **Intrusion Detection and Prevention:** Intrusion detection and prevention systems are safety pioneers. That doesn't make them any less important to the big data community. The importance of Big Data and the Distributed Architecture lend themselves to intrusion attempts. IPS allows security administrators to protect the Big Data Platform from intrusion and, should an intrusion succeed, IDS quarantines the intrusion until it causes serious harm.

- **Physical Security:** Do not neglect physical security. Plug it in when you install your big data infrastructure in your own data center, or carefully monitor the security of your cloud provider's data center. Physical security systems can deny access to data centers to strangers or staffs who has no business in sensitive areas. Video monitoring and security reports should do the same thing.

7. CONCLUSIONS

The purpose of Big Data Analytics is to obtain achievable information in real time. Big data is going to have a huge effect on current industry. This paper discusses the various security concerns and challenges in the areas of Big data analytics and cyber security.

REFERENCES

1. Big Data Analytics for Detection of Frauds in Matrimonial Websites Vemula Geeta et al | International Journal of Computer Science Engineering and Technology (IJCSET) | March 2015 | Vol 5, Issue 3, 57-61

2. Big Data Cyber security Analytics Research Report - Ponemon Institute© Research Report Date: August 2016

3. International Research of Scientific Research - Cyber security a challenge to developers, is face recognition technique a solution to this problem? Volume: 2 | Issue: 7 | July 2013 • ISSN No 2277 – 8179

4. RSA "INTELLIGENCE DRIVEN THREAT DETECTION & RESPONSE", retrieved from https://www.emc.com/collateral/whitepaper/h1304-intelligence-driven-threat-detection-response-wp.pdf on 12-Jan-2015

5. CLOUD SECURITY ALLIANCE. "Big Data Analytics for Security Intelligence", September 2013, retrieved from https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Big_Dat a_Analytics_for_Security_Intelligence.pdf on 5-Jan-2015

6. A. Jakóbik, Big Data Security. Cham: Springer International Publishing, 2016, pp. 241–261.

7. D. Mondek, R. B. BlaÅ¿ek, and T. ZahradnickÃ¡, "Security analytics in the big data era," in 2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C), 2017, pp. 605–606

8. R. Patgiri, "Issues and challenges in big data: A survey," in The 14th International Conference on Distributed Computing and Internet Technology (ICDCIT-2018), ser. Lecture Notes in Computer Science, vol. 10722. Berlin, Germany: Springer, 2018, in-press

9. Big Data Working Group, "Expanded top ten big data security and privacy challenges - cloud security alliance," Accessed on 10/12/2017 from https://downloads.cloudsecurityalliance.org/initiatives/bdwg/ Expanded_Top_Ten_Big_Data_Security_and_Privacy_Challenge s. pdf, 2013.

10. Cyber-Security Definitions; National Initiative of Cyber-security Careers and Studies (NICCS), USA. https://niccs.us-cert.gov/glossary; Access date :31/03/2016.

11. J. Oltsik, An-Analytics-based Approach to Cybersecurity, May 2015: Enterprise Strategy Group (EGS).