International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-10, August 2019

Ensure Security for Mapreduce-Hadoop Distributed File System using Encryption Method Over Big Data

Gunjan V. Keswani

Abstract*: Big data security is the most focused research issue nowadays due to their increased size and the complexity involved in handling of large volume of data. It is more difficult to ensure security on big data handling due to its characteristics 4V's. With the aim of ensuring security and flexible encryption computation on big data with reduced computation overhead in this work, framework with encryption (MRS) is presented with Hadoop Distributed file System (HDFS). Development of the MapReduce paradigm needs networked attached storage in addition to parallel processing. For storing as well as handling big data, HDFS are extensively utilized. This proposed method creates a framework for obtaining data from client and after that examining the received data, excerpt privacy policy and after that find the sensitive data. The security is guaranteed in this framework using key rotation algorithm which is an efficient encryption and decryption technique for safeguarding the data over big data. Data encryption is a means to protect data in storage with containing a key encryption saved and accessible to reuse the data while required. The outcome shows that the research method guarantees greater security for enormous amount of data and gives beneficial info to related clients. Therefore the outcome concluded that the proposed method is superior to the previous method. Finally, this research can be applied effectively on the various domains such as health care domains, educational domains, social networking domains, etc which require more security and increased volume of data.

***Index Terms*: Big data, security, Hadoop, MapReduce, encryption**

I. INTRODUCTION

As big data turn out to be increasingly accessible, Security and privacy aspects are continuing. The important sources of Big Data are in the finance and business in which vast volume of banking, stock exchange, online and onsite purchasing data goes via computerized systems day by day and are taken and maintained for the purpose of customer behavior, inventory monitoring, and market behavior. We could as well seen big data in the life sciences in which big sets of data for instance clinical data, genome sequencing, and patient data are examined and utilized to advance inventions in science and research. The increase of big data epoch in the Internet world has resulted in the instable development of data size. On the other hand, trust problem has turn out to be the main issue of big data, resulting in the trouble in safe data dissemination as well as industrial growth. The block chain technology for this kind of issue offers a novel solution by means of uniting traceable, non-tampering, characters with smart contracts, which helps in running default instructions automatically. With the intension of guaranteeing the safe circulation of data resources, it provides a trustworthy big data sharing model dependent upon block chain technology and smart contract [1]. At present plentiful information security (IS) incidents in large networks of organizations have turn out to be more complex and as well destructive. Therefore the systems containing appropriate services related to security in position to alleviate and punctually reply to IS threats by serving organizations well recognize their present state of network, in addition to carry out repetitive job in large IS-related data processing unit in automatic style are wanted as not ever earlier. Such centers named as Security Intelligence Centers (SICs) and Security Operations Centers (SOCs) since their subsequent development step. The summary of main characteristics of SICs is discussed [2]. The business logic of SIC as well as data architecture are presented. These outcomes bring about the key region of advance research. A rising amount of individuals wish to upload or post their day-to-day activity info on the internet media. At the period in-between, social network has been introduced in big data epoch creates personal info security issues more apparent. The research focuses on big data, investigates the present condition of secrecy protection on social network as well as examines the causes of secrecy violation on social network, and after that presents equivalent counter measures according to the viewpoints of information regarding users security related literacy, laws and regulations along with security protection technology, anticipate assuring personal info security applicable on social network in the epoch of big data [3]. In [4], for identifying progressive attacks over virtualized infrastructures, a new big data related security analytics method was utilized. Network logs and user application logs gathered once in a while from the host virtual machines (VMs) are kept in the HDFS. After that, by means of MapReduce parser based detection of possible attack paths, and graph-based event correlation, taking out of attack features is carried out. Subsequently, the presence of attack identification takes place through two-step machine learning, called belief propagation and logistic regression. Belief propagation is used for computing the trust in presence of an attack built on them and logistic regression is used for computing conditional probabilities of attack regarding the attributes.

Manuscript published on 30 August 2019.
*Correspondence Author(s)
Gunjan V. Keswani, Department of Computer Application, Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, India. (Email: keswanigv@rknec.edu)

In order to assess the proposed method, Experimentations are carried out by means of utilizing renowned malware and in contrast with previous security methods for virtualized infrastructure. Hadoop MapReduce is a programming structure in order to efficiently constituting requests that make boundless measures of info (multi-terabyte info sets) in parallel on wide bunches (numerous hubs) of merchandise fittings in a trustworthy, shortcoming forbearing manner [5, 6]. A MapReduce skeleton consists of two portions. They are "mapper" and "reducer". Basically this proposed work maintains tabs on MapReduce modifying model, planning activities, supervision and re-running of the fizzled assignments. Hadoop clusters are utilized to handle vast volume of data inside as well as outside of organizations. Accordingly, it has turn out to be significant to be capable of locating and eliminating data efficiently as well as resourcefully. It utilizes Safe Delete, a complete framework [5], which transmits file info to the block management layer through an ancillary communication route. The framework finds data blocks which are not deleted and changes the regular deletion process in the HDFS. It defines a new safe deletion method in HDFS, which produces an arbitrary pattern and also writes numerous times to the disk location of data block. The main goal of proposed research method is to introduce the encryption techniques over big data which is processed over hadoop map reduce environment. This research method attempts to ensure the secured handling of big data regardless of increased data arrival rate over periodic time. This research method introduces the encryption for the big data which is executed on HDFS and mapreduce framework to ensure the increased security rate.

This section provides the detailed introduction regarding the role of big data and the security needs. In section 2, various methods introduced by different research to efficient handle the big data is discussed with their working procedure. In section 3, the proposed research methodology is discussed along with suitable examples. In section 4, discussion about the overall research method in terms of their simulation environment is given. Finally in section 5, overall conclusion of the research work based on simulation outcome obtained is given.

II. RELATED WORK

Submit your manuscript electronically for review. Alsuhibany et al (2016) applies the big data technology, and period in-between, take the semantic processing approaches. It unites the processing of big data with semantic techniques and utilizes a structure for doing analysis on security in big data epoch with semantics. Using this structure, it could unite security data from multi-sources, and then process and examine data by means of semantic connection knowledge that is assumed to know semantically about security data and also process data with inference techniques as well as semantic association. The structure could be operated with present threat intelligence sharing in addition to swapping technique. Achana et al (2015) presenting a big data security technique in which it initially chooses some attributes, which contains a greater value compared to the remaining and protect them that consecutively offers security to the complete big data. As it is utilizing a selection technique, the relevance amid a dataset's attributes is extremely significant. As a result it concentrates on two foremost things- primarily, protecting big data by means of safeguarding beneficial info inside. Next, it is difficult to safeguard entire big data along with its attributes. It takes big data that contains its own attributes as a single object. It takes up that an attribute that contains a greater significance is extremely significant than other attributes. After that it concentrates on protecting the data by means of utilizing the MOBAT method. Data masking is a method wherein the real set of data should be substitute with a different set of data, which isn't actual on the other hand genuine. The big data is masked by means of utilizing a mathematical formula, which utilizes modulus operator. Therefore, it offers security to big data by means of these methods. Toga et al (2015) have utilized a structure to share vast bio medical data. This structure converses policies for data sharing, which might fulfill the requirements for safely sharing big data scientifically such as logging access, security from humans, managing PHI, handling other supervisory acquiescence problems, models and policies for data accessibility, sharing, managing agreements on data usage, and dealing cost efficiency in addition to sociological challenges. Liu et al (2014) recommended a technique for verifying integrity of big data based on cloud while a third party is responsible for auditing. The auditing technique is dependent upon BLS signature facilitating the technique to aid fine grained requests. The technique as well integrates authorized auditing that comprises an extra authorization method to filter illegal audits. The issue of enhancing efficiency in confirming recurrent updates is as well conversed. Yoon et al (2014) applied a solution for identifying compromised MapReduce workers via log analysis. In a Hadoop ecosystem, MapReduce nodes are susceptible to attacks from malevolent strangers and are vulnerable to cyber-attacks in which the assailants might utilize distributed system resources. Furthermore, logical errors as well as data miscalculations are probable to go unnoticed while such malevolent servers/aggressors are engaged. The method presented in this research recommends semantic analysis of Hadoop logs with the intension of identifying malevolently distressed nodes or normal nodes, which aren't creating the anticipated result: It identifies "cheating nodes" or normal nodes, which might avoid computations in an attempt to protect computational resources - Malevolent nodes that might anticipated to carry out further computations, the input data and output data might be tried to be read and maximizing the cost. In 2015, Gadepally et al utilized tool known as Computing on Masked Data (CMD) that unites improvements in cryptographic tools as well as database technologies in order to offer a low overhead technique to unburden some mathematical procedures safely to the cloud. Since the tendency of moving storage required for data as well as computation of data to the cloud rises, homeland security tasks must know the effect of security on main signal processing kernels for instance thresholding or correlation. This research defines the design as well as CMD tool development.

Retrieval Number: J88830881019/19©BEIESP
DOI: 10.35940/ijitee.J8883.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

In 2014, Islam et al designed a method to offer satisfactory security to the unstructured data by taking the kinds of the data and their sensitivity levels. It has revised the diverse analytics techniques of big data that provides the proficiency to construct a data node of databases of diverse kinds of data. Every kind of data is additionally categorized to offer satisfactory security and improve the overhead of the security system. A security suite was developed by integrating diverse security standards and algorithms with the aim of offering security to data node. The appropriate security standards or algorithms are triggered by means of utilizing an algorithm that is interfaced with the data node. It is proved that data classification regarding sensitivity levels improve the system' performance. All the existing research works discussed above provides the way of efficiently handling the big data. However none of the research work focuses on the security violation which might arise while handling mixed variety large incoming data for the period of time. This needs to be focused more to ensure the security of the big data handling.

III. PROPOSED METHODOLOGY

MapReduce framework is proposed with encryption mechanism in this presented technique. It is utilized for enhancing the security by utilizing earthquake dataset over big data.

A. Problem Statement

Big data denotes a framework, which lets the examination and handling of a vast volume of data compared to the conventional data processing methods. Big data presumes a modification from the conventional methods in three diverse means: the quantity of data (volume), the degree of data generation and transmission (velocity), and the kinds of structured and unstructured data (variety). One among the most significant portions of the big data world is the usage of brand novel techniques with the intention of taking out beneficial info from data and the capability to unite data from diverse sources and diverse formats.

The examination of the foremost challenges and issues identified regarding big data security. The goal is to find the key security dimensions on which investigators are concentrating on their efforts. Lastly, it is to find out diverse methods has already been implemented with the intention of handling these issues. MapReduce: The framework of MapReduce is a method to process data paralleled by the dissemination of data as trivial chunks crosswise the clusters. The vast amount data split into chunks must be verified for interdependencies to evade serious issues when these ensuing sets are added to acquire the needed structured data. The data must be clustered based data dependent upon their target arranged for deciding priorities, processing in addition to data dependencies. When one data needs to be processed with the outcome of additional data as its input, then it is united together in an attempt to create a cluster. The clusters are created based on priority and processing of the data clusters. MapReduce method is mostly utilized for parallel wise processing of data sets crosswise numerous clusters called filtering that is the map function and producing computation outcome via aggregation , carried out by the reduce function. The heterogeneous data items are processed by the Map Join Reduce method. It doesn't shuffle the intermediary outcomes, which needs to be transferred from mapper to reducer and it evades check pointing of outcomes often. In MapReduce, the map tasks as well as incomplete reduce tasks are re-implemented rather than the complete map and reduce tasks in the failure of a single node. It could attain the least time of execution. It is able to recognize data semantics, making simpler the writing of applications based on analytics and by minimizing phases of MapReduce, possibly enhancing performance (Sehrish et al. 2013). Reduce job includes two subtasks known as joiner and reducer. Here, the joiner possesses the union of the intermediary outcomes from the map jobs and for carrying out aggregation, reducer subtask is utilized. Next to the map and reduce jobs, the final outcome is kept in HDFS. MapReduce is used in Earthquake Datasets that comprises info such as Earthquake ID, Date time, Latitude, Longitude, Magnitude, Depth, NST and Region. Therefore the Earthquake dataset is provided to map-reduce process as input for the fine-tuning and well structuring. The Hadoop MapReduce program is provided with the above earthquake datasets as input and the MapReduce process is executed for the 'N' number of data in the dataset. This process is implemented for inputs regardless of any size. It aids quick and effective processing of the data so that unstructured data of any quantity is structured effectively. The MapReduce method produced output file that must be eliminated consistently beforehand executing it with the intention of evading file previously exist exception. Two stages are there in the MapReduce framework '"Map" and "reduce" shown in Fig.1. Every phase contains key-value pairs for both input as well as output [12]. For developing these stages, it requires to state two functions: a Mapper that is a map function and a Reducer is a reducing function. As a result, a master node is essential to execute the services needed to manage the communication amid Mappers and Reducers. After that, an input file is divided up into pieces of fixed sized. This is known as input splits. Then, these splits are delivered to the Mappers and separately work to process data comprised within every split. After that the Mappers would be able to process the data and divide the output. Every Reducer collects the data partition from every Mapper, unites every Mapper, processes them, and creates the output file that designated for them. Every Map task contains an input file and produces result files denoted by r. Every Reduce task contains m input files that are produced by m map tasks. Usually the input files provided for map tasks are exist previous to the execution of job, therefore the dimension of every map input file is identified beforehand scheduling. On the other hand, map tasks vigorously produced the output files for the period of execution; henceforth it is hard to identify the dimension of these output files previous to the execution of job.

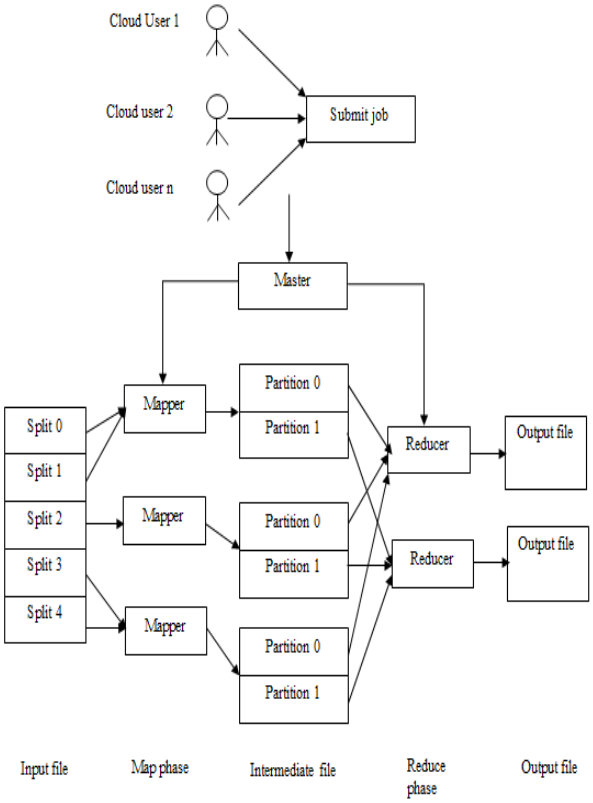

Fig. 1 MapReduce diagram

MapReduce is a programming model designed for files to deal with big volume of data. It hands out the workload amongst numerous machines and they parallel work on that data. MapReduce is a comparatively simple means to make distributed applications. The Fig 2 displays complete block diagram of the research method.

HDFS: HDFS is the main storage system utilized by Hadoop based applications. It is a distributed file system, which offers greater throughput access to application data producing numerous imitations of data blocks and dispensing them on compute nodes all through a cluster to facilitate consistent as well as fast computations. HDFS structural design mostly comprises two components known as NameNode and DataNode. A HDFS cluster contains two kinds of node: a NameNode referred as the master and an amount of DataNodes referred as workers which are operating in a master-worker pattern. The NameNode handles the file system namespace. The pillars of the file system are Datanodes. They store as well as retrieve blocks while they are expressed to (by clients or the NameNode), and they report back to the NameNode once in a while with lists of blocks that they are storing.

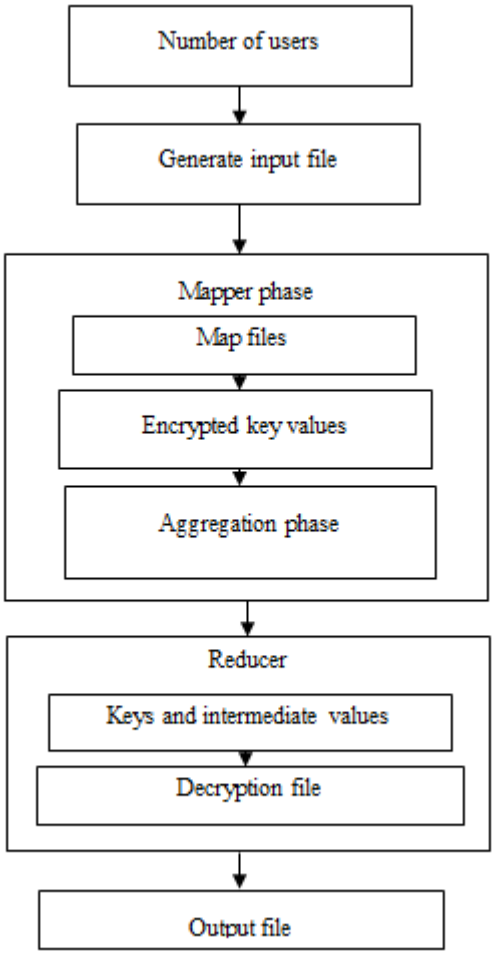

Fig. 2 overall block diagram of the proposed system

Name Node prefers regarding imitation of data blocks. In a HDFS, the size of block is 64MB and replication factor is 3. The HDFS namespace is a pyramid of files and directories, which are signified by inodes on the NameNode that record attributes such as modification, namespace, access times and permissions and disk space quotas. The content of file is divided into huge blocks (normally128 megabytes, nevertheless file-by-file that is user selectable) and each file's block is self-reliantly imitated at numerous DataNodes (normally three, but then file-by-file that is user selectable). A client of HDFS primarily contacts the NameNode which requiring reading a file for the places of data blocks encompassing the file and after that contents of reads block from the DataNode nearby the client. While the client requests the NameNode for writing data to recommend a set of three DataNodes to host the block replicas. Finally, the client writes data in a pipeline style to the DataNodes.

HDFS maintains the complete namespace in RAM. The list of blocks and inode data are in every file encompass the metadata of the name system known as the image. The insistent record of the image stored in the local hosts native files system is known as a checkpoint.

Retrieval Number: J88830881019/19©BEIESP
DOI: 10.35940/ijitee.J8883.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

The NameNode keeps the modification log of the image known as the journal in the local hosts native file system. For enhanced durability, redundant copies of journal and checkpoint are created at other servers. For the period of restarts the NameNode renovates the namespace by reading the namespace and restating the journal.

Two files in the local hosts native file system are used to represent every block replica on a DataNode. The first file comprises the data itself and the next file is blocks metadata comprising blocks generation stamp in addition to the checksums for the block data. In the course of startup every DataNode links to the NameNode and does a handshake in order to confirm the namespace ID as well as the software version of the DataNode. The DataNode automatically shuts down when either doesn't match that of the NameNode. The DataNode registers with the NameNode subsequent to the handshake. DataNodes obstinately keep their exclusive storage IDs. An internal ID of the DataNode is storage ID that helps to make it familiar though it is resumed using a diverse IP address or port.

By utilizing the HDFS, User applications of client access the file system, the file system interface of HDFS is exported by a code library. As compared to several traditional file systems, HDFS aids operations to create and delete directories and processes to read, write and delete files. By paths in the namespace, files and directories are referenced by user. Usually, the user application doesn't want to understand file system metadata and storage are on various servers, or that blocks contain several replicas. HDFS offers an API, which reveals the places of blocks of a file. This API enlightens the read performance by allowing applications such as the MapReduce method to schedule a task to where the data are situated. Normally, the replication factor of a file is three. It lets an application to set a file's replication factor. A greater replication factor for critical files or files that are accessed very frequently enhances tolerance in contradiction of faults and maximizes their bandwidth of read process.

MapReduce with Security: By means of utilizing MapReduce for parallelizing encryption processes could definitely enhance performance, since rather than encrypting blocks one after another, several mappers could perform on diverse blocks for encrypting, subsequently which all blocks are united by the reducer and keep those blocks back in HDFS. For this process, a number of deviations are probable. For example, diverse keys are utilized for individual blocks to encrypt or the identical keys are utilized for all blocks to encrypt. A big amount of keys could definitely delay the process of decryption. The mapper in the MapReduce process would consider the pair in the form of <block_id, data>, where the block is a portion of the data kept in HDFS and block_id exclusively recognizes the same block.

The mapper would create the outcome in the form of <block_id, c_data> here c_data signifies the content of the block which is encrypted. Then, the reducer would consider the input in the form of pairs and keep them as adjacent blocks. When encryption process required various keys to be utilized, every reducer would process those blocks encypted with the similar key exists in HDFS files in consecutive manner. The amount of reducers based upon the entire amount of keys being utilized in encryption in addition to size of data. The reducers deal with the encryption in addition to the decryption via writing the output file to the HDFS.

In MapReduce framework, the security is guaranteed using key rotation algorithm. MapReduce keys are arbitrary strings. MapReduce computation is employed to group the words and then apply COUNT to evaluate the number of groups. The sensitivity of COUNT is identical to the maximum number of distinct keys that a mapper can output after operating on any input record.

Key rotation algorithm is an effective encryption and decryption technique for safeguarding the data over big data. It is the finest algorithm wherein it could store data or direct data on the big data environment by means of utilizing a shared symmetric key. Here the data owner offers the authorization to encode the data by utilizing the shared symmetric key technique. A user could store data utilizing the encryption technique. The service provider is with the assistance of the data owner who offers assistance to transform plain data to cipher data and store in database. By similar means he offers the permission to take out data from the cloud data centre in an encrypted way and provides it back to the user in a decrypted way.

Procedure

1. Obtain the file
2. Encrypt the data utilizing keys
3. The authorized user sends the data access request to server and third party
4. Third party verifies the authorized user, when the user is verified then, it transfers the authorization signal to the server
5. It sends the key values and encrypted file to the requested user.
6. Server sends encrypted file to the user.
7. User compute the key values and encrypted file received from the server then verifies with key values
8. When both values are verified then user gets a data decryption key and decrypts the data.

IV. RESULT AND DISCUSSON

In this section, simulation evaluation of the proposed and existing research methods has been given. This comparison is done on the Hadoop environment for the large volume of data. This comparison is made to ensure the efficiency of the proposed research method than the existing methods. Here the parameters that are considered for the proving the security level of proposed method are, "Accuracy, Reliability, Time complexity and Error rate". Here the comparison is made between the proposed research method MRS-HDFS and the existing work Map Reduce framework [15].

A. Accuracy

It is the exactness of the model and is assessed as the over-all actual classification parameters $(T_p + T_n)$ that is categorized by the totality of the classification parameters $(T_p + T_n + F_p + F_n)$. It is calculated as given in Eq (1) :

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

$$\text{Accuracy} = \frac{T_p + T_n}{(T_p + T_n + F_p + F_n)} \quad (1)$$

Here T_P is called the number of exact predictions that an instance is negative, T_n is known as the number of improper predictions that an instance is positive, F_p is called the number of improper predictions that an instance negative, and F_n is called the number of exact predictions that an instance is positive.

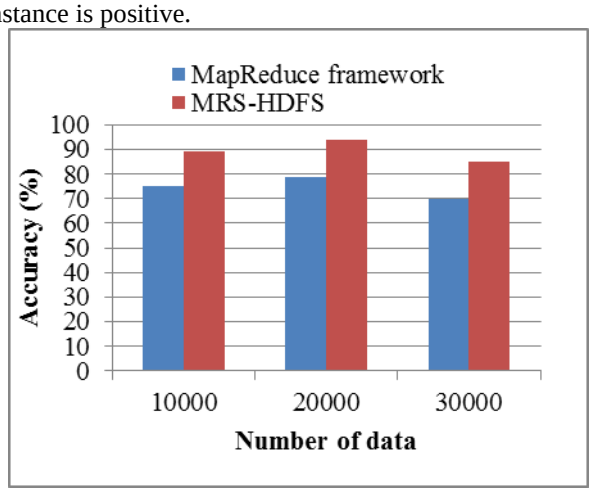

Fig. 3 Accuracy metric

According to Fig. 3, it is proved that the comparison metric is calculated amid the previous and proposed technique regarding accurateness. The algorithms are taken on the x-axis and the accuracy is taken on y-axis. The previous technique provides less accuracy but the proposed method provides greater accurateness for the provided Earthquake sample input. The proposed MRS-HDFS method selects best relevant files. The outcome proves that the proposed MRS-HDFS method gains better security outcomes than the previous MapReduce framework where it shows 19.64% better accuracy rate.

B. Reliability

While the technique offers lesser communication overhead, the system is known as superior.

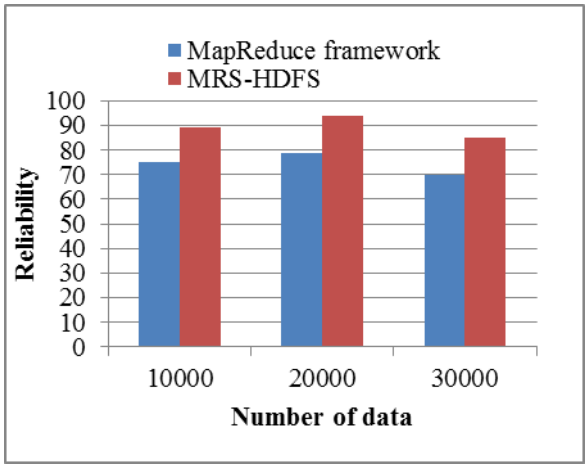

Fig. 4 Reliability

According to Fig. 4, it is proved that the performance metric is evaluated in regard to reliability. The amount of data is considered on the x-axis, and reliability is considered on the y-axis. The previous technique of MapReduce framework provided less reliability and the proposed MRS-HDFS technique gives greater reliability for the provided Earthquake dataset. Henceforth the experimentation outcome shows that the proposed MRS-HDFS method offers improved security level in big data. The reliability of the proposed research method MRS-HDFS shows 19.64% better performance than the Map reduce framework.

C. Time Complexity

While the algorithm offers lesser time complexity, the system is improved.

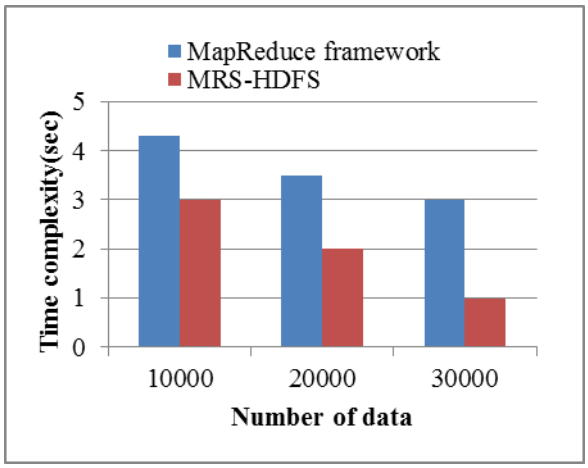

Fig. 5 Time complexity

According to Fig. 5, it is proved that the performance metric is evaluated in regard to time complexity. The amount of data is considered on the x-axis, and time complexity is considered on the y-axis. The previous technique MapReduce framework provided greater time complexity and the proposed MRS-HDFS technique offers lesser complexity time for the provided Earthquake dataset.

Henceforth the experimentation outcome shows that the proposed MRS-HDFS method offers improved security level in big data. The time complexity of the proposed method is lesser than existing method Mapreduce framework which is 44.44% lesser.

D. Error rate

While the method offers lesser error rate, the system is superior

Retrieval Number: J88830881019/19©BEIESP
DOI: 10.35940/ijitee.J8883.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

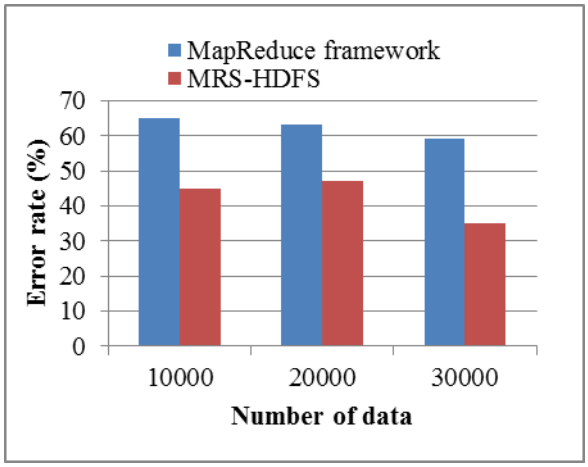

Fig. 6 Error rate

According to Fig. 6, it is proved that the performance metric of previous method is compared with proposed method in regard to error rate. The false rejection rate is taken on the x-axis, and the error rate is considered on the y-axis. According to the outcomes, it is proved that the previous method provided greater error rate and proposed MRS-HDFS technique contains lesser error rate. The outcome displays that the proposed MRS-HDFS technique offers improved security outcomes for the provided Earthquake dataset. The proposed method MRS-HDFS achieves lesser error rate where it is 32% lesser than the existing method Map reduce framework.

Table 1 shows the improvement in the proposed method MRS-HDFS with respect to the existing method Map Reduce framework based on simulation parameters.

Table1: Improvement in MRS-HDFS with respect to Map Reduce framework based on simulation parameters

Simulation Parameters	Improvement in MRS-HDFS with respect to Map Reduce framework
Accuracy (%)	19.64% more
Reliability (%)	19.64% more
Time complexity (sec)	44.44% less
Error rate (%)	32% less

V. CONCLUSION AND FUTURE WORK

In the proposed system, MRS-HDFS method is introduced for the handling security in the big data framework. In order to store and manage big data Hadoop and HDFS are utilized. The Earthquake dataset is assessed by means of utilizing MRS-HDFS technique for guaranteeing greater security. It is beneficial for big, long-running jobs and it safeguards sensitive information for the provided dataset. This research method introduced the encryption method such as key rotation algorithm on the map reduce and HDFS framework and analyzed their effectiveness in terms of data protectiveness. The experimentation outcome shows that the presented MRS-HDFS method guarantees improved security for big data and offers essential info to respective clients. Therefore the outcome proved that the presented method is superior in regards to greater accuracy, reliability and lower error rate, time complexity more willingly than the previous technique. Thus this research work can be applied effectively on the various domains such as health care domains, educational domains, social networking domains which require more security and increased volume of data. However this research work doesn't focus on various security threats which might occur due to data unavailability problem. This can be focused in future work scenario for the better handling of data. And also more recent encryption techniques need to be introduced in future for adapting the dynamic arrival of large volume of data without affecting their originality.

REFERENCES

1. Y. Li, et al., "Big Data Model of Security Sharing Based on Blockchain", *Big Data Computing and Communications (BIGCOM)*, 2017 3rd International Conference, IEEE, 2017.
2. N. Miloslavskaya, "Security Intelligence Centers for Big Data Processing", *Future Internet of Things and Cloud Workshops (FiCloudW)*, 2017, 5th International Conference, IEEE, 2017.
3. L. Qilian, et al, "Security in big data", *Security and Communication Networks*, vol. 8(14), 2015, pp. 2383-2385.
4. W. T. Yein, H. Tianfield, and Q. Mair, "Big Data Based Security Analytics for Protecting Virtualized Infrastructures in Cloud Computing", *IEEE Transactions on Big Data*, 2017.
5. A. Bikash, et al., "SD-HDFS: Secure Deletion in Hadoop Distributed File System", *Big Data (BigData Congress)*, 2016, IEEE International Congress, 2016.
6. J. Dittrich, and J. Arnulfo and Q. Ruiz, "Efficient big data processing in Hadoop MapReduce", *Proceedings of the VLDB Endowment*, vol. 5(12), 2012, pp. 2014-2015.
7. A. A. Suliman, "A space-and-time efficient technique for big data security analytics", *Information Technology (Big Data Analysis) (KACSTIT), Saudi International Conference, IEEE*, 2016.
8. R.A. Achana, S. Ravindra, S. Hegadi, and T. N. Manjunath, "A novel data security framework using E-MOD for big data", *Electrical and Computer Engineering (WIECON-ECE)*, 2015, IEEE International WIE Conference, 2015.
9. T.W. Arthur and I.D Dinov, "Sharing big biomedical data", *Journal of big data*, vol. 2(1), 2015, pp.7.
10. L. Chang, et al., "Authorized public auditing of dynamic big data storage on cloud with efficient verifiable fine-grained updates", *IEEE Transactions on Parallel and Distributed Systems*, vol. 25(9), 2014, pp. 2234-2244.
11. R Farah Sayeed, S Princey, S Priyanka, "Deployment of Multicloud Environment with Avoidance of DDOS Attack and Secured Data Privacy", International Journal of Applied Engineering Research, Vol 10 (9), 8121-8124.
12. G. Vijay, et al., "Computing on masked data to improve the security of big data", *Technologies for Homeland Security (HST)*, 2015 IEEE International Symposium.
13. Islam, Md Rafiqul, and Md Ezazul Islam, "An approach to provide security to unstructured Big Data", *Software, Knowledge, Information Management and Applications* (SKIMA), 2014 8th International Conference (IEEE).
14. S. Sehrish, G. Mackey, P. Shang, J. Wang, and J. Bent, "Supporting HPC analytics applications with access patterns using data restructuring and data-centric scheduling techniques in MapReduce", *IEEE Transactions on Parallel and Distributed Systems*, vol. 24(1), 2013, pp.158-168.
15. W. Wang, Y. Tan, Q. Wu, Q and Y. Zhang, "micMR: An efficient MapReduce framework for CPU–MIC heterogeneous architecture", Journal *of Parallel and Distributed Computing*, vol. 93, 2016, pp. 120-131.

AUTHORS PROFILE

Ms. Gunjan Vinay Keswani is currently working as as Assistant Professor in the Department of Computer Application with Shri Ramdeobaba College of Engineering and Management since 2015. She has completed her B.E. in Computer Technology from Rashtrasant Tukadoji Maharaj Nagpur University. She has an MTech degree in Computer Science and Engineering from RTMNU. She has published and presented altogether five papers in reputed International Journals including Thomson Reuters (ESCI) and International Conferences. In addition to this she has attended AICTE-NPTEL Faculty Development Programme on Big Data Computing. Her areas of specialization include Big data Computing, Data Mining, Network security and Wireless Sensor Networks. She has total experience of ten years which includes teaching experience of six years and four years of industrial experience.

Retrieval Number: J88830881019/19©BEIESP
DOI: 10.35940/ijitee.J8883.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)