INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

DATA STORAGE, PROCESSING, AND MINIMAL REDUNDANCY IN A CLOUD COMPUTING ENVIRONMENT

Chandu Rohanth[1], Dr. Devireddy Srinivasa Kumar[2], Devireddy Pravallika[3], Anusha Yanamala[4] and Aravind Yanamala[5]

[1]Department of Electronics and Computer Engineering, K L University, Vaddeswaram, Guntur (Dist.), Andhra Pradesh, India, [2]Professor & Director, Malineni Perumallu Educational Society's Group of Institutions, Guntur Andhra Pradesh, India, [3]Test Automation Specialist, IBM India PVT.LTD., Hyderabad, India, [4]Department of Computer Engineering, RVR&JC College of Engineering, GUNTUR, Andhra Pradesh, India, [5]Department of Computer Engineering, KKR & KSR Institute of Technology and Sciences, GUNTUR, Andhra Pradesh, India, chandurohanth@gmail.com

Abstract: Cloud computing is a system that, to put it another way, enables the quick transfer of resources such as servers, storage, networks, and services from one place to another with little effort required for management. For the most part, it is an architecture for large-scale distributed computing that is built on processing power, storage, applications, and services that are virtualized and managed in a dynamically scalable and controlled manner. Users have the ability to access all of their projected computing needs via a single point of access that is provided by the cloud. Customers do not need to be concerned about the infrastructure at their physical locations since they may store their data in the cloud and utilize software that is available on demand from the cloud. When it comes to computers, dependability and safety are two of the most important problems that need to be addressed immediately. In order to address the concerns raised by this research about the reliability of data centers and the potential vulnerabilities associated with them, a ground-breaking new platform has been designed. The Cloud Storage platform uses data redundancy solutions such as replication, RAID, and erasure coding to make sure users always have access to the resources they need (reliability). Installation procedures for systems that use cloud computing may be broken down into three categories. There have been instances of public, private, and even a combination of all three types. In the proposed system,

ISSN: 2057-5688

there is already an existing private cloud that is maintained by the specific business and serves as the basis for the cloud's development. Erase Code was included into the architecture of the private cloud computing system in order to reduce the negative effects of replication and RAID. It offers a number of advantages, some of which are low storage costs, low normalization costs, quick processing speeds, simplicity of data center recovery in the case of a catastrophe, and support in building up a secure cloud storage system. These benefits are just some of the things it offers.

Keywords: Data replication, RAID, Erase code, computing power, and storage cost.

I. Introduction

It makes a significant number of copies of this item to keep in its various locations. The data center provides access to a substantial amount of storage capacity. Given that all of the data might be delivered locally, replication also delivers improved or more indirect performance for virtual servers. When a replication system is utilized in the storage system, the machine's longevity, reliability, and accessibility are all increased, as are the speed at which queries are processed and the correctness of the answers they provide. On the other side, a centralized solution would need for a larger amount of storage space. The cost of updating is high, and verifying data integrity across the multiple security components is challenging because to the enormous normalization throw. The expense of updating is high. Cloud computing platforms make it possible for on-demand, on-demand system entry into a common pool of configurable computing tools (such as a server, storage, systems, and services), which can therefore be instantly constructed and moved with a minimal direction campaign [3]. Customers are able to store their data in the cloud and get the remote-requirement call-loud software they require without being restricted by infrastructure constraints. The data partitioning strategies, including Replication and Erasure Code, are used by the cloud storage platform, which enables the tools to be immediately accessible at any time.

Traditional encryption methods such as DES or AES may be used to encode the

ISSN: 2057-5688

message using either Reed-Solomon code (RS) or even Tornado before preserving the information, thereby reducing storage disc space and cloud data center procurement. Encryption and decryption are performed using a symmetric-key, whose primary is protected from virtual servers. After the data owner's permission, users may use the key to decrypt the information. An overview of the various approaches and how they favor shorter storage distances is provided in this chapter. The most significant contribution of the job is the calculation and formation of various storage sizes and calculation durations of the combination of multiple erasure codes with varying safety calculations. In comparison to replication, erasure codes are more cost-effective, require less storage space, are more durable, and are easier to retrieve. Each block of k message emblems is encoded straight to some code sentence symbols, with the addition of n-k test symbols generated from the message emblems. The n-character code is kept in a different data center. Using Reed-Solomon code and erasure code, this chapter aims to make it easier to store data in cloud data centers.

II. Literature Review

Peter Kunsz and his colleagues (2005) Provide some instances of early data storage strategies that made use of data replication in order to maintain high availability, fault tolerance, and very low access times over a large power-spread database. In order to cut down on the amount of time needed to obtain information, the data will be copied and stored in a number of different data centers. In addition to controlling the durations of their cycles, it is essential to keep these replication copies consistent and ensure that they always include the most recent information. Through the use of replica management, document transmission protocols and services (grid FTP), copies, and metadata catalogues have all been meticulously handled and organized. The user is able to do their own copy selection procedure for the purpose of acquiring user information using this grid FTP. This page discusses a variety of modules, including replication management service center components,

ISSN: 2057-5688

optimization, subscription, balancing, session management, and security.

Cong Wang et al (2012) offered a supply storage ethical audit system without the need of server hardware or application involvement. ' Users may reorganize cloud storage operations using nominal and coded data using this design technique. As a result, the malfunctioning function cloud, as well as r storage correctness, have been discovered. Data corruption and host misbehavior may be rapidly identified by the Administrator during the information storage correctness confirmation, even if the same level of storage correctness confidence has been kept by using token computation and erasure-coded data. Dispersed 'm+k' servers have been set up to prevent the information file from collapsing. In this case, the authors inserted tokens on the vectors before the file supply. Additionally, the faulty data is reconstructed while these encrypted components are saved on various cloud servers.

They depend on pre-computed token confirmation to assure data storage accuracy and to pinpoint the location of data malfunctions. Before the provision of documents, a person estimates a certain number of tokens to be used. Each token consists of two blocks of data. It takes a user a long time to earn the storage and accuracy of the data because of all the blocks that cloud servers make while calculating and bringing difficulties to an individual. Individuals have the option of either storing tokens on their own or encrypting them and maintaining them on cloud-based servers.

Song tao Liang et al (2014) Recommendations for reducing the fixing bandwidth and disc I/O costs from the storage strategies were made by using hybrid storage re-generating codes. A repair theory may be categorized into three groups. There are three stages of repair: operational, accurate, and hybrid vehicle. The authors used just two items in their list of repair methods: exact restoration HMSR codes. Using a quick regeneration technique, these HMS Janin codes outperform ordered M SR codes in terms of data downloading. The HMSR algorithm completes rehabilitation by

ISSN: 2057-5688

repairing the storage system under the small dimensions of the limited region.

In 2016, Yingxun Fuet al Only disc failures should be retrieved using stack-level retrieval, rather than strip-level recovery. Both the greedy algorithm and the rotational retrieval method are extensively used in this technique's two regaining y mechanics. Both a balance-priority strategy and a search-period priority approach are used by the greedy algorithm. This fine-grained policy-based access control and document deletion system was presented by Yang tang and others in 2012. It's nothing more than a safe and secure way to store data on the cloud. Fine-grained policy set access control and document deletion have been achieved with this method. Key managers are used to maintain FADE's in-built library of cryptographic keys. Attribute-based encryption is used to provide fine-grained access control, and the fault-tolerant main direction is used to ensure deletion of files. An automated system Employing cloud-based key management agencies, managers and companies must dedicate FADE to their employees. Working with a quorum technique, on the other hand, may help strengthen the network. The files were removed as the coverage reached page 139.

III. Error correction methods

Only data that cannot be read from the disc should be recovered using the stack-level retrieval method, not the strip-level recovery method. The greedy algorithm and the rotational retrieval method are both put to great use in the two regaining y mechanics that are a part of this strategy. The greedy algorithm employs not only a search-period priority strategy but also a balance priority strategy in its decision-making process. In 2012, Yang Tang and colleagues introduced a system that provides fine-grained policy-based access control and deletes documents. There is nothing more to it than a reliable and risk-free method of storing data on the cloud. Utilizing this strategy allowed for the successful execution of fine-grained policy set access control as well as document deletion. Key managers are what are utilized to keep FADE's built-in library of cryptographic keys organized and secure. In order to offer fine-grained access control, attribute-based encryption is used, and fault-tolerant main direction is utilized, since both

ISSN: 2057-5688

of these are required to guarantee the deletion of data. A computerized program or system Employing cloud-based key management means that agencies, managers, and enterprises have a responsibility to devote FADE to their staff. On the other hand, using a quorum method might be one way to assist with the consolidation of the network. At the point when the coverage reached page 139, the files were deleted.

The erasure-coded cloud storage system has N discs, of which M contain n code word symbols and K carry n-k code information. The N discs are divided into M and K discs (check bits). In order to get the check bits from the k message symbols and the co de word symbols, the erasure code must first be utilized. Erasure codes are characterized by two unique characteristics. In the first place, it ought to be MDS (maximum Distance Separable), which indicates that if any K of the N discs are out of order, their information may be recomputed from an active M disc that is still alive. In other words, the information can be salvaged. Rebuilding whatever was on the damaged disc is the responsibility of the new disc, which is located in the data center. Due to the fact that erasure codes are methodical, the M data discs also include the raw data. The RS codes in the MDS are familiar to most people.

Operations	erasure code	Optimal erasure code
Time for encoding	$(k+ł)\ln(\frac{1}{e})^p$	$(kł)^p$
Time for decoding	$(k+ł)\ln(\frac{1}{e})^p$	$(kł)^p$
Inefficiency of decoding	1	1+e
General calculation	polynomial	X-OR

There is a Low-Density Parity Check (LDPC) chart that is part of tornado code, and all of the tornado codes have been made for reliable multicast. To create parity nodes, the data logos and parity symbols of each node have been kept. All tornado codes are derived from bipartite graphs that cascade. The number of steps and nodes required to produce a Tornado Code might vary widely. However, the fault tolerance qualities are owned by the specific border degree distribution that results in a cascaded bipartite LDPC chart. Erasure codes, as opposed to replication, provide a higher level of mistake tolerance. [8].

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

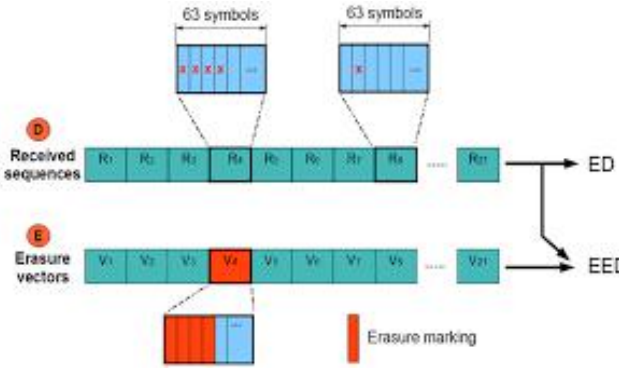


Figure 1: Function of encoding system

All procedures are carried out individually by the datacenters in order to function in a scattered environment. Adware, adware, and message routing are all necessary components of this strategy. It is possible for cloud storage systems to satisfy certain criteria of information security and data resilience and storage due to the integration of monitoring and partitioning operations.

IV. Model setup

SS inch, SS 2 are two of the n datacenters (storage servers) in this system. In the simulation, your investment's return is represented by eight data centers and four key managers. These data owners are protected by the symmetric secret that is managed by the important managers. Erasure's cloud-storage technology has three phases: data storage, data storage, and data recovery. The data recovery phase is the last phase. During the data storage time, data owners use any of their symmetric techniques to detach the communication and send it to data centers. A message M is broken up into M1, M2, M3, M4.... Mk (k=8) because it is encrypted into cypher text ci utilizing 8 data centers. SSi(i=8 storage server) with identification I D randomly selects data centers to receive these encrypted text blocks. It is saved at each data center after receiving encrypted messages from the data owners. k encoded message blocks cannot be received by any storage data center. Users may destroy received messages from symmetric-key (which is stored in primary Managers) with the permission of the information owner after the data shredding time by receiving decoded messages from information centers using identifier i-d. The administrator's petition is used to restore pieces of data that are critical to key supervisors and live data centers when the information center collapses.

V. Proposed method

ISSN: 2057-5688

Replicating data in real time may improve the performance of an application system. We are currently calculating the prevalence level and copy element in order to find the most suitable document to replicate and decide on the different replicas. To figure out where to put the copies, we're using fuzzy logic. We also use a round-robin method to deploy the replicas from the identified systems.

Algorithm for Guaranteed Reliability

1: t ← GetTime()

2: A ← Proc − (H)Digitalize sequence

3: for tier ← Client tier − 1 down to RootTier + 1 do

4:A ← Aggregate(A)

5:for all record r€A do

6:if r,numOfAccesses ≥ threasholds

7:Update − Ctime(r. fileID, r.nodeID,t)

8:Getreplicate (r.fileID, r.nodeID, t)

9:end if

10: end if

11: end for

12: end for

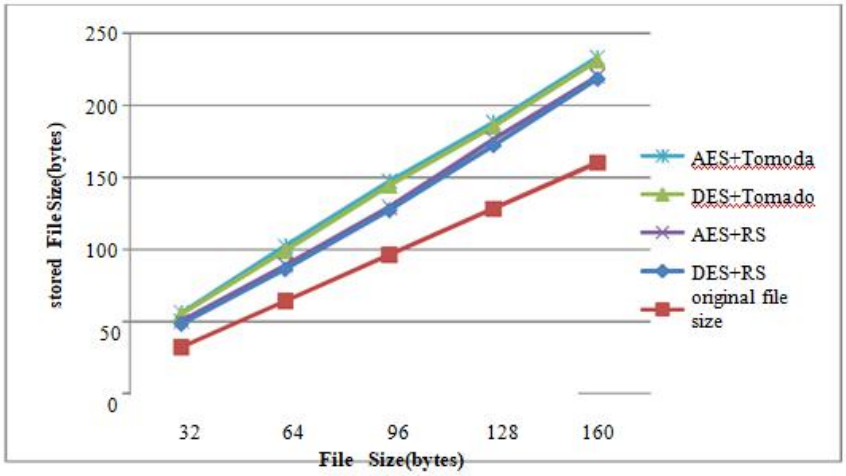
Data from A is combined into the current grade in the first algorithm lineup 4. The specifics of Aggregate's work will be studied to a considerable extent using a large data set. All recordings' node IDs are derived from the current processing grade after the aggregate. It will process further for each and every album of RIN A if r.numOfAccess is more than the current tier's limit (line 6). r.fileID exists at the node of all r.nodeIDs in the event. This is followed by the current replication session duration being updated and ep being removed from A. Alternately and regardless of whether or not the r.nodeID node has sufficient distance for document r.fileID to replicate and expel record dtc from the node (lines 10—1 2). It's probable that the remaining recordings in A will be flashed into a better inline 4 when the internal loop is finished. As previously stated, the improved option A will also be handled. Calculating the ratio between your seldom sought file on the host and the full collection of often hunted files allows you to rank the optimal document. There is an order to the most often used files. Most of your user's replies are saved in the main cloud,

ISSN: 2057-5688

which is where the files are often accessed. A significant amount of calculation and article processing may be required as a result of the host's new status. In addition, it is widely accepted that cloud computing systems often include a large number of servers. The mechanism used to determine rank makes it easy for anybody with a valid domain name to have access to cloud server data. On each server, an estimate of the number of files that can be found is computed. rnk calculates low-status documents based on the cloud host's prediction and the minimal status files. As a result, the file's reduced status chooses just the bare minimum of options. The next step is to choose a 'I' ranking of services that adheres to the predetermined shrewd preferences only insofar as a viable standing structure has been recognized. Replicated data is removed from the second host depending on the brink of a shutdown of the selected software and saved only on two servers. Additionally, the status values may be calculated and deleted from files based on RAM.

VI. Results

Storage Size: Simulator findings must be used to verify the functioning of the erasure codes platform. After shredding (preparing to save) from the information centers, the chunked file size hasn't improved substantially. As a result of this, there are minor differences across techniques. The plain-text cube's dimensions are the same as the ciphertext's size because of encryption.



ISSN: 2057-5688

Figure 2: Comparison between original block and various encodes block size

Simulator findings must be used to verify the functioning of the erasure codes platform. After shredding (preparing to save) from the information centers, the chunked file size hasn't improved substantially. As a result of this, there are minor differences across techniques. The plain-text cube's dimensions are the same as the ciphertext's size because of encryption. Even if AES uses various keywords (such as 128,192,256), the distance between this ciphertext and the preceding block of plain text is not affected by the length of the key. The size of the ciphertext and the plaintext will be the same. It is also included in the chunks when the information size is expected to increase during the communication. In storage locations where data replication mechanisms are not used, this variation is mirrored. Techniques that replicate data need n data centers, but methods that utilize erasure codes only need n -k acceptable data centers, the results of each approach vary. RS codes take up less store space when compared to T-O tornado codes since RS codes analyses just a restricted number of elements throughout the document retrieval process. The tornado code, on the other hand, requires a vast quantity of data. The size of the charts is continually increasing as more data is added to the system, and as a result, more charts are parallel.

Reed Solomon codes, as predicted, perform better throughout the processing phase. The tornado code's performance is greater to that of the RS code, however the tornado co p is bigger and constructed utilizing probabilistic assumptions and indications. Because of this, Tornado code, on average, requires more data to recover a random file than Reed Solomon code. The table below provides information on the processing time required to save data from the data center for each of the many procedures.

Table 1: processing time comparison table

File size (Bytes)	DES and RS (ms)	DES and Tornado (ms)	AES and RS (ms)	AES and Tornado (ms)

ISSN: 2057-5688

32	0.731297	0.537067	0.621 283	0.48319
64	0.853517	0.603454	0.709 421	0.57473
96	1.135577	0.759629	0.905 823	0.71143
128	1.312048	0.879186	1.045 932	0.83172
160	1.628691	0.985965	1.168 921	0.89668

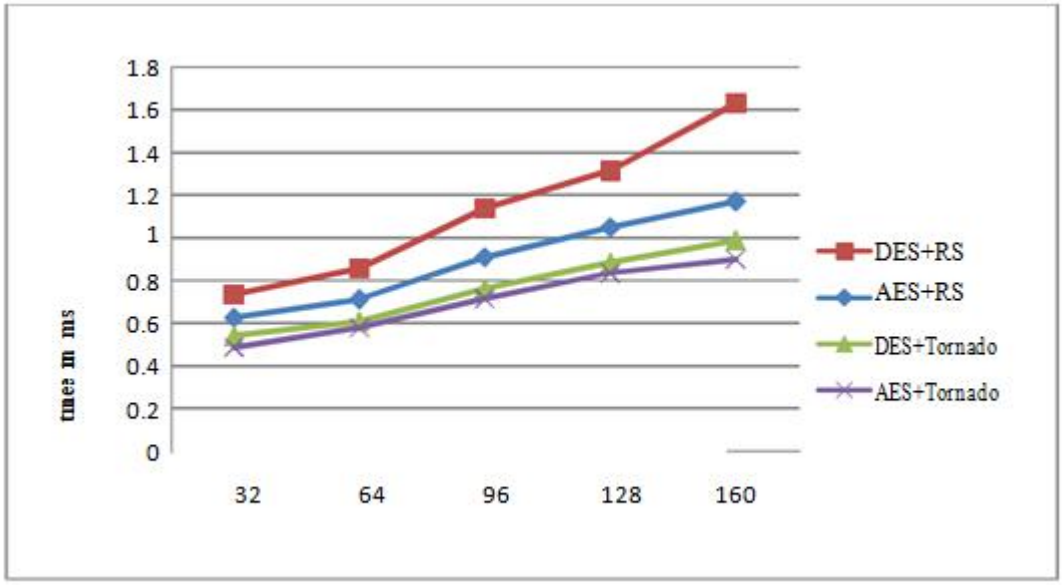


Figure 3: Comparison between various schemes storing times

Probabilistic and sophisticated graphs were utilized in R S code during the latest era of tornadoes. Reed Solomon codes, one of the four techniques of AES, deliver secured information with a low processing time.

VII. Discussion And Conclusion

This technology, which is able to convey and retain data in a safe manner, has proven to be quite successful for our company. Because of the security implications that AES256 has with its own, specific programming libraries, and applications that are a good fit for it, it

ISSN: 2057-5688

has become one of the algorithms that is employed the most often among its peers. The method may be simple to implement, but it requires a degree of security for writing that is difficult to give without also necessitating a level of resilience that is hard to do. The AES technique considers 128 separate blocks as a single unit, unlike the DES algorithm, which regards each block as a separate entity. Despite this, the AES approach is nonetheless slower to develop and less sensitive to strikes. Erasure coding uses a lot less resources than replication does, making it a far more viable option. The program finds out how long it takes AES to encrypt a protected chunked file and stores that information in the information center by using Reed-Sensible Solomon's Storage Space in conjunction with the R S code that is located on the network.

VIII. Future Scope

This article solely focused on the successful reconstruction of a single data center after it had been destroyed. It's possible that businesses offering community cloud security solutions and services would profit from this paradigm. There is a chance that T-study will It's possible that O's needed retrieval from more than one data center, and T-use operating system may come from a wide range of software applications.

IX. References

[1]. HaiyingShen, Guoxin Liu, “Swarm Intelligence based File Replication and Consistency Maintenance in Structured P2P File Sharing Systems” IEEE Transactions on Computers, Vol. 64, No. 10, Oct 2015.

[2]. SameeUllah Khan, Ishfaq Ahmad “Comparison and analysis of ten static heuristics-based Internet data replication techniques” Parallel Distrib. Comput. 68 (2008)

[3]. Zheng Yan, Lifang Zhang, Wenxiu Ding, and QinghuaZheng, “Heterogeneous Data Storage Management with Deduplication in Cloud Computing” IEEE Transactions on Big Data, Vol. pp, No.99, May 2017

[4]. Jing Zhao,XuejunZhuo, “Contact Duration Aware Data Replication in DTNs with Licensed and Unlicensed Spectrum” IEEE Transactions On Mobile Computing, Vol. 15, No. 4, April 2016

ISSN: 2057-5688

[5]. Jenn-Wei Lin, Chien-Hung Chen "QoS-Aware Data Replication for Data Intensive Applications in Cloud Computing Systems" IEEE Transactions on Cloud Computing May 2014

[6]. Rodrigo N. Calheiros, Rajkumar Buyya "Meeting Deadlines of Scientific Workflows in Public Clouds with Tasks Replication" IEEE Transactions on Parallel and Distributed Systems, Vol. 25, No. 7, July 2014

[7]. Han Hu, Yonggang Wen, Tat-Seng Chua, Jian Huang, Wenwu Zhu and Xuelong Li "Joint Content Replication and Request Routing for Social Video Distribution over Cloud CDN: A Community Clustering Method" IEEE Transactions on Circuits and Systems for Video Technology, Vol. 26, No. 7, July 2016.

[8]. S.Annal Ezhil Selvi and Dr. R. Anbuselvi, "Ranking Algorithm Based on File's Accessing Frequency for Cloud Storage System", International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 4, Issue 9, Sep 2017.

[9]. Jonathan L. Krein, Lutz Prechelt "Multi-Site Joint Replication of a Design Patterns Experiment using Moderator Variables to Generalize across Contexts" IEEE Transactions On Software Engineering, Vol. X, No. X, Month 2015

[10]. Wenhao Li, Yun Yang, Dong Yuan, "Ensuring Cloud Data Reliability with Minimum Replication by Proactive Replica Checking", IEEE Trans. Computers 65(5): 1494-1506 (2016)

[11]. YaserMansouri, Adel NadjaranToosi, and Rajkumar Buyya "Cost Optimization for Dynamic Replication and Migration of Data in Cloud Data Centers" IEEE Transactions on Cloud Computing, Vol. pp, No. 99, January 2017

[12]. Runhui Li, Yuchong Hu, and Patrick P. C. Lee "Enabling Efficient and Reliable Transition from Replication to Erasure Coding for Clustered File Systems" IEEE Transactions on Parallel And Distributed Systems, Vol. pp, No. 99, March 2017.

[13]. Jerry Chou, Ting-Hsuan Lai "Exploiting Replication for Energy-Aware Scheduling in Disk Storage Systems" IEEE

ISSN: 2057-5688

Transaction on Parallel and Distributed Systems, Volume 26, No 10, Oct 2015.

[14]. Guoxin Liu, HaiyingShen, Harrison Chandler "Selective Data replication for Online Social Networks with Distributed Datacenters" IEEE Transactions on Parallel and Distributed Systems, Vol. 24, No. 8, August 2016

[15]. Amina Mseddi, Mohammad Ali Salahuddin "On Optimizing Replica Migration in Distributed Cloud Storage Systems" 4th IEEE International Conference on Cloud Networking (IEEE CloudNet 2015)

[16]. Kan Yang &XiaohuaJia 2012, 'Data storage auditing service in c loud computing: challenges, methods and opportunities', Springer, World Wide Web, vol. 15, pp. 409-428.

[17]. Kan Yang & XiaohuaJia 2013, 'A n Efficient and Se cure Dynamic Audi ting Protocol for Data Storage in Cloud Computing', IEEE Transactions on Parallel and Distributed Systems, vol. 24, n o. 9, pp. 1717-1726.

[18]. Rashmi, KV, Nihar B Shah, Kannan Ramchandran & Vijay Kumar, P 2018, 'Information-Theoretically Secure Erasure Codes for Distributed Storage', IEEE Transactions on Information Theory, vol. 64, no. 3, pp. 1621-1646.