INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 11, November 2023, pp : 334-338

AN ANALYSIS ON PRIVACY AND SECURITY ISSUSES IN THE ERA OF BIG DATA

Dr M Saraswathi[1], Mr. Vemparala Venkata Surya[2], Mr. Aravind S[3]

[1]Assistant Professor, Department of CSE, SCSVMV Deemed to be University, India.

[2,3]UG Scholars, Department of CSE, IT SCSVMV Deemed to be University, India.

DOI: https://www.doi.org/10.58257/IJPREMS32290

ABSTRACT

In recent times, Big Data computing has become an important asset for firms. In fact, almost all industries are generating data in large amounts of data. Even though it has significant potential, it is still insecure and is a target of different security issues and problems. It has been identified in this research that Big Data computing requires good protection from various privacy and security challenges. It will enable firms to identify issue and resolve them quickly before confidential data have accessed. It is an appropriate aspect to be address because users share more and more personal data and content through their devices and computers to social networks and public clouds.

Keywords: Big Data Computing, Security Issues, Cloud, Confidential Data, Etc.

1. INTRODUCTION

The Big Data is an arising area applied to manage datasets whose size is beyond the capability of generally used software tools to capture, manage, and timely assay that large quantum of data. All these data are veritably frequently unshaped and from colourful sources similar as social media, detectors, scientific operations, surveillance. It's estimated that 90 percent of the total data in recorded mortal history is created in the last recent times. Indeed, though there has been a significant interest in big data and a large number of enterprises have espoused it, there are major security challenges associated with it. In fact, due to its security issues, businesses are concerned and they calculate on the use of different ways for perfecting security. In this paper, there will be a focus on the analysis of big data computing, its security enterprises, and how these security enterprises can be addressed. Let us discusses the most important challenges to the aspects of information security and sequestration assessed by the new conditions of Big Data operations. Being Big Data such an important, it's nearly natural that immense security and sequestration challenges will arise. Big Data has specific characteristics that affect information security like variety, volume, haste, value, variability, and veracity is represented in fig 1

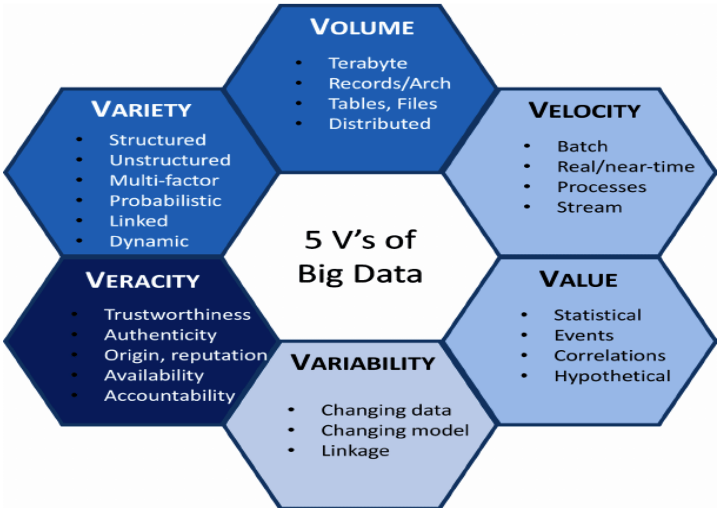


Fig 1: Bigdata Characteristics

They are top 10 security problems identified by OWASP

- **Insecure Web Interface** which can allow a bushwhacker to exploit an administration web interface (through cross-site scripting, cross-site request phony and SQL injection) and gain unauthorized access to control the IoT device.
- **Insufficient Authentication/ Authorization** can allow a bushwhacker to exploit a bad word policy, break weak watchwords and access to privileged modes on the IoT device.
- **Insecure Network Services** which can lead to a bushwhacker exploiting gratuitous or weak services running on the device, or use those services as a jumping point to attack other bias on the IoT network.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 11, November 2023, pp : 334-338

- **Lack of Transport Encryption** allowing a bushwhacker to listen in data in conveyance between IoT bias and support systems.
- **sequestration** raised from the fact the most IoT bias and support systems collect particular data from druggies and fail to cover that data.
- **Insecure pall Interface** without proper security controls a bushwhacker can use multiple attack vectors (inadequate authentication, lack of transport encryption, account recitation) to pierce data or controls via the pall website.
- **Insecure Mobile Interface** without proper security controls a bushwhacker can use multiple attack vectors (inadequate authentication, lack of transport encryption, account recitation) to pierce data or controls via the mobile interface.
- **Insufficient Security** Configurability due to the lack or poor configuration mechanisms a bushwhacker can pierce data or controls on the device.
- **Insecure Software/ Firmware** bushwhackers can take advantage of unencrypted and unauthenticated connections to commandeer IoT bias updates, and perform vicious update that can compromise the device, a network of bias and the data they hold.
- **Poor Physical Security** if the IoT device is physically accessible than a bushwhacker can use USB anchorages, SD cards or other storehouse means to pierce the device zilch's and potentially any data stored on the device

2. BIG DATA PRIVACY

Privacy of big data faces many issues which are classified into 4 categories including Integrity Security, Data Administration, Data Privacy, and Data Security [1].

- **Framework Security**: The technology of Big Data follows the infrastructure of distributed computing and various users work in parallel in it. It implies that the identification of intruders is very important. At present, most of the institutions have transferred to NoSQL databases from the traditional ones to handle semi structured and unstructured data. NoSQL appears to offer architecture flexibility for the data that is multi sourced but it is vulnerable to attacks.
- **Data Privacy**: Various sources are used for collecting the data privacy has to be maintained in the analytic stage. Encryption techniques can be utilized for protecting data.
- **Data Administration**: BD or big data is collected from countless sources making it contain numerous end users. Gradually, complexity in big data increases. In big data, complexity will be concerned with provenance metadata because of the provenance graph.
- **Integrity Security**: Filtering process and input validation pose a significant challenge to the application of big data. Due to the data size, it is quite tough to determine whether the data is derived from a valid source or not. If the source is legit then the data has to be eliminated so that it doesn't possess a risk to the whole system.
- **Anonymization**: A common technique that is used by firms and organizations for the protection of data is data anonymization. It helps in protecting data across distributed and cloud systems. A number of solutions and models are utilized for the implementation of this technique including l-diversity, k-anonymity, m-invariance, and t-closeness. The sub-techniques are based on Bottom-Up Generalization and Top-Down Specialization
- **Data Cryptography**: Another common technique that is used for the protection of data is data encryption. It is utilized for ensuring the confidentiality of Big Data. In contrast with typical techniques for encryption, it should be noted that Homographic Cryptography allows computation even on the data that is encrypted.

3. RELATED WORK

This paper provides about [1] Privacy and security in the context of Big Data present a critical challenge in today's data-driven world. The rapid growth of data generation and sharing across various platforms and industries has amplified the risk of unauthorized data access, breaches, and privacy infringements.

The problem statement revolves around the need to develop robust security measures and privacy-preserving techniques tailored to the unique requirements of Big Data. It is imperative to address issues related to Big data Analytics such as data encryption, access control, anomaly detection, and compliance with evolving data protection regulations. Ensuring privacy involves preserving individual rights and data confidentiality, while security encompasses protection against unauthorized access, data breaches, and cyber threats. This involves striking a delicate balance between privacy, security, and data utility, which remains a critical concern for researchers and practitioners alike.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 11, November 2023, pp : 334-338

4. METHODOLOGY

To address the complex challenges of privacy and security in the realm of Big Data, a multifaceted methodology is essential. This methodology comprises several key components:

1. **Literature Review**: According to literature review involves an extensive serve to identify the evolving trends, emerging threats, and existing solutions in the domain of privacy and security in Big Data. It provides a foundation for understanding the current state of knowledge in this field.
2. **Data Collection**: Gathering relevant data is large amount to assess the actual security and privacy concerns that organizations face in Big Data environments. This data collection phase includes surveys, interviews, and data breach incident reports. Real-world data is crucial for assessing the practical implications of security and privacy issues.
3. **Data Analysis**: The collected data is then subjected to rigorous analysis. Quantitative and qualitative techniques are employed to identify patterns, vulnerabilities, and the impact of security breaches on Big Data systems. This analysis helps in understanding the severity and prevalence of different threats.
4. **Case Studies**: A significant component of our methodology involves the examination of real-world case studies. These studies offer insights into specific incidents, strategies employed by organizations to mitigate risks, and lessons learned.
5. **Continuous Improvement**: Our methodology acknowledges the dynamic nature of the security and privacy landscape in Big Data. Therefore, continuous improvement and adaptation are fundamental. Ongoing research and collaboration with industry experts and stakeholders ensure that the framework remains relevant and effective in the face of evolving threats.

Big Data Analytics Tools

- **Hadoop** - helps in storing and analysing data
- **MongoDB** - used on datasets that change frequently
- **Talend** - used for data integration and management for synchronize big data
- **Cassandra** - a distributed database used to handle chunks of data
- **Spark** - used for real-time processing and analysing large amounts of data
- **STORM** - an open-source real-time computational system
- **Kafka** - a distributed streaming platform that is used for fault-tolerant storage

PROPOSED WORK /METHODS

There are many ways to deal with these challenges that come up with big data privacy and security. [3] Companies often use different techniques like De-identification, Encryption etc. to deal with them so that no-one could extract sensitive information from the datasets.

De-identification is a traditional technique which is used to prevent someone's personal identity from being revealed [1]. This process is adopted as one of the main approaches towards data privacy protection and it is commonly used in many fields of multimedia, communication, biometrics, big data etc.

Encryption: The fundamental idea behind the concept of encryption is that whenever we transfer some encrypted information then our computer converts that data into a cipher text i.e., the result of simple text after encryption or simply defined as an unreadable output of an encryption algorithm which can only be put back into a readable form after decoding it. This concept is actually in use since long time even before computer was invented to send secret messages.

Applications of using Encryption:

Encryption Protects Privacy: Encryption is a technique for safeguarding confidential data, such as personal information. This helps to ensure anonymity and privacy and also reducing opportunities for surveillance by criminals. It also ensures the security of communication between client apps and servers.

Data is quite unsafe when travelling from one location to another i.e., the path through which data travels and Encryption works during this time only, ensuring no matter where data has been kept or how it is used.

TECHNIQUES

- Application Software Security
- Maintenance, Monitoring, and Analysis of Audit Logs
- Secure Configurations for Hardware and Software
- Account Monitoring and Control

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

www.ijprems.com
editor@ijprems.com

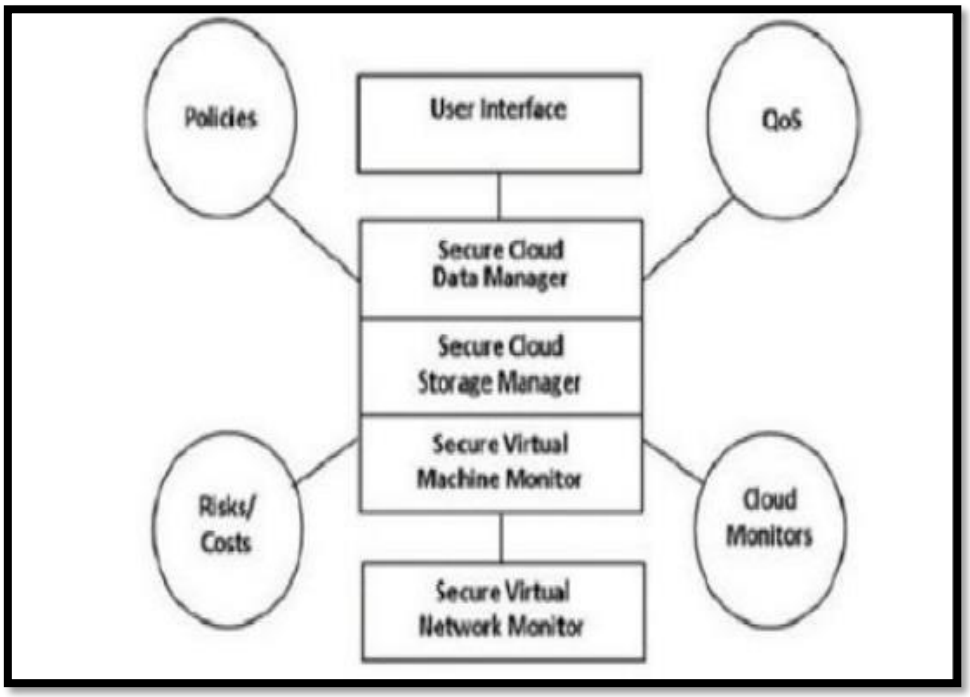
Vol. 03, Issue 11, November 2023, pp : 334-338

Impact Factor : 5.725

Application software security is a paramount concern in today's digital landscape, encompassing a range of crucial features to safeguard sensitive information and maintain the integrity of systems.[1] Key elements include the maintenance, monitoring, and analysis of audit logs. This ensures that every action within an application is tracked, allowing for swift detection and response to any suspicious activities or breaches. Additionally, securing configurations for both hardware and software plays a pivotal role in preventing vulnerabilities and unauthorized access. This involves rigorous adherence to best practices in configuring systems, networks, and applications to minimize security risks. Furthermore, account monitoring and control are essential features in the protection of application software. It involves continuously overseeing user accounts, permissions, and privileges, which helps in promptly identifying any anomalies or unauthorized account activity

The current state in [3] Big Data includes designing network topology, distributed algorithms, integration of software defined networks (SDN), scheduling, optimization and load balancing among different commodity computer. Data-driven information security dates back to bank fraud detection and anomaly-based intrusion detection systems. Fraud detection is oneof the most visible uses for big data analytics. However, the custom-built infrastructure to mine Big Data for fraud detection was not economical to adapt for other fraud detection uses Privacy-preserving techniques, including privacy-preserving aggregation, operations over encrypted data, and de-identification techniques. Following security measures should be taken to ensure the security in a cloud environment: File Encryption, Network Encryption, Logging, Software Format and Node Maintenance, Nodes Authentication, Rigorous System Testing of Map Reduce Jobs, Honeypot Nodes, Layered Framework for Assuring Cloud, Third Party Secure Data Publication to Cloud, Access Control.

Propelled Encryption Standard (AES) is most well-known calculation that bolster square figure, henceforth it is appropriate for HDFS pieces. AES accessible with 128- piece AES, 192- piece AES and 256-piece AES, 128-piece AES is utilized the greater part of times in light of its straightforwardness. There are distinctive methods of operations of AES: ECB, OFB, CTR, XTS and CBC. It is accounted for that AES: ECB is great decision of encryption or unscrambling calculation since its simultaneously played out a calculation in a disseminated domain.



5. CONCLUSION

Overall, it can be said that while big data offers some significant benefits to firms, it also poses some significant issues and challenges. One of these challenges is concerned with the sheer number of privacy risks that are experienced by it. The use of big data is quite risky in the sense that the identity of consumers can be exposed, and it can eliminate all the possible trust they have in a specific brand or a firm. For the protection of big data, there are generally a number of steps that are required to be taken by firms.The study of various methodologies by many researchers are making the data secured and provide privacy which made clear about the various methods, its merits and demerits and inabilities for providing security and privacy in Big Data. With this, we can come to conclude that we required some new technologies or the considerable modifications in the available technology.

6. FUTURE SCOPE

The following are some of the future scope which I have found while referring these papers. To reinforce big data security- focus on software protection, in location of tool safety. Isolate gadgets and servers containing important facts. Introduce real-time security data and event control. Provide reactive and proactive protectionIntegrating with a trust infrastructure (security of MapReduce). There are several domains of trust that must be made explicit and verified for MapReduce framework. Processing on encrypted data (security and privacy of MapReduce).

7. REFERENCES

[1] Gupta, A., & Jain, S. (2018). "Privacy and Security Challenges in Big Data: A Review." Journal of Big Data, 5(1), 29. This paper provides a comprehensive overview of privacy and security challenges in the Big Data

[2] The field of privacy and security in Big Data has garnered significant attention in recent years. "Privacy-Preserving Data Mining: A Survey" by V. S. Verykios et al. (2019) is an early work that laid the foundation for privacy-preserving techniques in Big Data analytics.

[3] Jayesh surana, akshaykhandelwal, avanikothari, himanshisolanki, meenalsankhla, big data privacy methods 2017 ijedr, volume 5, issue 2.

[4] Sharma, S., & Chen, Y. (2020). "Security and Privacy in Big Data: A Comprehensive Review." Journal of Ambient Intelligence and Humanized Computing.

[5] W. El-Hajj, “The most recent SSL security attacks: origins, implementation, evaluation, and suggested countermeasures,” Security and Communication Networks, vol. 5, no. 1, pp. 113-124, 2012.

[6] F. L. Greitzer, A. P. Moore, D. M. Cappelli, D. H. Andrews, L. A. Carroll and T. D. Hull, “Combating the insider cyber threat,” IEEE Security & Privacy, vol. 6, no. 1, pp. 61-64, 2008.

[7] M. D. Viji, K. Saravanan and D. Hemavathi, “A Journey on Privacy protection strategies in big data,” IEEE, pp. 1344-1347, 2017.