International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)

All-Inclusive Hadoop Framework Privacy Preservation Strategy in the Cloud Over Big Data.

Dr. A. Kanimozhi[1], Dr. A. Kamatchi[2]

[1]*Assistant Professor, Department of Computer Science, Sri Krishna Adithya College of Arts and Science, Coimbatore, India*
[2]*Assistant Professor, Department of Decision and Computing Science, Coimbatore Institute of Technology, Coimbatore, India*

***Abstract*-- In big data research, data security is one of the most crucial elements. Most cloud systems handle sensitive data in a sequential fashion, including personal, business, and health-related data. The emergence of cloud storage systems threatens this kind of information. Large data flows, however, are not protected by conventional security techniques. An efficient approach to data collection and protection that protects privacy. A Hadoop-based map-reduce mechanism is used to balance the various security protocols that are initially grouped with cloud-based large data. One of the most important aspects of big data research is data security. Applications for the cloud system handle private data, including corporate, medical, and personal records. The cloud systems that contain this kind of data may be at risk from threats. The definition of big data technology is software-utility. Information analysis, processing, and extraction from vast amounts of data and incredibly complicated structures are the main functions of this technology. Conventional data processing software finds it extremely challenging to handle this.**

Big data technologies are frequently linked to other technologies like deep learning, machine learning, artificial intelligence (AI), and the Internet of Things (IoT) that are greatly enhanced, making them one of the more expansive ideas of rage in technology. Big data technologies work in tandem with these technologies to analyze and manage vast volumes of batch-related and real-time data.

***Keywords*-- Big data, encryption, cloud storage, classification, clustering, optimization**

I. INTRODUCTION

A large amount of data that keeps growing exponentially over time is known as "big data." No standard data management technology can efficiently store or process this enormous and complex data set. Regular data and big data are comparable, but big data is substantially bigger. Big data is used by medical researchers and physicians to identify risk factors and disease indicators, as well as to diagnose patients' ailments and medical issues. Additionally, information on infectious disease outbreaks and dangers is updated for government agencies and healthcare organizations by combining data from the internet, social media, electronic health records, and other sources.

The combination of confidentiality, which forbids unauthorized disclosure of information, integrity, which forbids unauthorized alteration or deletion of information, availability, which forbids unauthorized withholding of information, and privacy, which refers to an individual's or group's capacity to selectively disclose information about themselves, constitutes data security.

Hadoop is an open-source software framework released under the Apache v2 license that makes data-intensive distributed applications easier to use. Simple programming models abstract and streamline the processing and storing of large and/or rapidly expanding data sets, including both structured and unstructured data. It uses commodity technology with low fault tolerance and redundancy to deliver high availability and scalability. Hadoop puts computation ahead of data. The bulk of the system's processing and storage capacity is located in the cluster's primary nodes.

In this study, the effectiveness of privacy-preserving big data security in the cloud using Hadoop has been Exclusive Enhanced in a number of ways. The first approach uses the IDEA framework with Hadoop and specific methods with the MapReduce programming paradigm to explore and realize privacy preserving knowledge discovery from massive data. The second approach is the Optimal Support Vector Machine. Using a categorization strategy for cloud privacy, the Hadoop architecture was used to minimize repetitive encrypting and decryption processes for public and hybrid files. There is a noticeable improvement in the categorization technique.

The effectiveness of cloud-based privacy-preserving large data security through The eIn this study, the effectiveness of privacy-preserving big data security in the cloud using Hadoop has been Exclusive Enhanced in a number of ways. The initial In this study, the effectiveness of privacy-preserving big data security in the cloud using Hadoop has been Exclusive Enhanced in a number of ways. The first approach uses the IDEA framework with Hadoop and specific methods with the MapReduce programming paradigm to explore and realize privacy preserving knowledge discovery from massive data.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)

The Optimal Support Vector Machine is the second strategy. To reduce the amount of time spent encrypting and decrypting public and hybrid data, the Hadoop architecture was utilized in conjunction with a cloud privacy classification technique. The categorization process has improved noticeably.

II. Related Works

Zhao et al. [11] have introduced a safe authentication mechanism for users across many servers based on passwords. For user authentication, Elliptic Curve Cryptography (ECC) is presented, which handles two security attacks namely Offline Password Guessing and Impersonation Attack. Experimentations were carried out to examine the performance in terms of efficiency and security. The proposed plan utilizes four stages, the First stage is the Initialization stage where the RC focuses on the initialization of the framework as indicated by a security parameter. Second is the Registration Phase.

In this stage, every system user Ui registers with the enrolment centre RC (the Registration Center) to get a smart card containing a secret key, as a qualification of Ui to demonstrate his/her realness to service provider (SP). Each SP Sj registers with RC to acquire a secret key, as a qualification of Sj to demonstrate its legitimateness to the system user. Next is the Authentication Phase. By running a confirmation strategy between a system user Ui and a service provider Sj, they can check the legitimacy of one another and set up a safe channel. That is, Sj guarantees that Ui is an enlisted client, and Ui accepts that the administration gave by Sj is lawful. The last one is the Password Update Phase, then a client Ui needs to refresh his/her unique secret key PWi. The suggested method is also suitable for real-time applications. ECC is an asymmetric encryption technique that creates smaller key pairs (public and private). This is one of the disadvantages, as well as the fact that the processing and communication costs are higher than with hashing approaches.

win et al. [12] have conducted a security analytics service that runs on a cloud of virtualized equipment and stores data in HDFS A two-step machine learning approach was described in this study, which included logistic regression (to calculate the conditional probabilities of assaults through characteristics) and belief propagation (to estimate the belief in the presence of an attack).The utilization of logistic regression empowers the quick count of assault's contingent probabilities. All the more significantly, calculated relapse likewise empowers the retraining of the individual logistic regression classifiers utilizing the new assault includes as they are obtained from assault identification.

III. Objectives

Enhancing large data security and privacy preservation in the cloud using the Hadoop framework is the aim of the research. Enhancing large data security and privacy preservation in the cloud using the Hadoop framework is the aim of the study.

The suggested techniques were created to improve transmission performance and reduce data uploading time.

Lower the error rate

Increase data processing time and storage capacity; prevent unnecessary transmissions; and lengthen the time needed for data classification.

High-level data throughput and Accuracy

Increase data processing time and storage capacity while preventing needless transmission. Increase the time it takes to classify data and achieve high levels of data accuracy and throughput.

Preventing Needless Transmissions

Increase the time it takes to classify data and achieve high levels of data accuracy and throughput.

1. Design And Development Of Improved Techniques For Preserving Privacy

IDEA Algorithm based on Map Reduce and weighted Auto Encoder KNN classifier for privacy preservation. Using the Incremental Density Based K-means Clustering (IDK-means) algorithm, this section describes the suggested methodology for clustering data collected from the cloud source. With the suggested approach, processing, encrypting, and decrypting take relatively little time. Privacy Preservation with Weighted Auto Encoder KNN classifier and Map Reduce based IDEA Algorithm. Using the Incremental Density Based K-means Clustering (IDK-means) algorithm, this section describes the suggested methodology for clustering data collected from the cloud source. With the suggested approach, processing, encrypting, and decrypting take relatively little time.

Optimal Support Vector Machine Privacy Preservation Using a weighted FCM algorithm based on density peaks instead of an upgraded word auto key encryption classifier, this classification technique preserves anonymity. The findings of this technique's examination show how useful and applicable the suggested integrated methodology is for protecting sensitive data while it is being transferred over several cloud nodes. The deep learning technique is used to classify the encrypted data. IDEA Algorithm based on Map Reduce and weighted Auto Encoder KNN classifier. Within this section, the suggested.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)

(A) K-Modes based Fast Mutation Artificial Bee Colony clustering methodology over an Advanced super encryption classifier is used in this intelligent classification method. The Advanced SET approach using Hadoop framework is discussed in the proposed method, which uses the KDD dataset to implement it. The census-income KDD data set was used to evaluate the proposed method of efficient convolution for privacy-preserving across enormous data employing map reduction idea in a cloud computing environment.

2. *Performance Assessment*

Below is a comparison of the outcomes of the suggested method with the current methodology. The following is a summary of the research work's findings based on the results

IV. Shortening The Time Needed To Upload Data

One of the primary functions of online data transmission from one system to another memory is data uploading. The suggested technique takes the least amount of time to upload the specified amount of data, as shown in Table 1 and Fig 1.

Table 1: Overall comparison of the suggested and current data uploading time methods

Performance Analysis	AES	Triple DES	RC2	Waek classifier with Map reduce	Enhanced Word key encryption	Advaned Super Encryption
Data Uploading Time	40 (sec)	35 (sec)	38 (sec)	11.8 (sec)	10.3 (sec)	9.6 (sec)
Encryption Time	39 (sec)	8.337 (sec)	9.4 (sec)	19 (sec)	20.1 (sec)	24.1 (sec)
Decryption Time	0.12 (sec)	0.23 (sec)	0.20 (sec)	17.4 (sec)	18.6 (sec)	23.7 (sec)

(A) Shortening The Time Needed To Upload Data

One of the primary functions of online data transmission from one system to another memory is data uploading. The suggested technique takes the least amount of time to upload the specified amount of data, as shown in Table 1 and Fig 1. The time delay for the advanced super encryption technique with K-Modes based Fast Mutation Artificial Bee Colony clustering is reduced to 9.6 seconds. In the current WAEK classifier with map reduce, the data uploading time is 11.8

(B) Enhancing Transmission Performance

This categorization technique evaluates the performance of big data security techniques and shows a significant improvement. Table 1 displays the performance analysis.

(C) Lower The Rate Of Errors

Data uploading is one of the main purposes of online data transmission from one system to another memory. According to Table 1 and Figure 1, the recommended method uploads the required quantity of data in the shortest amount of time. With K-Modes based Fast Mutation Artificial Bee Colony clustering, the time delay for the sophisticated super encryption approach is lowered to 9.6 seconds. The current WAEK classifier with map reduce takes 11.8 seconds to upload data.This suggested approach uploads data processing to the convolution process in order to lower the error rate and improve data accuracy. To increase data security and reduce error rates, a sophisticated super encryption technology is employed.

(D) Increase The Amount Of Time Spent Processing And Storing Data.

The data processing time for the current Exclusive Enhanced word auto key encryption classifier with density peak-based weighted FCM algorithm using the data is 2000 mb (434 secs). Data processing time is increased to 486 secs for the advanced super encryption technique with K-Modes based Fast Mutation Artificial Bee Colony clustering. Compared to the current

(E) Extend Data Processing Time And Storage Capacity

The current Exclusive Enhanced word auto key encryption classifier with density peak-based weighted FCM algorithm uses 2000 MB of data and takes 434 seconds to process. This method increases the data processing time to 486 seconds for the advanced super encryption technique with K-Modes based Fast Mutation Artificial Bee Colony clustering.

(F) Long Time For Data Classification

Table 3 shows that the suggested methodology's time consumption for data classification is higher than that of the current techniques, AES and Triple AES. The performance of data security and accuracy is significantly higher due to the display.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)

TABLE 2: OVERALL COMPARISON OF DATA PROCESSING TIME

File size (MB)	100	300	500	1000	2000
GMPLS/MPLS Networks	58	76	100	135	190
IDEA with hadoop framework	17.04	51.12	85.2	170.4	340.8
Enhanced WAKE	18.6	53	86	272	434
Advanced SET	21.7	61.4	93.3	311	486

(G) PREVENT NEEDLESS TRANSMISSIONS:

This classification method is distinguished by the needless transmission of the encrypted data via the convolution process and classification following a proper decryption of the data.

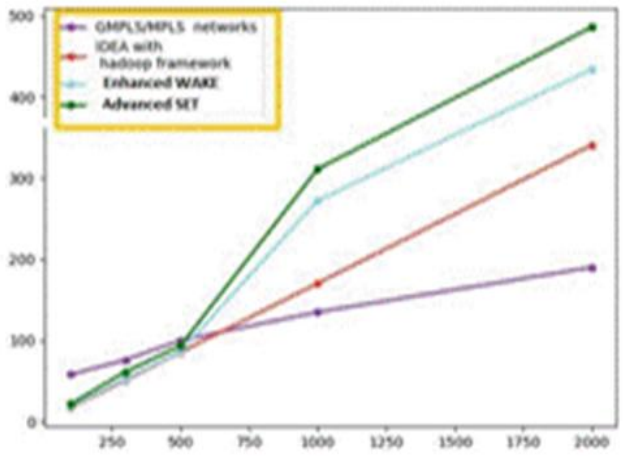

Fig 1 : Comparison of Data Processing Time

(H) Excellent throughput and data accuracy

The accuracy of the classification indicates how efficient the suggested categorization method is. In order to improve data accuracy and throughput, the encryption and decryption processes were carried out more effectively than other current methods. The accuracy of the results displayed in Table 4 is demonstrated by the suggested methodology for processing changes in data transmission volume to take the longest possible time.

Overall Accuracy, Precision, Recall

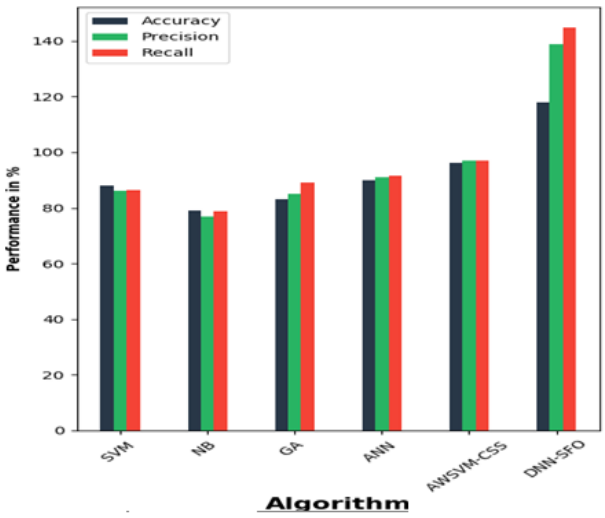

TABLE 3 :COMPARISON OF DATA CLASSIFICATION TIME

File size(MB)	*100*	*300*	*500*	*1000*	*2000*
GMPLS/MPLS Networks	*75*	*120*	*300*	*750*	*975*
IDEA	*8.77*	*26.3*	*43.85*	*87.7*	*175.4*
HDFS security	*9.4*	*28.2*	*47*	*94*	*190.5*
Exclusive Enhanced WAKE	*10*	*29*	*48*	*96*	*191*
Advanced SET	*12*	*33*	*57*	*136*	*241*

V. CONCLUSION AND PROSPECTIVE IMPROVEMENT

This study has offered effective methods for enhancing big data security and privacy preservation in a cloud setting by utilizing Hadoop. In order to increase throughput and use clustered data of map reduce algorithms to avoid time delays, this research has presented effective methods to improve the privacy preservation of big data security utilizing Hadoop in a cloud setting. The massive amount of data generated and its security implications required an efficient privacy-preserving solution.

The data will undergo another convolution procedure if the encryption is insufficient. The big data-based privacy-preserving methodology outperforms current methods when compared to the experimental results of the proposed framework. The suggested methodology encrypts and decrypts data created in the cloud. The suggested methodology processes, encrypts, and decodes data in a comparatively short period of time.

This work has suggested practical approaches for artificial intelligence, and innovations in machine learning may be incorporated into the strategy in the future. However, because these files are different from txt, csv, log, xls, and sql files, they need to be handled differently. Therefore, in order to handle file characteristics for such big data sets, additional methods will be needed.

REFERENCES

[1] Roy, S., Shovon, A. R., & Whaiduzzaman, M. (2017, December). Combined approach of tokenization and mining to secure and optimize big data in cloud storage. In 2017 IEEE region 10 Humanitarian Technology Conference (R1s0-HTC) (pp. 83-88). IEEE.

[2] Wu, J., Dong, M., Ota, K., Li, J., & Guan, Z. (2018). Big data analysis-based secure cluster management for optimized control plane in software-defined networks. IEEE Transactions on Network and Service Management, 15(1), 27-38.

[3] Ji, C., Li, Y., Qiu, W., Jin, Y., Xu, Y., Awada, U., .&Qu, W. (2012). Big data processing: Big challenges and opportunities. Journal of Interconnection Networks, 13(03n04), 1250009.

[4] Chaudhary, R., Aujla, G. S., Kumar, N., & Rodrigues, J. J. (2018). Optimized big data management across multi-cloud data centers: Software-defined-network-based analysis. IEEE Communications Magazine, 56(2), 118-126.

[5] Li, Z., Xu, W., Shi, H., Zhang, Y., & Yan, Y. (2021). Security and Privacy Risk Assessment of Energy Big Data in Cloud Environment. Computational Intelligence and Neuroscience, 2021.

[6] Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big Data and cloud computing:innovation opportunities and challenges. International Journal of Digital Earth, 10(1), 13-53.

[7] Gupta, B. B., Yamaguchi, S., & Agrawal, D. P. (2018). Advances in security and privacy of multimedia big data in mobile and cloud computing. Multimedia Tools and Applications, 77(7), 9203-9208.

[8] Zhang, L., Wu, C., Li, Z., Guo, C., Chen, M., & Lau, F. C. (2013). Moving big data to the cloud: An online cost-minimizing approach. IEEE Journal on Selected Areas in Communications, 31(12), 2710-2721.

[9] Subramanian, E. K., & Tamilselvan, L. (2020). Elliptic curve Diffie–Hellman cryptosystem in big data cloud security. Cluster Computing, 23(4), 3057-3067.

[10] Guo, L., &Qiu, J. (2018).Optimization technology in cloud manufacturing. The International Journal of Advanced Manufacturing Technology,97(1), 1181-1193.

[11] Zhao, Y., Li, S., & Jiang, L. (2018) Secure and Efficient User Authentication Scheme Base on Password and Smart Card Environment. Security and Communication Networks,2018,1–13. doi:10.1155/2018/9178941

[12] Win, T. Y., Tianfield, H., &Mair, Q. (2018). Big Data Based Security Analytics For Protecting Virtualized Infrastructures in Cloud Computing. IEEE Transactions on Big Data, 4(1),11–25. doi:10.1109/tbdata.2017.2715335

[13] Jun Wu, Kaoru Ota, Mianxion Dong, Jianhua Li, Hongkai Wang (2018),Big Data Analysis-based Security Situational Awareness for Smart Grid, IEEE Transactions on Big Data, Vol. 4, Issue.3, PP. 408.

[14] Lu, C. C., & Tseng, S. Y. (2002, July). Integrated design of AES (Advanced Encryption Standard) encrypter and decrypter. In Proceedings IEEE International Conference on Application-Specific Systems, Architectures, and Processors (pp. 277-285). IEEE.

[15] Del Rosal, E., & Kumar, S. (2017). A fast FPGA implementation for triple DES encryption scheme. Circuits and Systems, 8(09).

[16] AbdElminaam, D. S., Abdual-Kader, H. M., & Hadhoud, M. M. (2010). Evaluating The Performance of Symmetric Encryption Algorithms. Int. J. Netw. Secur., 10(3),216-22.

[17] Suthaharan, S. (2016). Support vector machine. In Machine learning models and algorithms for big data classification (pp. 207-235). Springer, Boston, MA.

[18] Jiang, L., Zhang, L., Li, C., & Wu, J. (2018). A correlation-based feature weighting filter for Naive Bayes. IEEE transactions on knowledge and data engineering, 31(2), 201-213.

[19] Mirjalili, S. (2019). Genetic algorithm. In evolutionary algorithms and neural networks (page No (43-55). Springer, Cham.

[20] Walczak, S. (2018). Artificial neural networks. In Encyclopedia of Information Science and Technology, Fourth Edition (pp. 120-131). IGI Global.