PREMIER Safeguard Privacy in Public Administration for Big Data Environment

Ganesh D.Ghuge[1], Dnyandev S.Musale[2], Ganesh L.Borhade[3], Pandurang D.Dahiphale[4]

[1]Lecturer, Dept. of Computer Technology, Amrutvahini Polytechnic, Maharashtra, India,
[2]Lecturer, Dept. of Computer Technology, Amrutvahini Polytechnic, Maharashtra, India,
[3]Lecturer, Dept. of Electronics &Telecommunication, Amrutvahini Polytechnic, Maharashtra, India,
[4]Lecturer, Dept. of Electronics &Telecommunication, Amrutvahini Polytechnic, Maharashtra, India

---***---

Abstract –*data is very important in life .data are collectively collected from various streams like social network, website etc. increase the amount of data available provides opportunity and problem also. we have to manage data in very efficient manner and we have to provide security for this big data using various approaches like Access control, data mining solution, real time security etc.*

Keywords: Big data, data mining, security issue and Data Exfiltration Detection.

1.INTRODUCTION

The large, diverse sets of information that grow at ever-increasing rates is a big data. Big data often found from in various format from data mining and Big data can be divided a in two types like unstructured or structured. The information or data already managed by the organization in databases and spreadsheets is a Structured data; Structured data is frequently numeric in nature, while second types is Unstructured data and Unstructured data is information that is unorganized and does not fall into a predetermined model or format. It consists of a data which gathered from social media sources, which help institutions gather information on customer needs.

Publicly shared comments on social networks and websites are responsible for generation of big data. Data gathered from personal electronics devices and different apps, through questionnaires, feedback, product purchases, survey and electronic check-ins. The inputs devices and sensor in smart devices allows for data to be collected across a broad spectrum of situations and circumstances.

1.1 Pros and cons of big data

Opportunities and problems are by increase in the amount of data available. In general, having more data on potential customers should allow organization to better tailor products and marketing efforts in order to create the highest level of satisfaction and repeat business. Organizations that collect a large amount of data are provided with the opportunity to conduct deeper and richer analysis for the benefit of all stakeholders.

While better analysis is always positive, big data can also reducing its usefulness and create Noise, overhead Organizations should determine which data represents signals compared to noise and handle larger volumes of big data.

1.2 Security needs

Big data security is an under a single common category that includes all security measures and tools applied to analytics and data processes. For security needs these points are keep in mind big data system Attacks like , DDoS attacks information stolen , malicious code, DDoS attacks, ransomware .can generated from offline or online platforms .

The unpleasant of information Stolen can be even worse when organizations store sensitive or confidential data that is credit card numbers or customer related information. They failed to meet basic Information security measures to be in compliance with data loss protection and privacy mandates that are the GDPR (General Data Protection Regulation).

Six Big Data Security Challenges

Big Data competitive situation are not limited to on premise platforms. They are also affecting the cloud system. Following points are reviews the six most common competitive situation of big data on premises and in the cloud system.

1. Distributed Data

big data frameworks distribute data processing tasks every systems for better analysis. Hadoop, fg, is a framework used for storage and distributed data processing which is freely available.

Criminals (cyber) could try the Reduce Map, mapper to show wrong lists of values or key pairs, making the Reduce Map process no real value. Distributed processing might reduce the overhead on a system, but eventually lot of system means more security issues.

- Non-Relational Databases

Long-established RDB use tabular schema of rows and columns. As they cannot handle big data, cause it is diverse in structure and highly scalable. Non-RDB, also known as No

Structural query language databases, is designed to overcome the disadvantage of RDB.

Non-RDB does not use the tabular schema of columns and rows. Instead of, No Structural query language DBMS optimize storage models according to type of data. As a result, No Structural query language DBMS are flexible and scalable than their relational alternatives.

Structural query language DBMS favor flexibility and performance consideration of security. Organizations that adopt Structural query language DBMS have to set up the DBMS in a Confident environment with advance security measures.

4. Endpoint emotionally

Criminals (Cyber) can manipulate data on endpoint devices .to transmit the wrong data to data lakes. Security solutions that analysis for logs from endpoints need to validate the authenticity of those endpoints.

For example, hackers can access created systems that use lot of sensors to detect Viruses in the processes. After gaining access, hackers make the sensors show faulty results. Opportunity like that is usually solved the fraud detection methodologies.

5. Data Mining Solutions

Data mining is the main for many big data environments. Data mining tools find the patterns in unstructured data. The problem about data often contains financial information and personal information. For the reason, organizations need to add additional security layers to protect from external and internal threats.

6. Access Controls

Organizations prefer to restrict access to influences data medical records that include personal information. But people have not some access permission, like that medical researchers, still need to use this data. The solution for many organizations is to grant granular access to everyone. That means separate access and sees only the data they may need to see.

2. Big Data security issues

- Access Controls: It is mainly important for an organization to the system which is said to fully secure. exchange the data is the Permission it should be granted for authenticated users. Access control may needs to protect hacker to not hack the website by attackers, or by any intending activities. But to implement a fully secure and strong access control is a big issue for organizations as it integral part a big investment and a lot of maintenance.

- Non-RDBMS stores: Non- RDBMS like No Structural query language usually not having enough security.

- Storage: In Big data architecture, we store information on multiple tiers. Its storage depends on business needs in terms of cost and performance.

- Endpoints: Security solutions that usually draw logs from endpoints will need to validate the authenticity of those endpoints or the analysis is not going to do much.

- Real-time security compliance tools: Real-time tools generally created from large amount of data. The key issue is to find a rough to ignore wrong or rough data . So that human skill can be focused on correct breaches or valuable data.

- Data mining solutions: Data mining solutions generally check the pattern that arises from business strategies. That why, there is a need for ensuring that it is secured from both internal threats and external threats.

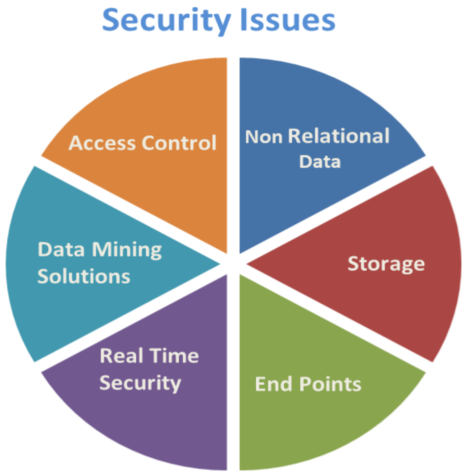

Fig -1: Security Issue

3. Safeguard solution for big data

Addressing Big Data Security Threats

The Security issue is not for new concept for big data. The security tools have ability and they have more scalable for data types. The bog data used following security techniques.

Encryption

Security for big data needs encryption tools across big data volumes for end to end. Organizations need to encrypt user as well as data which are created by machine. A response suggested that, encryption tools have to operate on multiple big data storage that is No Structural query language databases and DFS (Hadoop).

User Access Control

Network security tool are the basis of User access control. The not enough knowledge for access control measures is causing great damage for big data systems. A vigorous user control policy has to be depending on automated role-based settings as well as policies.

Prevention and Intrusion Detection

Data security thread includes intrusion. Intrusion should be detected and prevented at proper time. An IPS (Intrusion Prevention System) forces security teams to secure big data platforms from emotionally a bold by inspect network traffic. The Intrusion Prevention System work behind the firewall and separate the intrusion before it damage.

Centralized Key Management

The Concept of protecting cryptographic keys from misuse called Key management. It provides efficiency to application. Through CMS use access audit logs, secure keys as well as policies. To handling sensitive information a feasible key management system is essential for organizations.

4. Security Use Cases

1. Cloud Security Monitoring

Secure communication required for Cloud computing generally support increased financial gain for all businesses as well as efficient communication. Cloud application monitoring done by big data system. It provides to sensitive data for every host as well as to monitors cloud-hosted infrastructure. Solutions suggest supporting several relevant cloud platforms.

2. Network Traffic Analysis

In network Traffic continually moves from one end to other. For the high volume of data over the network, it was difficult to protect visibility for transactional over the network traffic.

Security analytics allow your enterprise to watch over this network traffic analysis team to protect enterprise to observed network traffic. is also make it easier in cloud security monitoring.

3. Insider Threat Detection

Attacks happen users and entities are easily difficult to detect for that reason cyber criminals can evade defences by applying legitimate credentials to access corporate resources. Link SYP handle the finder of these attacks with analytics teams. Some of the techniques, including S/UML, are applied to data from the network .

4. Threat Hunting

The Security analyst tea mostly doing work on threat hunting. Security analyst search for main indicators of a house types of threats and the field of IT infrastructure.

IT Engineer work on automate this threat of finding . It acts as an extra set of eyes for your threat hunting efforts. Threats hunting automation can help in detecting malware beaconing activity and thus alerts for its stoppage as soon as possible.

5. CONCLUSION

A new organizations use big data and its analytics tools and system to improve business strategies plan. That is giving cyber criminals to more opportunity to attack big data. Thus we list of big data security issues continue to grow. There are many more privacy and government rule & regulations for big data platforms. However, big organizations and private organizations users do not always know what is happening for that with their data and where the data is centrally stored. By chance, smart big data analytics tools can lead to new security strategies when given information. fg, security tools can reach resulting based on the correlation of security information for different systems. This is the ability to reinvent security is crucial place to the health of networks in a time of evolving cyberattacks.

REFERENCES

[1] M.N.I. Sarker, M. Wu, and M.A. Hossin, “Smart Governance through Bigdata: Digital Transformation of Public Agencies”, in Proceedings of International Conference on Artificial Intelligence and Big Data (ICAIBD), 2018, pp. 62-70.

[2] E. Agbozo, and K. Spassov, “Establishing Efficient Governance through Data-Driven e-Government”, in Proceedings of 11th International Conference on Theory and Practice of Electronic Governance (ICEGOV2018), 2018, pp. 662-664.

[3] M.Sh. Husain, and N. Khan, Securing Government Information and Data in Developing Countries, Chapter in the book: Big Data on E-Government, Hershey: IGI Global, 2017.

[4] Ya. Yang, “Research on the Opening of Government Data in Government Governance Reform in the Era of Big Data”, in Proceedings of 2nd International Conference on Judicial, Administrative and Humanitarian Problems of State Structures and Economical Subjects, 2017, pp. 246-251. DOI: https://doi.org/10.2991/jahp-17.2 [6] L. Ma, and X. Wu, “Citizen engagement and co-production of e-government services in China”, Journal of Chinese Governance, 2020, vol. 5(1), pp. 68-89.

[5] X. Xu, “FDI Forecasting in View of Big Data”, in Proceedings of 9th (2017) international conference on financial risk and corporate finance management, 2017, pp. 145-150.

[6] E. Hughes-Cromwick, and Ju. Coronado. “The Value of US Government Data to US Business Decisions”, Journal of Economic Perspectives, 2019, vol. 33(1), pp. 131-146.

[7] Y. Choi, H. Lee, and Z. Irani, "Big data-driven fuzzy cognitive map for prioritising IT service procurement in the public sector", Annals of Operations Research, 2018, vol. 270(1-2), pp. 75-104.

[8] K. Löfgren, and C.W.R. Webster, "The value of Big Data in government: The case of 'smart cities", Big Data & Society, 2020, pp. 9-14.

[9] [11] J. Sangki, "Vision of future e-government via new e-government maturity model: Based on Korea's e-government practices", Telecommunications Policy, 2018, vol. 42(10), pp. 860-871. DOI: 10.1016/j.telpol.2017.12.002

[10] [12] F.B.B. Nasution, and N.E.N. Bazin, "E-Government Maturity Model to Support System Dynamics in Public Policymaking", in Proceedings of 5th International Conference on Electrical Engineering Computer Science and Informatics, 2018, pp. 464-471.

[11] [13] G.V. Pereira, G. Eibl, C. Stylianou, G. Martinez, H. Neophytou, and P. Parycek, "The Role of Smart Technologies to Support Citizen Engagement and Decision Making: The SmartGov Case", International Journal of Electronic Government Research, 2018, vol. 14(4), pp. 1-17. DOI: 10.4018/IJEGR.2018100101

BIOGRAPHIES

Ganesh D.Ghuge,
Lecturer, Dept. of Computer Technology, Amrutvahini Polytechnic, Maharashtra, India

Dnyandev S.Musale
Lecturer, Dept. of Computer Technology, Amrutvahini Polytechnic, Maharashtra, India

Ganesh L.Borhade
Lecturer, Dept. of Electronics &Telecommunication, Amrutvahini Polytechnic, Maharashtra, India

Pandurang D.Dahiphale
Lecturer, Dept. of Electronics &Telecommunication, Amrutvahini Polytechnic, Maharashtra, India