International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

# Enhanced Naked Mole-Rat Optimization Based Data Replication in Cloud Computing

[1]Navneet Vishnoi-I, [2]Vikram Singh, [3]Sachin Sharma, [4]Pawan Kumar

**Submitted**: 22/08/2023 **Revised**: 09/10/2023 **Accepted**: 20/10/2023

**Abstract:** Over the last several years, there has been a significant increase in the study of cloud computing, including replication systems and their uses. As the number of replicas grows and they distribute out, the cost of maintaining the system data availability, performance, and consistency rises as well. This research optimizes the choice and placement of data replication on the cloud by an intelligence optimization algorithm with several goals. In this we proposed Enhanced Naked Mole-Rat Optimization (ENMRO). The most common data replication approach is utilized to identify the best-chosen data replicate first. Then it is used to determine the optimal location for a data copy based on the shortest space, the quantity of data transfers, and the accessibility of data replication. The recommended technique has been put through a simulation utilizing cloudsim. The Cloud is intended to mimic many kinds of data centers (DCs) with different architectures. Every DC is made up of a host that hosts a collection of virtual machines (VMs) that provide replications of accessible data blocks. Several well-known techniques, including Replica Selection and Placement (RSP), Genetic Algorithm (GA), Dynamic Cost aware Re-replication and Re-balancing Strategy (DCR2S), Dynamic Replica Selection Ant Colony Optimization (DRSACO) were used to compare the accomplishment of the suggested approach. The investigational findings demonstrate that ENMRO replicates data more effectively than comparative algorithms. In comparison to other algorithms, it also delivers greater data availability, reduced costs, and lower bandwidth usage.

***Keywords:*** *Data replication, cloud computing, data centers, Enhanced Naked Mole-Rat Optimization*

## 1. Introduction

Cloud computing allows for the rapid creation and scalability of a wide variety of useful services, such as on-demand virtual computers, network resources, and services. The Cloud is characterized by features such as scalability, data persistence, and accessibility [1]. Infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS) are all components of a cloud environment [2]. Additionally, the Cloud is more affordable than other conventional systems and offers subscription on-demand scalability, flexibility, and dynamicity [3]. Load balancing has a significant impact on the Cloud network's data transformation rate, throughput, and minimal overload [4]. It divides up jobs inside a VM to stabilize the loads in the cloud to maintain minimal overloads and gain the best possible access to the available resources [5].The majority of the time, data intensive apps like distribution from far-off places to nearby locations via nodes or geographic sites employ replication methods [6].In a cloud context, a replication model may be utilized as a cluster. The cluster enables scalability and data accessibility to guarantee the consistency and integrity of various replications across nodes. It also entails writing and reading about data replication using protocols by system necessities [7].

This research proposes an effective ENMRO-based dynamic data replication mechanism. The chosen and positioned data replicates in various data centers form the foundation of the suggested technique. It assesses data replication using seven factors, including the Zipf and geometric distributions as well as data replication access, distance, prices, and data availability. As a result, by using our expertise, we can strategically deploy copies of your data on the cloud. Utilizing CloudSim, the suggested model is verified, and the optimization ENMRO is assessed. The experimental findings demonstrate that the suggested techniques exceed the other options in terms of time savings, access speed acceleration, cost reduction, high availability, and locating the least expensive route across diverse Cloud datacenters.

## 2. Related works

The research [8] investigated the process of data replication on the cloud. Unlike other methods in the literature, the subject was on the system's energy efficiency and bandwidth requirements simultaneously. The purpose

[1]*Assistant Professor, College of Computing Science and Information Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India, Email id: vishnoi_navneet@yahoo.co.in*
[2]*Assistant professor, School of Computer Science & System, JAIPUR NAITONAL UNIVERSITY, JAIPUR, India, Email Id: vikramsingh@jnujaipur.ac.in*
[3]*Assistant Professor, School of Engineering and Computer, Dev Bhoomi Uttarakhand University, Uttarakhand, India, Email Id: socse.sachin@dbuu.ac.in*
[4]*Assistant Professor, Department of Computer Science and IT, Jain(Deemed-to-be University), Bangalore-27, India, Email Id: pawan.bhuphd@gmail.com*

of the article was to provide a dynamic data replication model along with a concise assessment of replication strategies that are appropriate for use in distributed computing settings [9]. The study [10] provided a thorough analysis and categorization of contemporary data replication schemes across several obtainable cloud setting systems in the form of categorization to characterize the present scheme on the subject and discuss unresolved challenges. The research [11] suggested estimates for the power requirements and data transfer rates associated with accessing databases in the cloud. To further increase the Quality of Service (QoS) while decreasing communication delays, they also suggested an energy-efficient replication technique based on the current models.

The work [12] provided a tabular presentation of all the elements of data replication and analyzes the main points of the most current techniques. The research [13] provided a CSO-based method to secure data replication (SDR), whereby a new replica's location is chosen based on four factors: centrality, energy use, storage utilization, and load. The goal of the work was to offer a comprehensive overview of these two categories of data replication methods and to examine the key characteristics of each category [14]. The research [15] presented a method of data replication that uses an adaptive selection of data files to replicate to enhance the accessibility of the system. Additionally, the suggested technique decides on the effective data nodes for replication as well as the number of copies in a dynamic manner. The article [16] explored an obstetric imaging diagnosis platform based on cloud computing technologies in light of that. First, a distributed file system, cloud computing technology, and caching technology were combined to create a medical imaging platform. Second, contrast-enhanced ultrasound technology suggests a more precise ultrasound picture for determining the shape, size, position, and placental developmental anomalies. The work [17] suggested a data replication technique that strikes a balance between the tenant's performance assurances and the cloud provider's bottom line. The study [18] provided a comprehensive overview of cloud-based data replication methods. They suggest a new categorization of replication techniques along the following five dimensions, based on the survey and review of current categorization: Distinguish between static, dynamic, providers, customer-centric, an optimum number of replicas, and goal function.

## 3. Methododolgy

This section discusses the structural model that has been proposed for choosing and positioning replications in cloud computing to share data across nodes. In our situation, we optimized the choice and location of data replications using heterogeneous system and swarm intelligence (SI) methodologies. The initial level comprises the highest data centers, which are most tightly centralized, and have superior data accessibility, storage capacity, and evaluation, as well as more hosts and virtual machines (VMs) than the other data centers. Middle data centers, which contain fewer entire elements than the initial level, make up the second level. The lowest data centers with fewer complete components than those in the second stage make up the third level. Overall, there are hierarchical connections between the data centers, whether on a similar stage or at higher levels. The suggested architecture comprises a broker, data centers, a replication catalog, a replica, a replication managing system, hosts, virtual machines, tasks, and users organized in a hierarchical network structure. Selecting and placing replications across nodes may be optimized with the help of ENMRO to guarantee that the most effective technique with replication may be employed in the most beneficial proposed system. Applying the ENMRO approach enhances availability and support, lowers costs, enables job execution, and is faster than the methods used when choosing and positioning replications via CloudSim. As a result, data centers enable us to achieve optimum replication and maximum utilization. The number of distinct data centers in the cloud environment, n, may be used to represent the data centers in the form of $DCs = \{dc1, dc2 \ldots dcn\}$. The physical machine may be expressed as $PM = \{pm1, pm2 \ldots pmx\}$, where $pmx$ denotes a collection of several PMs found in DCs. We provide our unique solution for diverse cloud data centers. The virtual machines are denoted by the notation $VM = \{vm1, vm2 \ldots vms\}$, where $s$ is the total value of VMs that are housed in each $PM$. There are several functions, like time and space shards, and various spaces. The data replication may be articulated as $F = \{f1, f2 \ldots fy\}$, where $y$ denotes the entire number of possible data replication that can be housed within DCs. The primary storage space unit is denied and is denoted by the formula $B = \{b1, b2 \ldots, by\}$, by, where y is a collection of various data replications kept in DCs. In the described approach, all replication files are dispersed randomly among high data centers. Using the replica catalog, which keeps track of every replica position in the various DCs, distinct probabilities of the form possibility = pro (bap) may be recorded in every data center.

### 3.1. Availability of data files

Generally, "readiness to offer appropriate services" may be used to describe availability. All users should have complete access to the data replications upon request. As a result, users' access to data on the Cloud is a key concern. A failure in the Cloud's data replication or node causes a system with unavailable data. In the cloud, data accessibility is a significant concern. To guarantee the availability of various blocks in the DCs, the replicas may be spread throughout various data centers and multiple replications can be housed in a single DC. Because high DCs are more expensive and provide greater data availability and dependability, the blocks within them are also more available. The blocks within low DCs are likely to have poor availability, and low DCs have lower prices, less availability, and worse dependability. They may be determined as follows by using Equations (1) through (4):

A general description of a genetic algorithms operation is given below:

$$pro(bab_i)_{hig}DC > pro(bab_i)_{mid\ DC} > pro(bab_i)_{low\ DC} \quad (1)$$

$$pro(fla_l) = \begin{cases} (1-\prod_{j=1}^{bnr_l}(1-pro(bab_i)j))^{nbk} \ for\ case\ 1 \\ \prod_{j=1}^{bnr_l}(1-\prod_{j=1}^{bnr_l}(1-pro(bap_i)j) \ \ for\ case\ 2 \end{cases} \quad (2)$$

$$\overline{pro(fla_l)} = \begin{cases} 1-(1-\prod_{j=1}^{bnr_l}(1-pro(bab_i)j))^{nbk} \ for\ case\ 1 \\ 1-\prod_{j=1}^{bnr_l}(1-\prod_{j=1}^{bnr_l}(1-pro(bap_i)j) \ \ for\ case\ 2 \end{cases} \quad (3)$$

$$high_{dc} = 0.9 > mid_{dc} = 0.6 > low_{dc} = 0.3 \quad (4)$$

Where, B blocks $pro(fla_l)$ Probability of file availability $nbk$ -Number of block $bnr_l$-Number of replica of a data file Block unavailability probability of block $nbk$ -Number of access task have request $hig^{DC}$ - High Datacenters $hig^{DC}$ - Middle Data centers $low^{DC}$-low Data centers

### 3.2. Costs of replication in datacenters

The velocities, costs, reliabilities, availability probabilities, and performance in the cloud are all included in the set of DC costs that differentiate the lowest, middle, and highest data centers. Costs associated with DCs have a significant role in selecting and positioning copies via DCs according to MOO criteria. To prevent replications via DCs and assurance that the budget is enough for users, the overall replication costs must be maintained as low as feasible.

$$cost_l(dc_t) = \sum_{y=1}^{x}(cost(cd_x).amq_l(cd_x)) \quad (5)$$

### 3.3. The shortest distance between data centers

The following equations may be used to determine the access to data replications, which is used to determine the distance between DCs (6) and (7). The link among those two DCs (I and j) is the ideal option. By demonstrating that this formulation guarantees them no passing in an unending ring, this formula shows that $c_{ij}$and $c_{ji}$are passed.

$$Min\ \sum_{j=1}^{m}\sum_{i=1}^{m} c_{ji}y_{ji} \quad (6)$$

$$\sum_{i=1}^{m} m_j y_{j\geq l.}\, y_j \in \{0,1\}(1 \leq j \leq m) \quad (7)$$

### 3.4. Difficulty with the knapsack

The knapsack issue is really difficult. Every object has a weight and a value. By utilizing the following Equation (8) and (9), it is possible to ensure that the budget for users is enough while still minimizing costs in the Cloud:

$$maximize\ oy = \sum_{i=1}^{m} o_i y_i \quad (8)$$

$$uy = \sum_{i=1}^{m} u_{ji} y_i \leq u_j \quad (9)$$

### 3.5. Enhanced Naked Mole-Rat Optimization (ENMRO)

Naked Mole-Rat Optimization (NMRO), which is straightforward and linear, has recently drawn attention. The revised and improved Naked Mole-Rat Optimization (ENMRO) intends to increase effectiveness by strengthening its core exploitation and exploration capabilities. By including GWO-inspired equations into the fundamental NMRA, exploitation is enriched and Local Neighborhood Search (LNS) and DE formulations are employed to further increase exploitation. Following is a discussion on the idea of LNS.

This search mode uses just the nearby physical and digital features. NMRO uses the existing solution in addition to contextual knowledge in the worker phase to expand the search area. The search capabilities are improved with faster convergence thanks to the use of the LNS paradigm.

$Y = (Y_1, Y_2,, Y_3,, \dots, Y_m,), Y_j, (j \in [1, \dots, m])$ is a vector (D-dimensional). For any vector $Y_j$, the neighborhood radius r (where 2r + 1 < n) is specified; hence, the neighborhood of $Y_j$ consists of $Y_j, -q, \dots, Y_j, \dots, Y_j, +q$. Based on their indices, the vectors in this instance are organized in a ring topology. Figure 1 depicts the LNS model, which is characterized as follows:

$$K_j = Y_j + n * \left(Y_{n_{opt}} - y_j\right) + m * (Y_o - Y_r) \quad (10)$$

Where $Y_{n_{opt}}$ is the optimum solution so far in the $Y_j i$; $n, m \in rand\ ()$ are the scaling factors and $o, r \in [j - q, j + q]$ $(o \neq r \neq j)$.Equation (10), which is used in the enhanced NMRA, modifies the optimum solution, which then performs the worker phase.

$$y_j^{s+1} = K_j^s + q * (y_l^s - y_n^s) \quad (11)$$

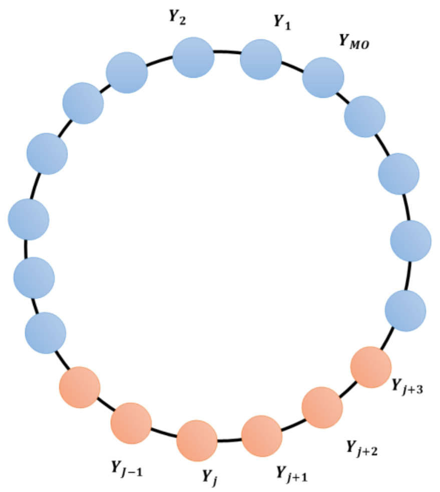


**Fig.1.** Topology of 3-ring neighborhoods.

To enhance the efficiency of regular NMRO, the worker phase is updated in ENMRO. Here, we do half of the entire number of iterations. In this first part, a new investigative equation is added. The average of the first three solutions derived from the optimal solution provides evidence for the following equation:

$$u_1 = u_j - B_1(D_1.c - u_j) u_2 = u_j - B_2(D_2.c - u_j), u_3 = u_j - B_3(D_3.c - u_j) \quad (12) \qquad u_{new} = \frac{u_1+u_2+u_3}{3} \quad (13)$$

Grey wolf optimization (GWO) search equations inspired this modification. At GWO, the consequence is constructed by adjusting the variables until the search agents are at optimal positions. With the use of GWO search equations, the method can overcome the poor exploration of traditional NMRO by increasing the unpredictability of the search equations used during the worker phase. The same definition is used to arrive at a different answer:

$$u_j^{s+1} = u_{new}^s + \lambda(u_i^s - u_l^s) \quad (14)$$

Where $B1, B2, B3$ and $D1, D2, D3$ belong to B and D, respectively, and are given as:

$$B = 2bq_1 - b, D = 2.q_3 \quad (15)$$

Where $q1$ and $q2$ are uniformly distributed random integers and $b$ is a random value that decreases linearly from [0, 2] as a function of the number of rounds. The NMRO's ability to explore is improved since the agents may move about inside the search space at any point thanks to the random number.

In this step, the resident is divided into three portions, and the last solution is assessed utilizing three different search algorithms to enhance the exploitation effectiveness of NMRO. For the first segment of the population, the following general equation, which is comparable to the conventional NMRO, is used:

$$a_j^{s+1} = \begin{cases} a_{new}^s + \lambda(a_i^s - a_l^s), if\ 0 \geq s \geq s_{max}/2 \\ a_j^s + \lambda(a_i^s - a_l^d), if\ \frac{s_{max}}{2} + 1 \geq s \geq s_{max} \end{cases} \quad (16)$$

The following LNS-based search equation, which takes its cue from Equation (10), is utilized for the 2nd segment of the population:

$$a_j^{s+1} = \begin{cases} K_{new}^s + \lambda(a_i^s - a_l^s), if\ 0 \geq s \geq s_{max}/2 \\ K_j^s + \lambda(a_i^s - a_l^d), if\ \frac{s_{max}}{2} + 1 \geq s \geq s_{max} \end{cases} \quad (17)$$

Here $K_j^s$ the better solution adapted using LNS and λ is a scale factor that is created evenly [0, 1]. Here$K_{new} = a_{new} + n * (a_{n_{opt}} - a_{new}) + m * (a_o - a_r)$ using formulation (4). The search equation based on DE is used for the third segment of the population and is provided by

$$a_j^{s+1} = \begin{cases} a_{new}^s + \lambda((a_i^s - a_l^s), +((a_o^s - a_r^s))\ if\ 0 \geq s \geq s_{max}/2 \\ a_j^s + \lambda((a_i^s - a_l^d) + ((a_o^s - a_r^s)), if\ \frac{s_{max}}{2} + 1 \geq s \geq s_{max} \end{cases} \quad (18)$$

Additionally, in the fundamental NMRO breeder phase, the search positions of the new breeders are just population update operations by the best-estimated NMRO. If the internationally optimum individuals merge into the local optima, the whole algorithm suddenly experiences premature convergence. Therefore, minimizing the likelihood that the algorithm enters the local optimum serves as the directing function of the better persons who may be located close to the optimal solution. In this study, the present optimum people are replaced with random persons located close to the best answer to drive the algorithm search and increase the likelihood that the algorithm would deviate from the local optima. The following is the most recent method for finding the top person in ENMRO's neighborhood:

$$a_j^{s+1} = (1-\lambda)a_j^s + \lambda(c(1 + \mu.unifrnd(-1,1)) - a_j^s) \quad (19)$$

Where $unifrnd\,(-1,1)\,(-1,1)$ is the uniformly circulated variable inside of (-1, 1) and is the disturbance coefficient. The method searches between segments $C(1+)\,(1+)\ and\ d(1+)\,(1+)$, using the current best answer as the middle and as the step size in the area search for the ideal person. In essence, this broadens the search space and increases the likelihood that the algorithm will depart from the local ideal. The following is how Algorithm 2 displays the ENMRA pseudo code:

Algorithm 2. Pseudocode of Enhanced NMRA

Input: Explain the function of objectives. ((NMR), $\mathrm{NMR} = NMR_1, NMR_2, \dots, NMR_C)$

Output: Find the current optimal solution c

Start: Start $NMR_t: m, C: \frac{m}{5}, U: C - m$

The Probability of a Breeding: co

While iteration $< s_{max}/2$

For $j = 1$: workers

Perform worker phase;

If $s \geq s_{max}/2$

$$u_1 = u_j - B_1(D_1.c - u_j) u_2 = u_j - B_2(D_2.c - u_j), u_3 = u_j - B_3(D_3.c - u_j)$$

$$u_{new} = \frac{u_1 + u_2 + u_3}{3}$$

$$u_j^{s+1} = u_{new}^s + \lambda(u_i^s - u_l^s)$$

Else

$$u_j^{s+1} = u_{new}^s + \lambda(u_i^s - u_l^s)$$

Evaluate $u_j^{s+1}$

End for

For $i = 1: breeders$

If rand>$aq$

To carry out the breeder phase, the population is split into three equal subsets.

$$a_j^{s+1} = \begin{cases} a_{new}^s + \lambda(a_i^s - a_l^s), if\ 0 \geq s \geq s_{max}/2 \\ a_j^s + \lambda(a_i^s - a_l^d), if\ \frac{s_{max}}{2} + 1 \geq s \geq s_{max} \end{cases}$$

For the 2nd part of population:

$$a_j^{s+1} = \begin{cases} K_{new}^s + \lambda(a_i^s - a_l^s), if\ 0 \geq s \geq s_{max}/2 \\ K_j^s + \lambda(a_i^s - a_l^d), if\ \frac{s_{max}}{2} + 1 \geq s \geq s_{max} \end{cases}$$

For the 3rd part of the population

$$a_j^{s+1} = \begin{cases} a_{new}^s + \lambda((a_i^s - a_l^s), +((a_o^s - a_r^s))\ if\ 0 \geq s \geq s_{max}/2 \\ a_j^s + \lambda((a_i^s - a_l^d) + ((a_o^s - a_r^s)), if\ \frac{s_{max}}{2} + 1 \geq s \geq s_{max} \end{cases}$$

End if

calculate $a_j^{s+1}$

End for

Enhancing policy for the neighborhood search

$$a_j^{s+1} = (1 - \lambda) a_j^s + \lambda(c(1 + \mu.unifrnd(-1,1)) - a_j^s)$$

Bring the current workforce and breeding pool together.

Calculate the population

Enhance the overall best c

Enhance iteration count

End while

Update final bestd

End

## 4. Result and discussion

This study examines the dynamic selection and placement of replicas in the cloud carried out using CloudSim. It employs the ENMRO algorithm and validates the experimental findings by comparing them with those of other algorithms. In Figure 2 and Table 1, we can see how the replication costs change as the number of jobs performed by each user does. When compared to the DCR2S and EFS, our ENMRO technique results in lower replication costs. Changing the particle count and the number of iterations in the ENMRO algorithm has been proven to affect the selection of replicas. When compared to the DCR2S, whose replication costs are within budget, the EFS has prohibitively high replication costs. There are three problems: user waiting time, optimized replicas, and fixed costs. Therefore, ENMRO helps minimize not only the cost of replication but also the availability of data. Finally, cloudlets are used to optimize the selection of replicas.

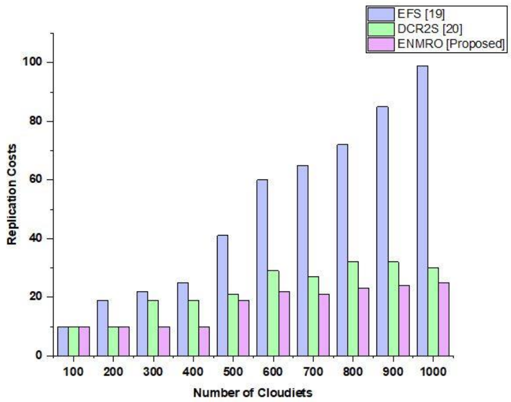


**Fig.2**. Costs of replication and the quantity of cloudlets

**Table 1.** Comparison of Energy Consumption

| Number of Cloudiets | Replication Costs | | |
|---|---|---|---|
| | EFS [19] | DCR2S [20] | ENMRO [Proposed] |
| 100 | 10 | 10 | 10 |
| 200 | 19 | 10 | 10 |
| 300 | 22 | 19 | 10 |

| | | | |
|---|---|---|---|
| 400 | 25 | 19 | 10 |
| 500 | 41 | 21 | 19 |
| 600 | 60 | 29 | 22 |
| 700 | 65 | 27 | 21 |
| 800 | 72 | 32 | 23 |
| 900 | 85 | 32 | 24 |
| 1000 | 99 | 30 | 25 |

Using the ENMRO method to choose the best replicas, the investigational findings are shown in Figure 3 and Table 2. Here we compared RSP and DRSACO with our suggested method ENMRO. As findings of these results, it is now clear that our technique has a shorter runtime requirement for gaining access to ideal copies than competing algorithms.

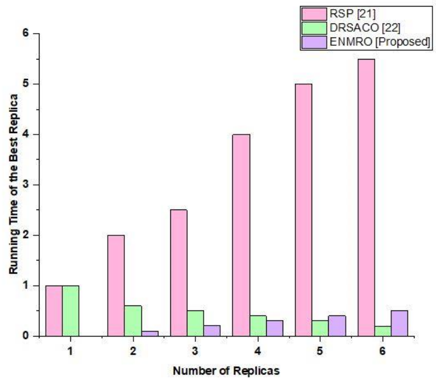


**Fig.3.** Probability of data availability and the number of replicas

**Table 2.** Comparison of retrofit cost

| Number of Replicas | Running Time of the Best Replica | | |
|---|---|---|---|
| | RSP [21] | DRSACO [22] | ENMRO [Proposed] |
| 1 | 1 | 1 | 0 |
| 2 | 2 | 0.6 | 0.1 |
| 3 | 2.5 | 0.5 | 0.2 |
| 4 | 4 | 0.4 | 0.3 |
| 5 | 5 | 0.3 | 0.4 |
| 6 | 5.5 | 0.2 | 0.5 |

The number of replications across DCs and the data transfer rate that saves time and money are determined using ENMRO in Figure 4. The results demonstrate that our approach is better than other algorithms like ACO and GA. When our method's findings are compared to those of other algorithms, it becomes clear that our technique outperforms others in terms of the time and money required to optimize the placement of replicas.

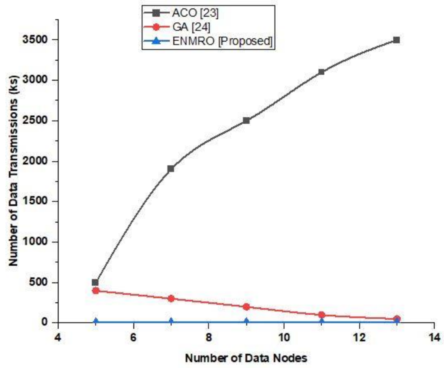


**Fig.4**. The data transfers between several data nodes

**Table 3:** various data nodes transmit data in various ways.

| Number of Data Nodes | Number of Data Transmissions (ks) | | |
|---|---|---|---|
| | ACO [23] | GA [24] | ENMRO [Proposed] |
| 5 | 500 | 400 | 8.2 |
| 7 | 1900 | 300 | 8.3 |
| 9 | 2500 | 200 | 8.4 |
| 11 | 3100 | 100 | 8.5 |
| 13 | 3500 | 50 | 8.6 |

## 5. Conclusion

One of the most useful settings for maximizing replica performance and achieving high availability is the cloud. ENMRO was introduced in this study to facilitate the variable duplication and relocation of data. The best option among the most popular data replications is determined by ENMRO. After determining the optimal locations for data copies using ENMRO, these locations are then optimized to be as close to end users as possible. CloudSim was used to build and implement the recommended system architecture. The effectiveness of the proposed model was assessed in comparison to other replication algorithms, including RD, RSP, GA, DCR2S, and DRSACO. The outcomes of the simulation demonstrated that the suggested algorithms were superior to and more efficient than the comparing methods. Future research will evaluate the recommended system design in a real computer environment. The knapsack issue will also be addressed to reduce expenses and maximize the speed and efficiency of data replications over the cloud.

## References


[1] Salem, R., Salam, M.A., Abdelkader, H. and Mohamed, A.A., 2019. An artificial bee colony algorithm for data replication optimization in cloud environments. IEEE Access, 8, pp.51841-51852.

[2] Calheiros, R.N., Ranjan, R., Beloglazov, A., De Rose, C.A. and Buyya, R., 2011. CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. Software: Practice and Experience, 41(1), pp.23-50.

[3] Rajeshirke, N., Sawant, R., Sawant, S. and Shaikh, H., 2017. Load balancing in cloud computing. Int J Recent Trends Eng Res, 3(3), pp.260-267.

[4] Haricha, K., Khiat, A., Issaoui, Y., Bahnasse, A. and Ouajji, H., 2023. Recent technological progress to empower Smart Manufacturing: Review and Potential Guidelines. IEEE Access.

[5] Ahn, H.Y., Lee, K.H. and Lee, Y.J., 2016. Dynamic erasure coding decision for modern block-oriented distributed storage systems. The Journal of Supercomputing, 72, pp.1312-1341.

[6] Maheshwari, N., Nanduri, R. and Varma, V., 2012. Dynamic energy efficient data placement and cluster reconfiguration algorithm for MapReduce framework. Future Generation Computer Systems, 28(1), pp.119-127.

[7] Milani, B.A. and Navimipour, N.J., 2016. A comprehensive review of the data replication techniques in the cloud environments: Major trends and future directions. Journal of Network and Computer Applications, 64, pp.229-238.

[8] Boru, D., Kliazovich, D., Granelli, F., Bouvry, P. and Zomaya, A.Y., 2015. Energy-efficient data replication in cloud computing data centers. Cluster computing, 18, pp.385-402.

[9] Sun, D.W., Chang, G.R., Gao, S., Jin, L.Z. and Wang, X.W., 2012. Modeling a dynamic data replication strategy to increase system availability in cloud computing environments. Journal of computer science and Technology, 27(2), pp.256-272.

[10] Shakarami, A., Ghobaei-Arani, M., Shahidinejad, A., Masdari, M. and Shakarami, H., 2021. Data replication schemes in cloud computing: a survey. Cluster Computing, 24, pp.2545-2579.

[11] Boru, D., Kliazovich, D., Granelli, F., Bouvry, P. and Zomaya, A.Y., 2015, June. Models for efficient data replication in cloud computing data centers. In 2015 IEEE international conference on Communications (ICC) (pp. 6056-6061). IEEE.

[12] Mansouri, N. and Javidi, M.M., 2020. A review of data replication based on a meta-heuristics approach in cloud computing and data grid. Soft computing, 24, pp.14503-14530.

[13] Mansouri, N., Javidi, M.M. and Mohammad Hasani Zade, B., 2021. A CSO-based approach for secure data replication in the cloud computing environment. The Journal of Supercomputing, 77, pp.5882-5933.

[14] Milani, B.A. and Navimipour, N.J., 2017. A systematic literature review of the data replication techniques in the cloud environments. Big Data Research, 10, pp.1-7.

[15] Hussein, M.K. and Mousa, M.H., 2012. A lightweight data replication for cloud data centers environment. International Journal of Engineering and Innovative Technology, 1(6), pp.169-175.

[16] Lie, W., Jiang, B. and Zhao, W., 2020. Obstetric imaging diagnostic platform based on cloud computing technology under the background of smart medical big data and deep learning. IEEE Access, 8, pp.78265-78278.

[17] Tos, Uras, Riad Mokadem, Abdelkader Hameurlain, Tolga Ayav, and Sebnem Bora. "A performance and profit-oriented data replication strategy for cloud systems." In 2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), pp. 780-787. IEEE, 2016.

[18] Tabet, K., Mokadem, R., Laouar, M.R. and Eom, S., 2017. Data replication in cloud systems: a survey. International Journal of Information Systems and Social Change (IJISSC), 8(3), pp.17-33.

[19] M. Bsoul, A. Al-Khasawneh, E. Abdallah, and Y. fKilani, “Enhanced Fast Spread Replication strategy for Data Grid”, Journal of Network and Computer Applications, Vol. 34, pp. 575–580, 2011.

[20] N. Gill and S. Singh, “A dynamic, cost-aware, optimized data replication strategy for heterogeneous cloud data centers”, Future Generation Computer Systems, Vol. 65, pp. 10– 32, 2016.

[21] Shao, Z.L., Huang, C. and Li, H., 2021. Replica selection and placement techniques on the IoT and edge computing: a deep study. Wireless

Networks, 27(7), pp.5039-5055.

[22] L. Wang, J. Luo, J. Shen, and F. Dong, "Cost and time aware ant colony algorithm for data replica in alpha magnetic spectrometer experiment", In Proc. of International Conf. on Big Data, Santa Clara, USA, Vol. 1, pp. 247-254,2013.

[23] N. Navimipour and B. Milani, "Replica selection in the cloud environments using an ant colony algorithm", In: Proc. of International Conf. on Control Engineering and Communication Technology, Moscow, Russia, pp. 1-9, 2016.

[24] L. Cui, J. Zhang, L. Yue, Y. Shi, H. Li, and D. Yuan, "A Genetic Algorithm Based Data Replica Placement Strategy for Scientific Applications in Clouds", IEEE Transaction on Services Computing, Vol. 11, pp. 727-739, 2018.

[25] Jadhav, S. B. ., & Kodavade, D. V. . (2023). Enhancing Flight Delay Prediction through Feature Engineering in Machine Learning Classifiers: A Real Time Data Streams Case Study. International Journal on Recent and Innovation Trends in Computing and Communication, 11(2s), 212–218. https://doi.org/10.17762/ijritcc.v11i2s.6064

[26] Moore, B., Clark, R., Martinez, J., Rodriguez, A., & Rodriguez, L. Anomaly Detection in Internet of Things (IoT) Data Streams. Kuwait Journal of Machine Learning, 1(4). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/151