Importance of Security in Big Data Log Files on Cloud

Madan Mohan, Aadarsh Malviya, Anuranjan Mishra

Abstract: ***Today cloud computing is a very popular technology, and many people use this technology in many ways. it's important to have it safe. This technology was primarily used to keep data safer and safer in the cloud, so in this article we suggest a security framework for large data logs in the cloud. There are many and many risks that threaten the integrity of this information in the great information. Therefore, in line with the development of technology, the level of security has also increased significantly over the years. Various technology techniques access several online activities, such as interaction with different internet sites and services, making the web more accessible to their plug-ins. As a result, these activities have created a global platform for malicious activities to add these devices that expose large data logs harmful attacks. Sky system is an online platform that requires proper security integration. In addition, the current state of online security threatens high data in the cloud, which has affected the performance and service model.***

Keywords: ***Log File, Security, Big data, Dataset, Data analysis, Malicious, Technology, Security, Big data, Dataset, Cloud.***

I. INTRODUCTION

Cloud computing is increasingly gaining momentum in the process and storage of big data. Especially, in a world where security and privacy are magnified by the diversity and scale of data being process and stored. The uses of traditional security mechanism have become obsolete in the modern technological era [1]. As such, new security frameworks are being implemented to ensure the safety to the high-volume data. Therefore, this essay will focus on the security frameworks being implemented to enhance security and privacy of big data log files on the cloud as well as addresses the major challenges associated with these frameworks in cloud computing. Big data is a term used to refer to the study and applications of data sets that are so big and complex that traditional data-processing application software are inadequate to deal with them. Big data challenges include capturing data, data storage, data analysis, sharing, transfer, visualization, querying, updating, information privacy and data source.

Manuscript received on September 10, 2021.
Revised Manuscript received on September 16, 2021.
Manuscript published on October 30, 2021.
***Correspondence Author**

Madan Mohan*, Ph.D. Scholar, Department of Computer Science and Engineering, Noida International University, Greater Noida (U.P), India. Email: mmphdcse@gmail.com

Aadarsh Malviya, Professor & Dean, GNIOT, Greater Noida (U.P), India. Email: amc290@gmail.com

Anuranjan Mishra, Assistant Professor, Department of Computer Science and Engineering, Noida International University, Greater Noida (U.P), India. Email: aadarsh.malviya@niu.edu.in

There are a number of concepts associated with big data: originally there were 3 concepts volume, variety, velocity. instrument that provides small scale, static and semi-isolated security. Logfiles are very important data in the cloud so these logfiles should be secure. In securing big data, security frameworks must be put in places such as logging, encryption, and honeypot to provide data protection and privacy.

II. HADOOP

Apache Hadoop is a collection of open-source software utilities that facilitate using a network of many computers to solve problems involving massive amounts of data and computation. It provides a software framework for distributed storage and processing of big data using the MapReduce programming model. It can run on thousands of terabytes of systems involving thousands of nodes. The distributed file system in Hadoop helps to achieve fast data transfer rates, and the system continues to function even in the event of some node failure. This approach reduces the risk of a total system failure, even in the case of a large number of node lack of success. Hadoop Make the calculation resolution can expand, economy, flexible and fault tolerant. In these days so many companies are using Hadoop Framework to support applications that involve big amounts of data. Hadoop has two main subprojects Map Reduce and Hadoop Distributed File System.

III. META CLOUD DATA STORAGE SECURITY FRAMEWORK

This security framework is used to protected big data from any form of intrusion. The Meta Cloud MC security framework provides various scalable algorithms security solutions to any data deployed in the cloud system [1]. The Meta Cloud system works by forwarding data in the cloud to a Grouping and Choosing architecture GC for security enhancement. This security framework organizes stored data sent in cloud n multiple center-based systems. These systems are categorized into three basic level of security that is sensitive, critical and normal. Whereby, each level redirects data log files to the appropriate data center in the cloud for the safety measures. The Meta Cloud security framework provides a unique storage path that is impossible decrypt.

IV. MAPREDUCE SECURITY FRAMEWORK

This form of security encryption of big data log file provides authorization and authentication only to valid account users. It uses the HDFS authorization mechanism to protect and secure data files on the cloud computing systems [4].

The Map Reduce security system splits data log files into several chunks that can be computed parallel to enhance security measures. Furthermore, this security framework performs two core security measures. That is, they provide securing to mappers input in the cloud system and secure data that are present in any untrusted mapper. Additionally, the MapReduce security framework help contains intentional and unintentional data log file leakage. The system using a Hadoop helps in computing of big data that might undermine the level of security of private record and user privacy.

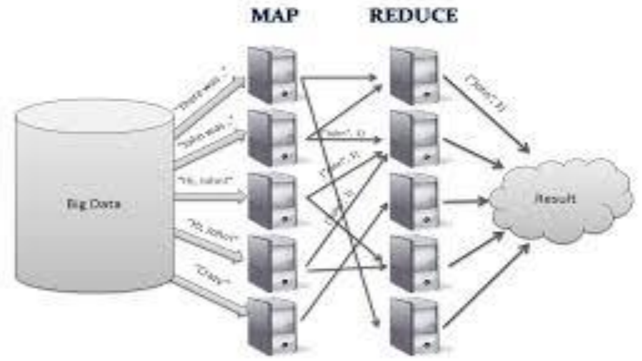


Fig. 1. The Aws CloudTrail Security Framework

Big data log files that are remitted to the cloud system especially those with rich information from high-level sources require an AWS Cloud Trail security system. This security framework provides the data log file with an Athena SQL engine power that rapidly responds to any form of security breach. Furthermore, the AWS Cloud Trail security framework record activities and publish them to the log files in the Amazon S3 where they can be tracked, and any insecurity activity obtained in the process [3]. The Amazon S3 help the AWS Cloud Trial system identify when an activity was performed, the resource affected and many significant details in case of data in a security breach. Therefore, any organization that uses the AWS Cloud Trail security framework systems can account for significant safety and privacy inconsistency in their cloud system. The AWS Cloud Trail is one of the most secure and flexible security frameworks. It helps protect the privacy of customer data hence recommended for most organizations.

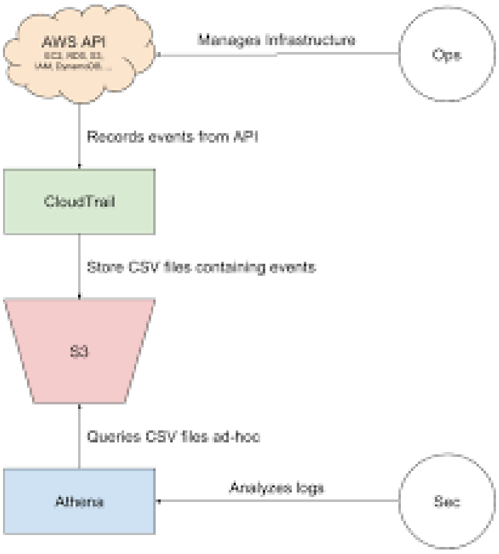
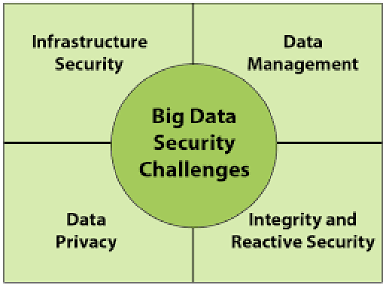


Fig. 2. Challenges of Securing Big Data

The primary challenge of big data is storage since the distributed storage is usually not enough given the rate at which data is collected. The storage systems are typically unable to store these big data. Therefore, there is a need for creating a storage system that is faster and has a capacity of processing big data. Additionally, scaling capacity must be taken into consideration for cloud storage of big data. The storage capacity and performance are a big challenge with clouds, and it influences others not to use clouds for storing and processing big data. In additions, big data require that all data must be collected, and the process of extracting information requires times and integrity. The privacy and security issues are some of the challenges that come with storing big data in clouds. Cloud computing gives a path to accessing stored data for analysis, and in during this analysis, private information may be shared that interfere with the data. Furthermore, obtaining personal information on clouds or IT architectures on cloud storages poses risks to the stored data [2]. “Parallel. Attackers often target cloud-computing system with new tactics, manage to access private information, and use the information for wrong reasons. Most cloud providers have failed to use proper network and architecture that will protect the privacy of the data being stored. Cloud providers must ensure they have the appropriate and secure infrastructure to operate the big data engine that receives and process data all the time. The traditional infrastructure used by most cloud providers in managing big data is slow and hard manage. Therefore, additional infrastructure must be added to speed up the management of big data. Proper framework and infrastructure will help in analyzing a large amount of data, which is always difficult to analyze [4]. Storing big data to clouds have experienced analytical and technical issues that limit its performance of cloud computing. Analyzing big data that comes in different forms require advanced skills given the sensitivity of the process. Technical problems such system fails and small processing system can lead loss of data or misinterpretation of data. Some clouds providers often have shared hardware resources, which increase workloads hence affecting the overall performance of the system. Storing big data in clouds can lead to data errors, which may be collected from shared resources, or data sets from the internet. It is necessary to limit sources of data to minimize data errors that may arise.



V. RESULT

Various technology appliances are gaining access to numerous online activities that means they are becoming more reliant on the internet and their plug-ins while interacting with different online sites and services.

Consequently, these activities have created a global platform for malicious activities to attach these appliances that expose big data log file to malicious attacks. The cloud system is an online platform that needs proper security integration. Moreover, with the current state of online insecurity, big data files on the cloud are compromised, and this has affected the pattern of performance and service.

In recent days there has been a rise in cases of reported surveillance and security gap which has compromised users' privacy in doubt of the current model where the third parties collect and control huge chunks of data which is personal. Although Big data does not necessarily represent the actual amount of data or size it cannot be processed via databases [5]. Due to this, it comes along with many privacy data issues which make it a major critical concern to any organization in our current companies and organizations of our world today. Therefore, these private issues must be addressed with greater and immediate effect to avoid more hazardous effects which may affect the companies or organizations either directly or indirectly. This issues also affect the performances of the company and its success may be limited to increase. This can lead to collapsing of companies and organizations. In order to avoid big losses which may lead to collapsing of companies and organizations, the companies are therefore putting across measures in order to curb any exploitation which may lead to huge losses.

VI. CONCLUSION

In summary, the cybercrime of today has increased causes a high degree of uncertainty and violation of cloud data privacy. Logfiles are very important data in the cloud so these logfiles should be secure. Therefore, guarantees must be guaranteed so that online information is safe and well protected against any risk. As such, the application of several security frameworks, such as the AWS Cloud example, plays an important role in offering security to data records in the cloud.

REFERENCES

1. Appelbaum, D. “Securing Big Data Provenance for Auditors: The Big Data Provenance Black Box as Reliable Evidence.” *Journal of Emerging Technologies in Accounting*, 2016, *13*[1], 17-36. doi:10.2308/jeta-51473
2. Dobre, C., & Xhafa, F. “Parallel Programming Paradigms and Frameworks in Big Data Era.” *International Journal of Parallel Programming*, 2014, *42*[5], 710-738. doi:10.1007/s10766-013-0272-7
3. Rodríguez-Mazahua, L., Rodríguez-Enríquez, C., Sánchez-Cervantes, J., Cervantes, J., García-Alcaraz, J., & Alor-Hernández, G. “A general perspective of Big Data: applications, tools, challenges, and trends.” *Journal of Supercomputing*,
4. 2016, *72*[8], 3073-3113. doi:10.1007/s11227-015-1501-1
5. Zhao, J., Tao, J., & Streit, A. “Enabling collaborative MapReduce on the Cloud with a single-sign-on mechanism.” *Computing*, 2016, *98*[1/2], 55-72. doi:10.1007/s00607-014-0390-0
6. A, Katal, Wazid M, and Goudar R.H. "Big data: Issues, challenges, tools and Good practices.". Noida:

a. 2013, pp. 404 – 409, 8-10 Aug. 2013.

7. Lu, Huang, Ting-tin Hu, and Hai-shan Chen. "Research on Hadoop Cloud Computing Model and its
8. Applications.". Hangzhou, China: 2012, pp. 59 – 63, 21-24 Oct. 2012.
9. Wie, Jiang , Ravi V.T, and Agrawal G. "A Map-Reduce System with an Alternate API for Multi-core
10. Environments.". Melbourne, VIC: 2010, pp. 84-93, 17-20 May. 2010
11. K, Chitharanjan, and Kala Karun A. "A review on Hadoop — HDFS infrastructure extensions.". JeJu Island: 2013, pp. 132-137, 11-12 Apr. 2013.
12. F.C.P, Muhtaroglu, Demir S, Obali M, and Girgin C. "Business model canvas perspective on big data applications." *Big Data, 2013 IEEE International Conference*, Silicon Valley, CA, Oct 6-9, 2013, pp.32 - 37.
13. Zhao, Yaxiong , and Jie Wu. "Dache: A data aware caching for big-data applications using the
14. MapReduce framework." *INFOCOM, 2013 Proceedings IEEE*, Turin, Apr 14-19, 2013, pp. 35 - 39.
15. Xu-bin, LI , JIANG Wen-rui, JIANG Yi, ZOU Quan "Hadoop Applications in Bioinformatics." *Open*
16. *Cirrus Summit (OCS), 2012 Seventh*, Beijing, Jun 19-20, 2012, pp. 48 - 52.
17. Bertino, Elisa, Silvana Castano, Elena Ferrari, and Marco Mesiti. "Specifying and enforcing access control policies for XML document sources." pp 139-151.
18. E , Bertino, Carminati B, Ferrari E, Gupta A , and Thuraisingham B. "Selective and Authentic Third- Party Distribution of XML Documents."2004, pp. 1263 - 1278.
19. Kilzer, Ann, Emmett Witchel, Indrajit Roy, Vitaly Shmatikov, and Srinath T.V. Setty. "Airavat: Security and Privacy for MapReduce."
20. "Securing Big Data: Security Recommendations for Hadoop and NoSQL Environments."*Securosis blog*, version 1.0 (2012)

AUTHOR PROFILE

Dr. Anuranjan Misra is Chairman of Computer society of India, Ghaziabad Chapter. He is Head MSME Business Incubation-GNIOT Centre(an Initiative of Ministry of MSME, Govt. of India), Head MSME Design Centre-GNIOT Center(an Initiative of Ministry of MSME, Govt. of India), President Institution's Innovation Council -GNIOT Centre(an Initiative of Ministry of Education, Govt. of India), Chair Unnat Bharat Abhiyan- GNIOT (an Initiative of Ministry of Education, Govt. of India), Chair Smart Campus Cloud Network- GNIOT (an Initiative of TERRI & AICTE , New Delhi) and Dean (Research, Innovation & Development) at Greater Noida Institute of Technology(GNIOT). He has 21 years of rich experience in academics, research and industry. He has delivered more than 25 expert talks around the world. He has more than 150 publications. He has handled research funding of 6+ crores. He has Senior Member of ACM, CSI, IACSIT, IACNG, IRACST, CSTA, ISOC, ICE, AEE, IFETS, ISMCDM, SIGSE. His research is in Big Data, Cloud Computing, Data Science, and Algorithms. He is passionate about quality of higher education in India.

Dr. Aadarsh Malviya is a research analyst who is currently holding the designation of Assistant Professor in the department of Computer Science, School of Sciences, Noida International University. He has more than ten years of working experience during which he has covered research, academics, and Industry. His research work includes cloud computing, expert systems, soft computing, image intensification, digital image processing and artificial intelligence. He submitted his research entitled "Towards Expert Systems for Enhancing Quality of Services in Cloud Computing" in 2017 for his Doctor of Philosophy. He has been Awarded as Young Researcher in Computer Science and Engineering from Uttar Pradesh, 2019 by GOREA. He has also covered the responsibilities of training and placement department and has formed a strong base in corporate and institute relationships. He is MTA (Networking fundamental) certified and has excellence in technical background. He has developed more than hundreds web application and has students across the world who are provided assistant in technical analysis. He aspire to develop himself such that he can take his nation to the best of his place.

Madan Mohan is a research scholar in Noida International University, Greater Noida, Uttar Pradesh, India. Having 20 years of experience in software development and teaching at various organizations. He is doing research on security framework for big data and cloud at NIU. Ph.D. Scholar (CSE), Noida International University, Greater Noida, (Up), India.

He has published the following papers:

- A comparison study about big data security and other technologies
- New challenges of big data security and its solutions
- Lack of privacy in big data and decentralized data in blockchain technology: going to be a big problem in future
- Security framework for big data log files on clouds
- Proposing a framework in big data to the hospitals for hypertension problem