International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6971
***Article Received**: 30 March 2023 **Revised**: 10 May 2023 **Accepted**: 28 May 2023*

Geo-Distance Based 2- Replica Maintaining Algorithm for Ensuring the Reliability forever Even During the Natural Disaster on Cloud Storage System

Dr. S.Annal Ezhil Selvi
Associate Professor, PG. Department of Computer Science,
Bishop Heber College, Affiliated to Bharathidasan University
Trichirrapalli, TamilNadu, India- 620017.
annalabel.cs@bhc.edu.in

Abstract: In today's digitalized and globalized scenario, everyone has moved to cloud computing for storing their information on cloud storage to access their data from anywhere at any time. The most significant feature of cloud storage is its high availability and reliability then it has the capability of reducing management factors as well as incurred lower storage cost compared with some other storing methods, it is most suitable for a high volume of data storage. In order to meet the requirements of high availability and reliability, the system adopts a replication system concept. In replicating systems, the objects are replicated many times, with each copy residing in a different geographical location. Though it is beneficial to the users, it leads to some issues like security, integrity, consistency and hidden storage and maintenance cost, etc. Therefore, it is exposed to a few threats to the Cloud Storage System (CSS) user and the provider as well. So, this research seeks to explore the mechanisms to rectify the above-mentioned issues. Thus, the predecessor of the research work has proposed an algorithm named as 2-Replica Placing (2RP) algorithm which is used to reduce the storage cost, maintenance cost; and maintenance overheads as well as increase the available storage spaces for the providers by placing the data files on two locations based on Geo-Distance. But it fails to address the recovery mechanism when a natural disaster happens because providing reliability with less than 2 replicas is a challenging task for the providers. Thus, the research proposed Geo-distance based 2-Replica Maintaining (2RM) algorithm which is used to consider that issue for ensuring reliability forever even during natural disasters

Keywords: Cloud Storage, Replication of Data, Data Availability, 2- Replica, Reliability and Storage Cost.

I. Introduction

In a globalization era, it is not an overstatement to say that today's world is data-driven. In such a predicament, we are all compelled to have all of the data in our palms at all times. It is simply not possible to maintain all of the information on hand. Concurrently, if the user digitizes all of the documentation, the storage system is the largest concern, and trying to maintain that data is also a very difficult and challenging task for all. As a result, all have migrated to a fantastic technology known as Cloud computing, which allows them to access all of their data from any location with a network device.

Cloud storage is a system that operates on an "on demand" or "Pay per Use" premise. All computational resources (such as storage and data) are distributed among users in cloud computing [1]. The service provider and the consumer are connected through the service level agreement (SLA). The terms of this agreement determines Quality of Service (QoS) such as availability, Reliability, Scalability, and cost, etc. Accessibility, outstanding reliability, and high fault tolerance are crucial criteria to take into account even though cloud hassles are often only significantly dissimilar [2]. To ensure high availability, performance, and fault tolerance, replication is being used [3].

In the interest of improving efficiency and stability, cloud services store multiple copies of a file's contents as replication. The client will be paid for using the public cloud due to it being an on-demand business. Customers will favour a service provider that can meet their highest digital storage requirements. As an outcome, the process of replication is employed to achieve the highest level of availability [4, 5]. Therefore, it is not obligatory for the advantages of replication to offset the costs associated. As a result, the cost of replication must be taken into account [8, 9]. Page ranking based on user preferences was proposed and it is effective in terms of relevance since it makes use of agents to assess the relevance of a page's content and takes user behaviour into account when ranking web sites. Thus, user preference-based page ranking facilitates users' easier and more satisfying navigation of search results to get the necessary information [13]. Software frameworks for cloud computing control cloud resources and deliver scalable, fault-tolerant computing services with globally consistent and hardware-independent user interfaces. The cloud

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6971
***Article Received**: 30 March 2023 **Revised**: 10 May 2023 **Accepted**: 28 May 2023*

service provider assumes responsibility for handling infrastructure-related problems [18].

This study suggested a methodology for cloud storage systems. First, the algorithm determines where the replica will be kept. The location is chosen based on its geographical proximity. This suggested technique needs to keep the data file in three distinct locations according to the geographic location, as detailed in the following sections.

The remaining part of the paper is structured as follows. Section 2 discusses about literature review and Section 3 describes about conventional method. The subject of proposed Geo-Distance-based 2-Replica Maintaining (2RM) algorithm is explained in Section 4. Section 5 discusses the interpretation and comparison of Expected results. Finally, section 6 concludes the work by discussing future plans.

II. EXISTING SYSTEM

Navneet et al. [15] had been utilized heterogeneous system perhaps instead of homogeneous. Data centers in one tier will be configured differently from those in other levels. Each information file's original version is kept in super data centers. Duplicates are dispersed from super to main data centers and from central to typical data centers in order to meet availability. The broker receives a file access request from the user and communicates internally with the Replica catalogue in addition. The location of the duplicates of the data request files is detailed in the replica catalogue. Brokers will review the list of data centers they have been given and scheduled the request for the closest data centre. Blocks are the fundamental storage devices. A file's blocks will have variable block availability probability at various types of data centers since a file might be replicated at more than one data centre. Super data centers will have the highest block because they are more expensive and have dependable hardware. This work was created using the Dynamic Cost- Aware Re-Replication and Redistribution Method to reduce the cost of duplication (DCR2S). This algorithm switches the location of the duplicates from superior to primary data centers and/or main to conventional data centers if the demand for availability increases rather than increasing the number of replicas to reduce the cost of duplication. Ensuring Cloud Data Reliability with Minimum Replication by Proactive Replica Checking was proposed to reduce the Cloud storage consumption while meeting the data reliability requirement [10]. Large-scale data management systems often use data replication which is a well-known technique that treats generated problems by storing multiple copies of data, called replicas, across multiple nodes [11]. This method has ability to guarantee failure tolerance, a capacity advertising copies data among different copies. These copies store the same set of information, so in case any of copies is lost, information may still be gotten and recouped from the other replicas.

File Accessing Frequency based Ranking (FAFR) Algorithm [12] and Dynamically Reduced Replica for Rarely Accessed files (DRRRA) algorithm work jointly proposed [14] and identify the rarely accessed files and retain the replica in two server other replicated files are deleted. To provide access to more request with 2 or 3-replica is a complicated issue. Therefore a Dynamic replica Creation for Availability enhanced Storage (DRCAES) algorithm which jointly work with FAFR algorithm to predict most frequently accessed files and automatically replicated to other server based on server memory. This enhances the availability, thereby reducing the request-response delay time. A dynamic replica creation strategy based on the gray Markov chain was proposed [16]. If the number of replicas needs to be increased, the newly added replicas need to be placed on the DataNodes. Considering the problem of load balancing of the DataNode during replica placement, a replica placement strategy based on the Fast Non-dominated Sorting Genetic algorithm was applied. Dynamic replica Creation for Availability enhanced Storage (DRCAES) algorithm which jointly work with FAFR algorithm to predict most frequently accessed files and automatically replicated to other server based on server memory [17].

III. PROPOSED METHODOLOGY

Accessibility, good reliability, and excellent performance are some of the crucial aspects of the cloud storage system. These crucial fundamental properties of CSS are attained through the replication mechanism. The cost of replication storage and maintenance may be slightly higher due to the duplication system. As a result, users could be reluctant to transition to cloud services. Despite the fact that there are now more people are increased for using cloud storage. The Life Cycle of the Cloud storage system is depicted in figure 1.

Data replication in cloud environments is used to increase application performance while providing a reliable and constant access control speed. Data ranking is considered as a significant factor for data object and it is utilized by various data centers for making decisions to replicating data. For instance, highly requested information is replicated to additional server nodes near the location of the data. Adversely, less frequently accessible data is not replicated and transferred to server nodes with poor system performance. Furthermore, when numerous users or apps request the same data or information of a particular sort simultaneously, the replication technique assesses the data correlation to assist the system in reducing latency and bandwidth consumption.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6971
***Article Received**: 30 March 2023 **Revised**: 10 May 2023 **Accepted**: 28 May 2023*

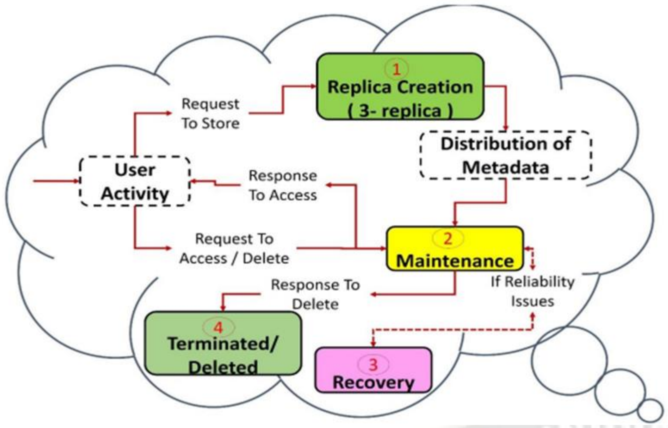


Figure 1. Cloud Storage Life Cycle

Therefore to reduce the storage space, cost, maintenance cost and overhead due to replication, this research has proposed an algorithm called 2-Replica Placing (2RP) algorithm which is also used to increase the available storage space without affecting the key features of CSS that will be discussed in next section.

IV. SYSTEM MODEL

The data files are kept in many data centers while using dispersed data storage. The data centers are spread throughout many regions. Other data centers, such as secondary or regular data centers, are available for each data centre. Also, the data warehouse is managed by a variety of managers and catalogues. With more than 500,000 sites around the world and an approximate capacity of even more than 1500 billion gigabytes, data centers are more powerful than it has ever been. The replication manager and replication catalogue are two crucial managers. All other data centers are connected to one another. This work does not contain major differences from the existing system models [7]. But the only difference, there is no variation in storage cost. It may be stored in super, primary, secondary or ordinary data centers but the storage cost is the same. The distance between the Data Centers (DC) 'DC' is calculated by using the following equation 1,

$$D_C = \sqrt{(x_i - x_j)^2 - (y_i - y_j)^2} \quad (1)$$

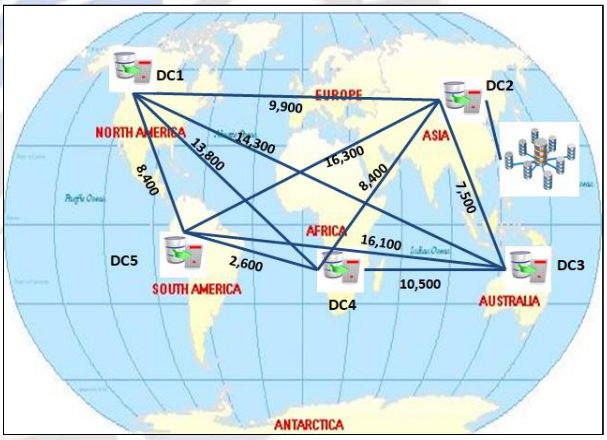
The DC's are located with the co-ordinates (xi, yi) whereas 'i' and 'j' represents the count of data centers {(i, j) = (1, 1) (2, 2),…,(n, n)}. As a result, each dynamic data center needs to be connected to an atlas by a data center.

Mathematical Notations Used in 2RM:

- m: Count of uploaded Data Files.
- n: Number of Data Centers (DC).
- Di : Distance between ith DC to jth DC, where i ← 1,⋯,n ∧ j ← 1,⋯,m.
- DFj : jth file, j ← 0,1 ∧ 2 (DF0 ← Original Data File).
  isWait, isDC (i): Variables to state the Data center's status.
- d : Selected distance ith DC to jth DC, where i ← 1,⋯,n ∧ ← 1,⋯,m. and d1 and d2 are the distance from the original data file residing DC to other two copies residing DCs
- s : Selection variable which is used to select the far geo-distanced DC to store the next DF
- DCi :ith Data Centers where i ← 1,2..n .
- The Data Center Configuration
- Di,j: Distance from ith DC to jth DC, and it is represented in following matrix (n X n) representation.

$$D_{i,j}: \begin{array}{c|cccc|} & DC_1 & DC_2 & \cdots & DC_n \\ \hline DC_1 & D_{1,1} & D_{1,2} & \cdots & D_{1,n} \\ DC_2 & D_{2,1} & D_{2,2} & \cdots & D_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ DC_n & D_{n,1} & D_{n,2} & \cdots & D_{n,n} \end{array}$$

For example, the sample system model configuration has shown here the following figure 2.



The Geo-distanced based 2-Replica Placing (2RP) algorithm which has proposed in previous work of this research that initiated when the user makes the request to store their data file on a cloud storage system. Then the newly proposed geo-distance based 2-Replica Maintaining (2RM) algorithm will take place for maintaining the 2-replica strategy forever to ensure the reliability even when the natural disaster occurred.

Proposed - Geo-distance based 2RM Algorithm:

{DC (1), DC (2), DC (3), … DC (n) set of Data Centre}

1: isWait ← False, isDC (i) ←True, isDC (j) ←True, isDC (k) ←True;{ Data center's status}

2: Monitor the DC's connection Status {Check the Status}

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6971
***Article Received**: 30 March 2023 **Revised**: 10 May 2023 **Accepted**: 28 May 2023*

3: if(isDC(i) =True & isDC(j) =True & isDC(k) = True & isWait=False) then Goto 2;

{Check any one of DC's loss their connection /destroyed }

4: Else If ((isDC(i) = False || isDC(j) = False || isDC(k) = False) & isWait=Ture) then goto 6;

5: Else wait 24 hr, isWait ← True, Goto 2;

6: Replicate the DF (Data File) from any one of the other DC

7: Find the actual distance of 'd' remaining two Data center

8: D ← d;

9: for (new←1; new<=n; new++) do

10: if (new!= i & new!=j & new!=k) then go to 11 else go to 9

11: Select next new DC (new)

12: Find the Distances of 'd1 and d2' from DC (new) to remaining two DC's

13: if (d1>=d & d2>=d) then go to 32 else go to 9

14: End For

15 for (new←1; new<=n; new++) do

16: if (new!= i & new!=j & new!=k) then go to 17 else go to 15

17: Select next new DC (new)

18: Find the Distances of 'd1 and d2' from DC (new) to remaining two DC's

19: d ← d-500

20: if (d>=D/2) then go to 21 else 15

21: if (d1>=D & d2>=d) || (d1>=d & d2>=D) then go to 32 Else go to 19

22: End For;

23: d←D;

24: for (new←1; new<=n; new++)do

25: if (new!= i & new!=j & new!=k) then go to 26 else go to 24

26: Select the next new DC (new)

27: Find the Distances of 'd1 and d2' from DC (new) to the remaining two DCs

28: d ← d-500

29: if d>=D/2 then go to 30 else go to 24

30: if(d1>=d & d2>=d/2) then go to 32 Else go to 28

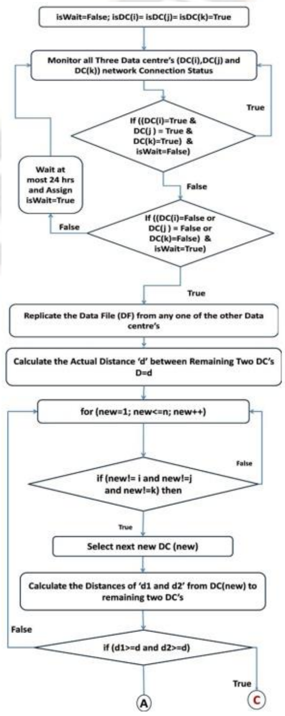
31: End For;

32: Store DF at DC (new)

The figure 3 shows the flow of the geo-distance based 2-replica maintaining algorithm in detail.

After completing the 2RP's work the 2RM's will be triggered to vigilant for ensuring the 2-replica forever. It monitors the DFs that have been stored in DCs until any disaster or some kind of connection problem occurs on that DC. If any problems are found the 2RM will wait for 24 hours, then after 24 hours it will go and re-check the connection status. If any improvement is noticed it is well and good. So, it is retaining its work. But if there is any inconvenience still retained, the remaining two DCs will be taken into consideration and their actual distance will be calculated. The DF from any one of the two DCs will be copied and it will get stored in the new DC which has the longest distance than the other two DCs. If no longest-distance DC is found the next process is started. It will consider that the new DC must be at least half the distance of the actual distance from the one DC among the two DCs. Another DC should be the longest distance from the new DC.

If the proposed system failed or not possible to find the far distanced new DC that time it uses a new mechanism to ensuring the far distance while selecting the New DC that the proposed methodology will reduce actual distance by 500 kilometers until to reach at least half distance of the actual distance. Then the DF will be stored in that new selected DC.



International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6971
***Article Received**: 30 March 2023 **Revised**: 10 May 2023 **Accepted**: 28 May 2023*

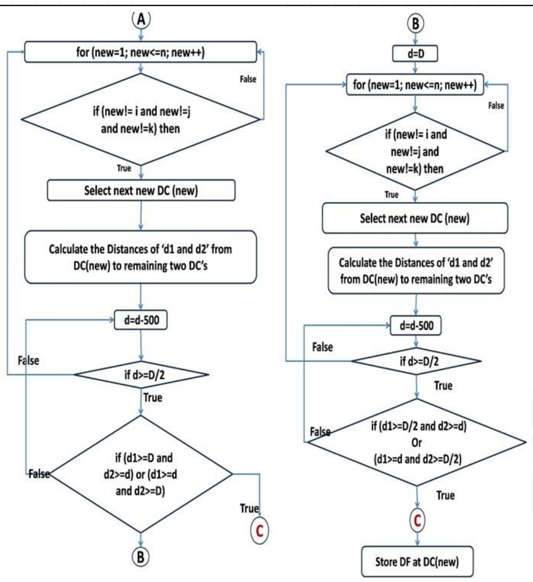


Figure 3. The Flow Diagram of the proposed Geo-distance based 2RM Algorithm

V. INTERPRETATION AND COMPARISON OF EXPECTED RESULTS

As mentioned in Figure 2.the System Model DC(1), DC(2), DC(3), DC(4), and DC(5) are data centers located in North America, Asia, Australia, Africa, and South America respectively. And the DCs connected like a complete graph. A collection of servers with different tiers makes up a data centre. It contains routers, computers, and storages. Many routers are used to connect the numerous servers to one another. Administrators can select alternative network topologies based on various strategies. The distance between the DC's are represented in Kilometers (KM) as a metrics form as follows,

Distances (D) =

	DC(1)	DC(2)	DC(3)	DC(4)	DC(5)
DC(1)	0	9,900	14,300	13,800	8,400
DC(2)	9,900	0	7,500	8,400	16,300
DC(3)	14,300	7,500	0	10,500	16,100
DC(4)	13,800	8,400	10,500	0	2,600
DC(5)	8,400	16,300	16,100	2,600	0

Figure 4. Distances in matrix

For an instance by assumption if DF is initiated to store on cloud storage from North America, then that DF is stored at the nearest DC which is North America (DC(1)) and makes copies of DF such as DF1,DF2. Then arrange the distances from DC(1) to all other Data centers in descending order. That Distances D(1), D(2), D(3),D(4) and D(5) are 14300 (DC(3)), 13800 (DC(4)), 9900 (DC(2)), 8400 (DC(5) and 0 (DC(1)) respectively. Then select the long distanced DC form DC(1) which is Australia(DC(3)(D(1)=14,300)). Then, the DF1 stores at DC(3).

After that, it selects the next long-distanced DC which is Africa (DC(4)(13,800)). After that, it calculates distance 'd' between Africa and Australia. The distance 'd' is 10,500. The distance 'd' is not greater than or equal to D(1)(14,300). So, the next distanced DC will be selected which is DC(2)(9900). Then, find distance 'd' between DC(2) to DC(3). The distance 'd' is 7,500 which is also not is not greater than or equal to D(1)(14,300).

For that, the new DC selection process goes until to reach 's' that is round of (n/2). Because the distances from DC(1) to all other DCs which were already arranged in descending order. So, the next half set of DC's could be close to the first DF residing DC which is DC(1).

Therefore the next step is to reduce the D(1)'s value by 500 kilometers since the reference threshold distance is considered here is 500km. Then the same process is going on until the desired DC is found which is at least half the distance far away from the DC(1). So, in above discussed example, the round of (5/2) is 3. So D(4) and D(2)'s data centers are only involved in the new DC selection process and their distances from DC(3) are 10,500 and 7,500 respectively. So, after eight reductions of D(1) by 500 kilometers, its value is 10,300. Then DC(4) satisfies the condition. So DF2 is stored at DC(4) which is in Africa.

Thus the DF, DF1, and DF2 are stored in North America, Australia and Africa respectively which are far away from one another.

Likewise, the 2RM will be involved to maintain the 2-Replica scheme forever even the DC's destroyed by the natural disaster for ensuring reliability.

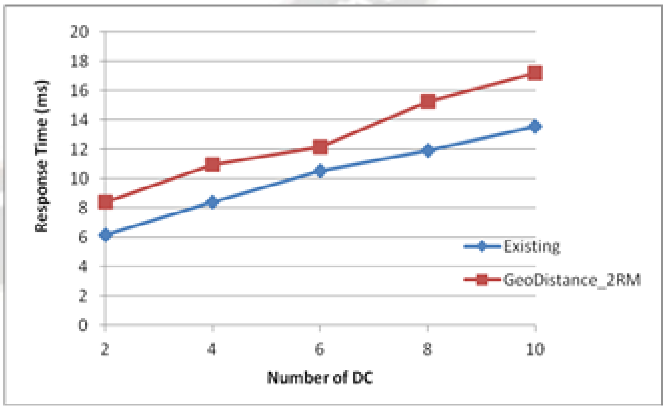


Figure 5. Response time

Figure 5 represents the graphical representation of response times with respect to randomly generated data object requests for both existing method and proposed Geo-distance based 2RM algorithm.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6971
***Article Received**: 30 March 2023 **Revised**: 10 May 2023 **Accepted**: 28 May 2023*

TABLE I COMPARISON OF PROPOSED ALGORITHMS (2RP & 2RM) WITH EXISTING

Comparison Factor	Dcr2s [7]	2rp & 2rm
Size Data Center	Super Data Center Main Data Center Ordinary Data Center	Size May vary for elasticity
Configuration Of Data Centre	Each has a different Configuration	May Different Configuration to Build the Architecture based on Provider's Specifications.
Cost Of The Data Center	(Varies) Super □ High Cost Main□ little bit lesser than Super Ordinary□ low cost	Same cost
Numbers Of Replication	Depends on User's needs, Based on the service offering by the providers and SLA, etc….	Only 2 replications with one Original Data file.
Storage Cost	Vary (Based on SLA and Type of Service)	Common to all (Because 2 Replicas and 1 original total of 3 files only stored)
Maintenance Cost	High(Based on SLA and Type of Service)	Less(Because 2 Replica and 1 original totally 3 files only stored)
Maintenance Overhead	High	Low
Reliability	Doesn't mention strongly	2RM is used to ensure
Available Storage Spaces	Decidable based on the service	Increased
Dynamic Replication (Future Work)	Replica Created based on Popularity and usage requirements.	Temporally Created based on DF Popularity and usage requirements after the usage it will be discarded.

VI. CONCLUSION

The proposed Geo-distance based 2-Replica Maintaining (2RM) algorithm is a successor of the 2-Replica Placing (2RP) Algorithm which has been used to reduce the replication storage cost, and maintenance cost and increase the storage space available for the provider's concern as well because it is used to store only two replications at two different geographical locations that mean the replicas are stored far away from each other. As a result 2RP both the users as well as the providers can be benefited which includes this proposed algorithm not only minimizing the cost but also increases the available storage spaces and reduce maintenance overheads for the service providers. The providers can provide an efficient CSS because this proposed algorithm is proven. But the only thread is recovery mechanism and reliability assurance during the natural disaster that has been solved by the proposed 2RM. Thus, it is proven the actual benefit of the 2-Replica strategy is obtained fully through this proposed 2RM algorithm. In the future, these two algorithms will be implemented and tested in an alpha testing environment and the main key feature of CSS will be assured through the enhancement of 2RP and 2RM with availability and reliability concerns forever even if the DCs destroyed by the natural disaster.

REFERENCES

[1] Mansouri, N., Javidi, M.M. & Zade, B.M.H. "Hierarchical data replication strategy to improve performance in cloud computing", Front. Comput. Sci. 15, 152501 (2021). https://doi.org/10.1007/s11704-019-9099-8.

[2] V. Hadzhiev, "SWOT Analysis of a Hybrid Model for Structuring, Storing and Processing Distributed Data on the Internet," 13th International Conference on Electrical and Electronics Engineering (ELECO), pp. 585-588, 2021 DOI: 10.23919/ELECO54474.2021.9677789

[3] S. Kianpisheh, M. Kargahi and N. M. Charkari, "Resource Availability Prediction in Distributed Systems: An Approach for Modeling Non-Stationary Transition Probabilities," in IEEE Transactions on Parallel and Distributed Systems, vol. 28, no. 8, pp. 2357-2372, 1 Aug. 2017, DOI: 10.1109/TPDS.2017.2659746.

[4] Prof. Sharayu Waghmare. (2012). Vedic Multiplier Implementation for High Speed Factorial Computation. International Journal of New Practices in Management and Engineering, 1(04), 01 - 06. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/8

[5] Mohammad A. Haque, Hakan Aydin, Dakai Zhu, "On Reliability Management of Energy-Aware Real-Time Systems Through Task Replication", IEEE Transactions on Parallel & Distributed Systems, vol. 28, no.3 , pp. 813- 825, 2017. DOI: 10.1109/TPDS.2016.2600595.

[6] S. Annal Ezhil Selvi and Dr. R. Anbuselvi, "Optimizing the Storage Space and Cost with Reliability Assurance by Replica Reduction on Cloud Storage System", International Journal of Advanced Research in Computer Science (IJARCS), ISSN: 2394-3785,Vol. 8, No. 8, pp. 327-333,2017 (ICI). DOI: https://doi.org/10.26483/ijarcs.v8i8.4708

[7] Dhablia, A. (2021). Integrated Sentimental Analysis with Machine Learning Model to Evaluate the Review of Viewers. Machine Learning Applications in Engineering Education and Management, 1(2), 07–12. Retrieved from http://yashikajournals.com/index.php/mlaeem/article/view/12

[8] 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, 2017. DOI: 10.1109/TCC.2017.2659728.

[9] Diwan , S. A. . (2023). Implementation Patterns of Natural Language Processing Using Pre-Trained Deep Learning Models. International Journal of Intelligent Systems and Applications in Engineering, 11(1), 33–38. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2441.

[10] Aral and T. Ovatman, "A Decentralized Replica Placement Algorithm for Edge Computing," in IEEE Transactions on

Network and Service Management, vol. 15, no. 2, pp. 516-529, June 2018, DOI: 10.1109/TNSM.2017.2788945.

[11] Changsong Liu, "A novel replica placement algorithm for minimising communication cost in distributed storage platform," International Journal of Networking and Virtual Organisations, Inderscience Enterprises Ltd, vol. 22(2), pages147161, 2020 .https://doi.org/10.1504/IJNVO.2020.105514.

[12] Gyawali, M. Y. P. ., Angurala, D. M. ., & Bala, D. M. . (2020). Cloud Blockchain Based Data Sharing by Secure Key Cryptographic Techniques with Internet of Things. Research Journal of Computer Systems and Engineering, 1(2), 07:12. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/5

[13] Zhen Huang, Jinhang Chen, Yisong Lin, Pengfei You and Yuxing Peng, "Minimizing Data Redundancy for High Reliable Cloud Storage Systems", Published by ELSEVIER (COMPNW 5513), No. of Pages 14, 2015. https://doi.org/10.1016/j.comnet.2015.02.013.

[14] 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. DOI: 10.1109/TC.2015.2451644.

[15] Abdenour Lazeb, Université Oran1, Ahmed Ben Bella, Oran, Algeria, "Towards a New Data Replication Management in Cloud Systems", International Journal of Strategic Information Technology and Applications, Volume 10 , Issue 2 , 2019. DOI: 10.4018/IJSITA.2019040101

[16] 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),ISSN:0976-5697, Vol. 4, No 9,pp. 29-33,2017 (UGC Approved). DOI: 10.20247/IJARTET.2017.0409007.

[17] Dr. Daya Gupta , Devika Singh ,"User Preference Based Page Ranking Algorithm ", published by IEEE International conference on Computing, Communication and Automation ,2017. DOI: 10.1109/CCAA.2016.7813711.

[18] S. Annal Ezhil Selvi and R. Anbuselvi, "Popularity (Hit Rate) Based Replica Creation for Enhancing the Availability in Cloud Storage", International Journal of Intelligent Engineering and systems (IJIES), ISSN: 2185-3118, Vol.11, No.2, pp.161-172, 2018 (Scopus). DOI: 10.22266/ijies2018.0430.18

[19] Navneet Kaur Gill and Sarbjeet Singh, "Dynamic Cost-Aware Rereplication and Rebalancing Strategy in Cloud System", © Springer International Publishing Switzerland 2015. DOI: 10.1007/978-3-319-12012-6_5.

[20] Anna, G., Hernandez, M., García, M., Fernández, M., & González, M. Optimizing Course Recommendations for Engineering Students Using Machine Learning. Kuwait Journal of Machine Learning, 1(1). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/104

[21] Chunlin Li, Mingyang Song, Min Zhang, Youlong Luo, "Effective replica management for improving reliability and availability in edge-cloud computing environment", Journal of Parallel and Distributed Computing, Volume 143, Pages 107-128, 2020, ISSN 0743-7315, https://doi.org/10.1016/j.jpdc.2020.04.012.

[22] Selvi, S. A. E., & Anbuselvi, R. (2018). Popularity (hit rate) based replica creation for enhancing the availability in cloud storage. Int J Intell Eng Syst, 11(2), 161-72.

[23] Ali Ahmed, Machine Learning in Healthcare: Applications and Challenges , Machine Learning Applications Conference Proceedings, Vol 1 2021.

[24] Selvi, S. A. E., & Anbuselvi, R. (2015, March). An Analysis of Data Replication Issues and Strategies on Cloud Storage System. In International Journal of Engineering Research & Technology (IJERT), NCICN-2015 Conference Proceedings, pp18-21.