RESEARCH ARTICLE OPEN ACCESS

Systematic Review of Recent Security Solutions in Hadoop and Cloud Environment

Anita V. Mithapalli, Prof. Swati S. Joshi
N.B.N. Sinhagad CoE, Solapur MS.
Asst. Professor N.B.N. Sinhagad CoE, Solapur MS.
Corresponding Author: Anita V. Mithapalli

ABSTRACT
The security issues for cloud computing, Huge data, Guide Diminish and Hadoop condition. Its key spotlight is on security issues in cloud computing that are connected with tremendous data. Cloud computing security is making at a snappy pace which joins PC security, network security, data security, and data protection.In this paper, we presented the systematic review of recent possible security solutions in Hadoop and cloud environment. The investigation and comparative analysis of recent studies related to the protection resolutions in Hadoop and cloud environment considered in this review. The outcome of this paper claims the various research gaps identified from the literature review.

Keywords - Cloud Computing, Big Data, Hadoop, Data protection.

Date Of Submission: 25-08-2019 Date Of Acceptance: 09-09-2019

I. INTRODUCTION

The Huge data is the term for data sets so colossal and trapped that it ends up difficult to process using customary data the executives instruments or preparing applications. The data and to distinguish designs it is fundamental to securely store, manage and share a great deal of complex data. Cloud goes with an express security challenge, for instance the data owner presumably won't have any direction over where the data is put. Apache's Hadoop appropriated record framework (HDFS) is creating as an unparalleled programming fragment for cloud computing united nearby consolidated parts, for instance, MapReduce. Hadoop, which is an open-source execution of Google MapReduce, including a scattered record framework, provides for the application programming engineer the impression of the guide and the decrease. With Hadoop, it is simpler for associations to take a few to get back some composure on the enormous volumes of data being produced every day, and yet can likewise make issues identified with security, data get to, checking, high accessibility and business congruity. Late advancement on exemplary enormous data networking advances, e.g., Hadoop and MapReduce, huge data innovations in could computing, huge data benchmarking undertakings, and portable huge data networking.

In Cloud Computing, "Cloud" signifies "The Internet", so Cloud Computing suggests a sort of computing wherein administrations are passed on through the Internet. The target of Cloud Computing is to utilize expanding computing capacity to execute a great many directions for consistently. Cloud Computing uses networks of a huge gathering of servers with specific relationship with pass on data getting ready among the servers. Cloud Computing involves a front end and back end. The front end client's PC and programming required to get to the cloud network. The back end contains various PCs, servers and database systems that make the cloud. The client can get to applications in the cloud network by partner with the cloud using the Web.

Hadoop is an open-source programming framework for securing and taking care of immense data in a dispersed way on tremendous clusters of ware hardware. Fundamentally, it accomplishes two assignments: huge data amassing and faster getting ready. Open-source programming: Open source programming differs from business programming as a result of the wide and open network of engineers that make and manage the undertakings. For the most part, it's permitted to download, use and add to, in any case an ever increasing number of business adaptations of Hadoop are getting to be accessible. Section II presents a review of recent methods. Section III presents the comparative study and research gaps. Section IV presents the Conclusion and future work

II. LITERATURE SURVEY

In this segment, we present the survey of related works upon two phases are security solutions in Hadoop and data security in the cloud environment.

Security Solutions in Hadoop:

In [6], Another security access control answer for Hadoop subject to CP-ABE, in has plan the CP-ABE use different attributes (gathering of properties) to perceive a client, as opposed to utilize a lone personality data, and hypothetical analysis demonstrated that our CP-ABE based arrangement can abstain from getting client complete data and upgraded security for client getting to document on Hadoop.

In [7], the productive confirmation protocol for Hadoop (Pile) lightened the serious issues of the current cutting edge validation mechanisms, specifically working system-based verification, secret word based methodology, and appointed token-based plans, individually, which are directly sent in Hadoop. Load pursued a two-server-based verification mechanism. Stack confirms the chief dependent on the digital mark age and check procedure using both propelled encryption standard and elliptic twist cryptography. The security analysis using both the formal security used the completely recognized real-or-random (ROR) model and the easygoing (non-numerical) security shows that Load ensures a few understood attacks. The formal security check utilized the generally utilized mechanized approval of Web security protocols and applications ensure that Stack is solid against replay and man-in-the-inside assaults.

In [8] the security dangers of the Hadoop environment, its security necessities, and structure contemplations. It additionally gave data stream charts. They expounded on the necessities and structure of RPC, HDFS, MapReduce, assignment token, square access token, these all are used for the security explanation behind the Hadoop framework.

In [9], Apache guard an open-source adventure by Cloudera is an approval module for Hadoop that offers the granular, work based approval required to give precise degrees of access to the right clients and applications. It support for occupation based approval, fine-grained approval, and multitenant organization. The Apache Knox Gateway is a framework that gives a singular reason for approval and access for various Hadoop benefits in a gathering. It offers an edge security response for Hadoop. The consequent favored position is it supports diverse affirmation and token check circumstances. It supervises security over different clusters and types of Hadoop. It furthermore gives SSO game plans and allows consolidating other character the board arrangements, for example, LDAP, Dynamic Index (Advertisement), what's more, SAML based SSO and other SSO systems [10].

In [11], the Hadoop security work, Undertaking Rhino offers a consolidated from beginning to end data security response for the Hadoop biological system. It gave a token-based affirmation and SSO plan. It offers Hadoop crypto codec framework and crypto codec utilization to give square level encryption to the data set away in Hadoop. It maintained key scattering and the board so MR can interpret data squares and execute the program as indicated by need. It also improved the security of HBase by offering cell level authentic

In [12], provided an HDFS square part segment that performs ARIA/AES encryption and decryption under the Hadoop coursed computing condition. In proposed variable-length data getting ready segments that performed encryption and decryption by including hoax data. It remained productive for different applications, for example, word tallying, sorting, k-Means, and progressive clustering.

In [13] presents a financially savvy strategy that anybody can use with their Hadoop bunch to given it three-dimensional security. The Kuber framework was not indicated to a specific encryption algorithm. It provided designers with the flexibility to use some other encryption calculation they approve of. All they need to was to give encryption and decryption techniques from their group documents as gave in the present execution of Kuber.

Data Security in Cloud Environment:

In [20] proposed another security and provenance model for data wrongdoing scene examination and post-appraisal in cloud computing. The framework proposed is addressed give the protection and security of secret files/archives that are piled up in the cloud.

In [21] outlined the one of a kind issues of security and protection challenges with the Cloud. They examined the blockades and responds in due order regarding giving a dependable Cloud computing condition. They explained the present security and protection plans, which must be essentially re-examined regarding their fittingness for Clouds.

In [22] proposed the structure of a framework that got the development and getting ready of the data kept on the Cloud. There is a necessity for security gotten gadgets on the Cloud, which would ensure clients that their data is checked and safe from security perils and strikes. The proposed use relied upon a relevant analysis and executed in a little Cloud computing condition.

In [23] has taken an elective point of view and proposed a data-driven perspective on Cloud security. The looks into investigated the security properties of secure data sharing among the applications encouraged on Clouds. They analyzed the data the board issues in appropriated question preparing, Legal what's more, framework analysis, and request correction affirmation. They proposed another security arrange for Cloud computing, which is named as Conclusive Secure Coursed Systems (DS2) [7]

In [8] The paper discussed out security risks in Cloud computing and enlightened advances that an undertaking can go out on a limb and ensure their resources. Cloud computing qualities, vulnerabilities, and significant territories in data chance administration. Moreover, they broke down the association security dangers, dangers, and useable countermeasures before adopting this technology

In [9] given a rundown of security dangers revealed in the chose writing that were connected with the Framework as an Administration (IaaS) Cloud computing administration commitments. This paper perceived security risks, which were balanced inside the Protection, Respectability, and Availability (CIA) security targets model. Eight perceived sorts of attacks are organized in association with three beginnings and mapped to the CIA model of security dangers.

In [14] Hash Message Authentication Code(HMAC) process, a riddle key and hash calculation, for instance, the secure hash algorithm (SHA) was used to make the message confirmation code. This check code steadily gave data decency and legitimacy in light of the fact that the mystery key was required to reproduce the code. HMAC can be helpful against man-in-the-middle ambushes on the message.

In [15], the proposed plan calls upon cryptography, expressly Public Key Foundation working together with SSO and LDAP, to ensure the affirmation, decency, and mystery of included data and interchanges. It displayed an even level of administration, available to each and every included substance, that comprehends a security work, inside which central trust was kept up.

In [16], centres around the specialized parts of digital crime scene investigation in circulated cloud situations. They contributed by surveying whether the client of cloud computing services can play out a conventional digital examination from a specialized perspective. Besides, they talked about potential arrangements and conceivable new strategies helping clients to perform such examinations.

In [17], for cloud security to dispense with the worries in regards to data misfortune, isolation, and privacy while getting to the web application on the cloud. Calculations like: RSA, DES, AES, Blowfish have been used and comparative examination among them have besides been shown to ensure the security of data on the cloud. DES, AES, Blowfish is symmetric key calculations, wherein a single key is used for both encryption/decryption of messages.

In [18], focused on the security and insurance issues in cloud computing in light of a quality-driven technique, and the data uprightness affirmation made dealing with encryption algorithm and the audit changed into executed with the assistance of hashing algorithm open in order to affirm the value delivered again even as checking the data decency available with the related record, here they have tackled different angles, for instance, customer record access approach, availability of data, data changing or dependability affirmation and the system ought to be security protecting so the data should not be break in the midst of the cloud execution.In [19], depicted that the overwhelming issue of cloud stockpiling service is to prompt clients to accept the cloud stockpiling supplier security and to transfer their touchy individual data. They proposed, cloud client driven key administration system for the assurance of data in the cloud, to take care of the issue of clients trust With the knowledge presented in the above studies, one may ask what combination phenomena would arise for Hadoop and cloud environment. To answer this question, in this work, we present a systemic investigation of possible security solutions.

II. COMPARATIVE ANALYSIS

Paper Title	**Year**	**Techniques**	**Highlights**	**References**
Tending to cloud computing security issues. Future Age Computer Frameworks	2012	Cloud computing security Confided in Outsider Public key framework Data and correspondence security Trust	This paper endeavors to assess cloud computing security. An answer is exhibited which endeavors to kill interesting threats. This paper presents a Confided in Outsider. TTP is entrusted with	[15]

			guaranteeing security qualities inside a cloud situation. The arrangement utilizes Open Key Foundation working together with SSO and LDAP	
Secure Client Data in Cloud Computing Utilizing Encryption Calculations	2013	AES, Blowfish, DES, Cloud Computing, RSA, Data Security	AES algorithm used the least time to execute cloud data, the Blowfish calculation has the least memory need. DES calculation eats up least encryption time. RSA consumes the longest memory size and encryption time.	[17]
Security and Protection in Cloud Computing	2013	cloud computing, security, protection, trust, classification, honesty, responsibility, accessibility	The associations among them, the vulnerabilities that may be abused by aggressors, the risk models, just as protection methodologies in a cloud situation	[18]
Development of an Archive The executives Framework for Private Cloud Condition	2013	Document Management, Collaboration, Synchronization, Private Cloud	This system provided clients in an association with a basic and productive mechanism to get to, oversee and share their data. It gave major report controls, synchronization, and sharing functionalities, and thinks about the help of heterogeneous customer devices	[19]
A new solution of data security accessing for Hadoop based on CP-ABE	2014	Encryption, Access Control, Public Key, Servers, Cloud Stockpiling, Data Security Access, Hadoop	ABE based solution can avoid obtaining user complete information and upgraded security for a client getting to a document on Hadoop	[6]
Structure and execution of HDFS data encryption plan utilizing ARIA calculation on Hadoop	2017	Hadoop Security, HDFS Data Encryption, Hadoop Encryption Codec, ARIA Encryption Calculation, Data Encryption,	Proposed variable-length data components and perform encryption and decryption including sham data. Furthermore, prevent data intruders from stealing user data	[12]
HEAP: A Proficient and Flaw tolerant Validation and Key Trade Convention for Hadoop-helped Huge Information Stage.	2018	Cloud computing, formal security, key understanding, huge information security, Hadoop, confirmation, formal security, AVISPA.	The chairman needs to login into the framework utilizing his character and secret word as it were. Therefore, the cluster enlistment arrangement is adaptable and easy to use in nature. Two-server based check and key exchange show does not have any likeness impediment	[7]

			with the standard single-server based methodologies The key rollover issue is a dire issue for customary client enrolment strategy, where during session key establishment, a whole deal riddle key mystery key) is used to create a puzzle channel among client and Enlistment Specialist (RA). Double server based session key foundation procedure makes the proposed verification framework increasingly strong, savvy and easy to use.	

Research Gaps

From the above literature review, we noticed some research gaps in order to design and study security solutions in Hadoop and cloud environment. As per the progress of research in this domain, we listed the research problems.

- The encryption asset supplier needs to get all applicable data of the client, it will harm the client's security unquestionably, and it will require more transfer speed and huge handling overhead.
- By information, encryption to secure the spillage of touchy information put away in Hadoop.
- The security and prosperity issues are the critical troubles before the cloud computing development. The well-regulated security and prosperity issue gives better organizations and creates trust in clients to move their data into the cloud.
- A network-level course of action with network protocols and network security, for instance, appropriated nodes, dispersed information, Internodes correspondence.
- User confirmation level arrangements with encryption/unscrambling techniques, validation strategies, for example, managerial rights for nodes, check of employments and nodes, and logging.

IV. CONCLUSION AND FUTURE WORK

In this paper, we displayed a precise audit of different studies over the Hadoop and cloud environment. The literature related to recent security solutions in Hadoop and cloud condition exhibited. In this paper, we contemplated the similar examination and analysis of the data security are provided for these challenges to overcome the risk involved in cloud computing. We showed the assessment and relative analysis recently examinations. The outcome of this paper claims the various research gaps identified from the literature review.

REFERENCES

[1]. K, Chitharanjan, and Kala Karun A. "A review on hadoop — HDFS infrastructure extensions.". JeJu Island: 2013, pp. 132-137, 11-12 Apr. 2013.

[2]. 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.

[3]. Zhao, Yaxiong , and Jie Wu. "Dache: A data aware caching for big-data applications using the MapReduce framework." INFOCOM, 2013 Proceedings IEEE, Turin, Apr 14-19, 2013, pp. 35 - 39.

[4]. Xu-bin, LI , JIANG Wen-rui, JIANG Yi, ZOU Quan "Hadoop Applications in Bioinformatics." Open Cirrus Summit (OCS), 2012 Seventh, Beijing, Jun 19-20, 2012, pp. 48 - 52.

[5]. Bertino, Elisa, Silvana Castano, Elena Ferrari, and Marco Mesiti. "Specifying and enforcing access control policies for XML document sources." pp 139-151.

[6]. Zhou, H., & Wen, Q. (2014). Another arrangement of data security getting to for Hadoop dependent on CP-ABE. 2014 IEEE

5th International Conference on Software Engineering and Service Science.

[7]. Chattaraj, D., Sarma, M., Das, A. K., Kumar, N., Rodrigues, J. J. P. C., & Park, Y. (2018). Store: A Proficient and Deficiency tolerant Verification and Key Trade Protocol for Hadoop-helped Huge Information Stage. IEEE Access, 1–1.

[8]. Owen O'Malley, Kan Zhang, Sanjay Radia, Ram Marti, and Christopher Harrell "Hadoop Security Design",Yahoo,2009.

[9]. Horton works "Technical Preview for Apache Knox Gateway", Hortonworks, November 2013.

[10]. M. Tim Jones "Hadoop Security and Sentry", IBM, January 2014.

[11]. Sudheesh Narayana, Packt Publishing "Securing Hadoop- Implement robust end-to-end security for your Hadoop ecosystem"

[12]. Youngho Song, Young-Sung Shin, Miyoung Jang, & Jae-Woo Chang. (2017). Structure and usage of HDFS information encryption plot using ARIA algorithm on Hadoop. 2017 IEEE International Conference on Big Data and Smart Computing (BigComp).

[13]. Parmar, R. R., Roy, S., Bhattacharyya, D., Bandyopadhyay, S. K., & Kim, T.-H. (2017). Large-Scale Encryption in the Hadoop Environment: Challenges and Solutions. IEEE Access, 5, 7156–7163.

[14]. Sobti, Rajeev, G. Geetha, Cryptographic hash functions: a review. International journal of computer science issues, vol.9 (2), pp. 461-479, 2012

[15]. Zissis, D., & Lekkas, D. (2012). Addressing cloud computing security issues. Future Generation Computer Systems, 28(3), 583–592. doi:10.1016/j.future.2010.12.006

[16]. Birk, D., & Wegener, C. (2011). Technical Issues of Forensic Investigations in Cloud Computing Environments. 2011 Sixth IEEE International Workshop on Systematic Approaches to Digital Forensic Engineering.

[17]. Rachna Arora, Anshu Parashar, "Secure User Data in Cloud Computing Using Encryption Algorithms" International Journal of Engineering Research and Applications (IJERA) Vol. 3, Issue 4, Jul-Aug 2013, pp.1922-1926

[18]. Xiao, Z., & Xiao, Y. (2013). Security and Privacy in Cloud Computing. IEEE Communications Surveys & Tutorials, 15(2), 843–859.

[19]. Kao, C. H., & Liu, S. T. (2013). Development of a Record The board System for Private Cloud Condition. Procedia - Social and Behavioral Sciences, 73, 424–429

[20]. Rongxing Lu. et.al (2010). "Secure Provenance: The Essential Bread and Butter of Data Forensics in Cloud Computing". ASIACCS '10 Proceedings of the 5th ACM Symposium on Information. Computer and Communications Security. pp. 282-292.

[21]. Hassan Takabi.et.al.(2010). "Security and Privacy Challenges in Cloud Computing Environments". IEEE security and privacy. www.computer.org/security. pp. 24 – 31.

[22]. R. La'Quata Sumter.(2010). "Cloud Computing: Security Risk Classification". ACMSE. Oxford. USA.pp.15-17

[23]. Wenchao Zhou.et.al.(2010)."Towards a Data-centric View of Cloud Security".CloudDB 2010.pp.1-8. 8. Anthony Bisong.et.al.(2010). "An overview of the security concerns in enterprise Cloud computing". International Journal of Network Security & its Applications. Vol.3. No.1. pp.30-31. 9. Paul Wooley.et.al.(2011). "Identifying Cloud Computing Security Risks" Capstone Report. University of Oregon Applied Information Management program. pp. 74.

Anita V. Mithapalli" Systematic Review of Recent Security Solutions in Hadoop and Cloud Environment" International Journal of Engineering Research and Applications (IJERA), Vol. 09, No.09, 2019, pp. 25-30