A Harmonized Dynamic Quality of Service Aware Data Replication Strategy in Cloud

Chandrakala H L, Loganathan R.

Abstract*: Cloud computing is that on-demand service in which the computing resources and the information technology like the storage, operating systems, databases, networks, hardware and databases for the whole software of applications have been delivered. Problem Statement: In case of data clouds, there are different algorithms of application which is needed for the various replication algorithms that have been seeking attention recently. This particular distinguishing feature of the algorithm that is simplicity and efficiency of search has been present in case of Harmony Search. Method: The main idea of a replication will be to be able to provide the data replication in several other locations. The Data access will be enhanced using the Data Replication Strategy (DRS). And the decision of where and what the replicate will be NP complete. The work further involves the problem of data replication which has been addressed in cloud using the Harmony search. Result: The harmonized replication performed better in terms of produce bandwidth and saving by about 12.5%. Conclusion: The storage effectiveness has been improved by means of data deduplication and for the experiments, the static, the adaptive and the methods of harmonized replication have been used.

Index Terms*: Cloud computing, Data Replication Strategy (DRS), Dynamic Data Replication problem and Harmony Search (HS).

I. INTRODUCTION

With the advent of the Cloud Computing (CC), this computing as its utility is now a dream come true and this has also the new age future of computing with a large part of the industry of IT is transformed and also has reshaped the purchase of using the IT software as well as hardware [1]. The Cloud computing may be alternatively defined at that large-scale and distributed computer paradigm driven using economies of scale that is present in the abstracted, scalable, available, virtualized and configurable or reconfigurable resources of computing (like that of the servers, the storage, the applications, the data and the networks) that are rapidly provisioned and has been released using minimal efforts for managing the data centers. For the external customers, such services have been provided based on the on-demand and the help of the high-speed internet with the X as the service (XaaS) the computing architecture that has been divided into three different segments: the applications, the infrastructure and the platforms.

The main objective of this for providing the users with some more flexible services in a manner that is transparent having cheaper and more powerful processors. Keeping the computation in view, the computing has been considered as the network of the data centers and has been considered to be energy efficient, low-cost and cost effective approach to its future computing. The large institutions like Google, Amazon and IBM is set up as the data centers and the platforms of cloud computing. Compared to the conventional systems of large scale storage, a focus of the clouds which have been sensitive to the workloads and the user behavior will provide as well as publish the storage service on that of the Internet [2].

The Distributed file systems like the GFS and the HDFS will be the cloud components and in case of the GFS cluster there are identified three different components that will be the multiple clients, the multiple chunk servers as well as the single master server. The files generally have been broken down into various small pieces or some chunks that have been stored in the datacenters that have been administered by that of the chunk servers. All of the metadata of the file systems have been maintained by that of the master server as well as the present location of such chunks. The interaction of such clients with that of the master for the operation of the metadata has been done and for the chunk servers the communication that is data bearing has been directed [3].

The data cloud objective will be to share as well as evaluate both the data as well as the resources of storage that are within a wide network that is dynamic, heterogeneous and distributive. Having a reasonable level of performance, the data has been distributed into the grid which is duly available as well as accessible for several applications. The main focus here is the evaluation of large data and the whole data in the cloud system will have to be enhanced for analyzing the dynamic as well as the real time data. The Access latency is duly reduced by the creation of a replica and a bandwidth consumption. The basis for such algorithms of data replication is the information of historical data access and the meta data. The data replication for a particular node will be in the cloud environment. The data replica has been well suited for a particular node in a different time. This replication for the cloud computing has been shown in Fig 1.

Revised Manuscript Received on September 06, 2019

Chandrakala H L, Associate professor, HKBK College of Engineering, Bangalore

Dr. Loganathan R. Professor and HOD/CSE HKBK College of Engineering, Bangalore

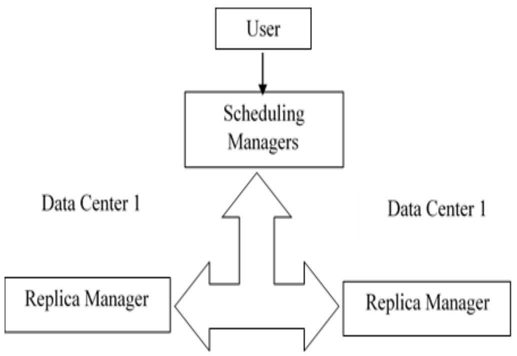

Fig 1 Replication in Cloud Computing

In recent times there are large data volumes that are shared with various resources in different scientific disciplines. The mass of the data will be measured in unit so terabytes and sometimes petabytes in most such fields. Such large volumes are normally kept in the cloud datacenters [3]. Therefore, a great amount of such data will be used for creating some identical copies of the data like the files and databases in the sites which are distributed geographically as replicas.

The process of speeding up of data access will be the benefit of the replication of data that will bring down the latency of access and the availability of increasing data. There is also a common technique that is used for the replicas that are geographically distributed for the dispersed clouds for increasing the response time of the users. The availability of guaranteeing replica and the features of data integrity are very important and they have to be similar to that of the original data with no inference or corruption.

The waiting time of the user is brought down by the decrease in data replication and there is also an increase in data availability by bringing down the consumption of system bandwidth in minimizing the cloud system. A common technique of optimization that was adopted by several conventional systems that include the Database Management Systems (DBMS) of the parallel and distributed systems, the mobile systems and certain other large scale systems that include the P2P data grid systems will be the data replication. In order to increase the availability of data and to enhance the performance and also for achieving fault tolerance are the main advantages of replication of data. In the data application strategy there are three primary questions that are what, when and where the replication has to be answered. In case of the conventional systems, there are several replicas that have been created using several other strategies of replication for achieving maximum performance. The classification of the techniques of data replication are divided into two main groups that include the dynamic and static replication mechanisms. In case of the static replication strategies, there are amy replicas with the host node that is determined even before they are properly defined [6]; and at the same time there are some dynamic strategies that are created in an automatic manner and the replicas are duly removed based on the changes in the access patterns of the users, their bandwidth and storage capacity.

On the basis of the alterations in the users and their access patterns along with their bandwidth, formation as well as the deletion of the replicas that are automatically done relating to the strategies of data replication in the cloud environment [7]. There are several intelligent choices that are made on the location of data that are based on the current environment and its knowledge. But at the same time there are also certain disadvantages like discrepancies in the runtime information of such data nodes in that of a complex cloud infrastructure along with the maintenance of the consistency of the data file. There are some stages that are involved in the strategy of dynamic data replication: for analyzing and modeling this relationship between that of the quantity of the replicas and the availability of the systems for being able to recognize any popular data as well as triggering the process of replication while passing the data for meeting a reasonable system byte based effective rate requirement and also for inserting the replicas among the data nodes in a manner that is balanced. For the purpose of designing this algorithm of a dynamic data replication, in the cloud there is in the strategy of such dynamic replica management a file which has some popular data which has been decided upon by means of a replica creation as well as an appropriate time for creation of a new replica having some popular data. An apt time for the creation of a new replica has been located by the creation of the replica which after an access recorder is duly assigned to each such data node and this is also used for the purpose of simultaneous user access for each such file that includes the name of the file, the size of the fine and the number of concurrent access, its file size and so on [8].

Owing to the large data volume in the cloud computing systems, the chances of failures in hardware are not important that have been based on that of the statistical analysis of the hardware failures. For providing a high availability of data, in case of cloud computing, the technique of data replication has been adopted that will be able to tolerate corruption of data [9].

This QoS requisite for any application is however not taken into consideration for the replication of data. At the time data corruption, the support of the QoS requirement of its application will not take place. The basic features are found to be heterogeneous in the nodes relating to the system of cloud computing. Owing to this in case of a high QoS application, such a data replication in case of a low performance node (a node having slow communication with latencies of the disk) will take place. Furthermore, in case of corruption of data application there is also a retrieval of such data application in a low performance node and this has slow communication and latencies of disc access leading to a violation of the QoS application.

The issue of what and as to where the needs of replication can solve the problem of constraint optimization which is NP complete is found and these issues will stretch the Harmony Search (HS) capacity to the limit. This will be the emerging metaheuristic optimization algorithm which will cope for a number of the challenging tasks in the last decade. As implied by the name the HS will be an algorithm that is music based.

This will have the inspiration which will be through that of the observation where the objective of the music will be to search for a state of harmony that is absolute.

Any optimality in the process of optimization this may be compared to that of

the process of improvisation of a jazz musician. For a problem of optimization, the optimal solution will be the best one that is available for the problem in certain specified objectives which will be able to transform the qualitative improvisations and this process into that of certain quantitative idealization rules and will thereby transform both the beauty and the harmony for the HS algorithm.
The main objective of the study is the problem of replication which is addressed in that of the cloud using the proposed algorithm of harmonized replication. The related literature for this work is discussed in Section 2, the techniques involved along with the methods are expressed in Section 3, the results are analyzed in Section 4 and the conclusion is made in Section 5.

II. RELATED WORK

There is a brief survey on the selection and the replication of the data center selection that was made by Jayalakshmi et al., [10]. For increasing the availability of data in an effective manner decrease the waiting time of the user by a small number of replicas with a new system model having two stages of data center selection and a dynamic data replication that had been proposed. Single point failure is to be addressed. There was also a strategy of data replication that was proposed by Tos et al., [11] for satisfying the guarantee of performance for the tenants with ensuring the profitability of the cloud provider. The queries and their response time will be estimated using this proposed strategy along with the expenditure which will hinder the cloud provider's profitability. By means of fulfilling such criteria, there is also a simulation study that is made with the validity of the proposed strategy. Implementing the proposed strategy in a real cloud environment is not easy.

There is an approach for replication of data which has been pro-posed by Rajalakshmi et al., [12]. Using a system of replica management, the replicas will be created, managed and also updated in case the original data has been modified. The main objective of the proposed work will be to concentrate on an algorithm design for the selection of an optimal selection of replica. The main phases that have been involved for this technique will be the file application and the replication operation. In the initial phase the location of replica and its creation using the catalog along with the index that will take place. And in the second phase it has been used for identifying if there is a sufficient space for the target for storing its stipulated file. The availability of the resources is the increasing by means of the replication in a minimum access cost, consumption of a shared bandwidth and the delay time by means of replicating data. These proposed systems have been developed in the Eucalyptus cloud environment. There is a better accessibility that is achieved by means of the algorithm of replica selection compared to that of the other techniques. The replica placement in geographical locations is not possible in this proposed method.

One such service that is provided is cloud storage which can reduce purchasing and maintenance cost. using an increased size of data, the data volume is brought down and this reduces the cost of running the large systems of storage. For improving the efficiency of storage in cloud, the techniques of deduplication are followed and this is read over a period of time. There are some datasets that are frequently accessed using a high redundancy level for the reliability need. So it becomes important to support this feature in case of a cloud storage. But the focus is on the static scheme which can re-strict the entire applicability in cloud storage. There is a dynamic scheme of deduplication in cloud storage that is proposed by Leesakul et al., [13] that will aim in the improving of the storage efficiency for maintaining the redundancy in fault tolerance. Need to monitor the changing of users' demand of files. Need to evaluate availability and performance of the system.

Cloud computing infrastructure will provide some scalable storage as well as computing resources for storage of big data. There is an effective management of data that includes the availability of data and its efficient access which is a very crucial for the applications. This is ideally achieved by means of using the replication of data by Shorfuzzaman [14] that will offer a reduced latency of data access and a higher availability of the data along with an enhanced system based load balancing. Furthermore, the QoS requisites for various applications can be different from one another. For offering continuous support to this QoS requirement there is a proposition of a very highly distributed technique of QoS aware replication that has been made which can compute an optimal location for the datacenters and for the replicas which have an overall cost of replication which has been minimized. Furthermore, a replication strategy will be to maximize this QoS satisfaction in the improvement of the data and the reduction of the latency of access. There is a dynamic programming that is used for formulating this problem. Finally, such experiments for the simulation have been performed that can be widely used in the observing of the patterns of data access in case of the demonstrating its effectiveness for the proposed technique. There is an existence of dynamic replica maintenance issues.

Using the replica of data for a cloud storage of large scale the applications will be minimized. There is a policy of a novel replica placement with the cloud storage load balancing has been made. This novel policy of replica placement for a cloud storage will be made. This novel replica placement policy along with bidding thought had been proposed by Zhang et al [15]. The features of the replica have been combined by means of the policy and the factors for the bidding mode, the bidding standard, the price and time and will start the replica bidding activity while getting the requisite which will not be met using the availability of file data. On the basis of the capacity as well as the probability of assessing of such data nodes the placement of replica will be done. The results of such study will show a better performance which has been seen as a proposed policy in the load balance and the response time compared to a static policy of replica and the CDRM scheme.

Some issues such as physical node failure, the relevant characteristics of user accessing mode influencing on replica policy are occurred in the existing method.

The Integrated file Replication and the consistency Maintenance mechanism was presented by Shalini and Surekha [16] that had integrated two techniques in a systematic and harmonized manner. There is a high replication that is achieved for the file replication which will update passively and determine the file replication and will update polling by the adapting of the dynamically varying file query and will update the rates which will avoid the unwanted file replications as well as the updates. The overhead will be dramatical-ly reduced and there will be some significant enhancements which will be yielded for the efficiency of file replication and the approaches of consistency maintenance. The IRM relying on polling file owners still cannot guarantee that all file requesters receive up-to-date files, although its performance is better than other consistency maintenance algorithms.

There is a new QoS-aware and Dynamic Data Replica Delete Strategy that has been proposed for the disc space as well as maintenance for the purpose of cost saving. The results of this study proved that the disc space can be saved using the DRDS technique along with the cost of maintenance for the distributed storage sys-tem at the same time having the assurance that is available for the QoS requirements.

There is a study on the data replication in case of cloud computing datacenters which was made by Boru et al., [18]. This was in addition to that of the enhanced QoS that was obtained because of the reduction of delay in communication. The results of this study from both the mathematical models and their extensive simulations have helped in the unveiling of the performance and the tradeoffs of en-ergy efficiency with guiding of the design for the solutions of future data replications. The proposed method reduces energy consumption, bandwidth usage, and communication delays significantly but the implementation is difficult.

In the cloud computing arena, the access to data will be enhanced using the DRS and the strategies of data replication will be pro-posed using the related studies. The strategies and their efficiency are related closely to the patterns of access for the users that work optimally in a specific pattern of data access. In such circumstances, there is an algorithm which has been proposed by Jeon et al., [19] for detecting the alterations in the pattern of data access for the users and this dynamically applies one more optimal strategy of replication. This technique that was proposed will have the advantage of maintaining the optimal performance by means to responding to the various patterns of data access and the algorithm has been tested and validate in terms of effectiveness. More data replication strategies and a wider variety of data access patterns should be observed and efforts will be made to progress an algorithm that dynamically identifies and applies an optimal ORS value. This proposed method consumes time to enhance its performance.

There are two QoS-Aware Data Replication (QADR) algorithms that were proposed by Lin et al., [9] in the cloud computing systems. The idea of the high-QoS first-replication (HQFR) was to be able to perform data replication on adopting the first algorithm using the greedy algorithm where the cost of replication cannot be reduced. The second algorithm had been proposed for achieving the two minimum objectives in which the second algorithm transforms the QADR problem into the well-known Minimum-Cost Maximum-Flow (MCMF) problem. For solving this problem of QADR the currently used MCMF algorithm had been applied; a minimum optimal solution to the problem of QADR in the polynomial time is achieved with the second algorithm and the time taken for computation is higher. Additionally, there are some large nodes that are present for the system of cloud computing. There is a time of large data application that was brought down by means of the node combinations techniques. There is a possibility of a large replication time for the data that has been reduced by the techniques of combination of nodes. Ultimately, the experiments of simulation had been done for demonstrating the effectiveness of the algorithm proposed which has been in the field of recovery and replication of data.

The technique of multi-objective offline optimization had been put forth by Long et al., [4] for the purpose of replica management, in which there are several factors influencing the decisions of replication that have been viewed which include the unavailability of the mean file, the mean service time, the load variance and the consumption of energy along with the mean access latency as their five objectives. The replication decisions had been made and laid out of the replication along with an enhanced, artificially immune algorithm that will spin around the solutions set of the candidates by means of mutation, clone and the process of selection. The technique was known as the Multi-objective Optimized Replication Management (MORM) that searches for the solutions which are near optimal using balancing of tradeoffs in five of the objectives of optimization. There were a series of experiments that have been reported that shoe the MORM and its effectiveness. The study's results conclude that the MORM was efficient in terms of energy and could outperform the management of default replication of HDFS (the Hadoop Distributed File System) and the MOE (the Multi-objective Evolutionary) algorithm in relation to the performance and the load balancing for storage clusters of a large scale.

A possible solution for decreasing the network traffic will be to replicate certain objects in various sites. Generally, this decision as to which one has to be replicated and where, there is required a solving constraint problem of optimization that is NP complete and the capacity of that of the Genetic Algorithm (GA) for the limit will be stretched by means of such problems. Furthermore, the GA had been propagated for identifying one such solution to this problem when there was a static read or a write demand for the superior quality of the solution.

However, there is a problem of high run time by that of the static GA approach that does not help in cases where there are continuous changes to the read/write demands in relation to its breaking news.

For the purpose of handling such issues, there is a hybrid GA technique which has been proposed by Loukopoulos and Ahmad [20] that will take the present replica distribution as its input with a new one computed using the knowledge of the traits of the network and all the changes that have taken place. Keeping this in mind the cur-rent pragmatic scenarios that are on the environments of distributed information, such protocols had been evaluated relating to the restriction of storage capacity for every site along with the variations. To lower the high running time, the proposed method presented many trade-off between running time/solution quality which leads to complexity.

III. METHODOLOGY

The Static allocation, the adaptive allocation and the harmonized search algorithm have been explained in detail by this section.

A. Static Replication Algorithm (SRA)

A Static Replication Algorithm (SRA) is that scheme of replication for objects assuming that there are no replicas that exist and the frequencies of the read or write are known and will remain static.

Data replication using a greedy algorithm

For every site $\mathrm{S}^{(i)}$ and the object O_k define the *replication benefit* value $\mathrm{B}_k^{(i)}$, in equation (1):

$$B_k^{(i)} = \frac{R_k^{(i)} - \left(\sum_{x=1}^{M} w_k^{(x)} O_k C(i, SP_k) - W_k^{(i)}\right)}{O_k} \quad (1)$$

The value above indicates the benefit expected in the Network Transfer Cost (NTC) terms, if the replication O_k is done at $\mathrm{S}^{(i)}$. A benefit is computed with difference between the NTC occurring from its current read requests and NTC arising owing to updates to the replica that is amortized to the size of the object and the negative of $\mathrm{B}_k^{(i)}$ k is replicating kth object, not efficient in "local view" of ith site. The local NTC that is observed from its ith site is thereby increased.

For proposing this algorithm a list $\mathrm{L}^{(i)}$ for $\mathrm{S}^{(i)}$ containing all replicated objects is maintained. The object O_k is replicated at $\mathrm{S}^{(i)}$ if the remaining capacity of storage $\mathrm{b}^{(i)}$ of its site is higher than the size and benefit value being positive. Keeping a list *LS* having all the sites with the "opportunity" for replicating an object. The site $\mathrm{S}^{(i)} \in LS$ should be $L^{(i)} \neq \varnothing$. This SRA Algorithm is performed in steps. In every step the site $\mathrm{S}^{(i)}$ will be chosen from the *LS* in that of a round-robin fashion with the benefit values of the objects that belong to $\mathrm{L}^{(i)}$ are duly computed. The one that has the highest benefit will be replicated along with the lists *LS*, $\mathrm{L}^{(i)}$ together by the nearest site value $SN_k^{(i)}$, which is accordingly updated. This SRA algorithm has outlined as below:

Initialize LS and all $L^{(i)}$

WHILE LS $\neq \varnothing$ DO

BMAX=0, OMAX=NULL /*BMAX holds the current max $\mathrm{B}_k^{(i)}$ value

OMAX holds the identity of the object for which $\mathrm{B}_k^{(i)}$ = **BMAX* */*

Pick up a site $\mathrm{S}^{(i)} \in LS$ in a Round-Robin way

FOR each $\mathrm{O}_k \in L^{(i)}$ DO

Compute $\mathrm{B}_k^{(i)}$

IF BMAX $\leq \mathrm{B}_k^{(i)}$ THEN BMAX=$\mathrm{B}_k^{(i)}$, $OMAX = k$

ELSEIF ($\mathrm{B}_k^{(i)} \leq 0$ OR $\mathrm{b}^{(i)} < O_k$) *THEN* $L^{(i)} = L^{(i)} - \{O_k\}$

Replicate O_{OMAX}

$\mathrm{L}^{(i)} = \mathrm{L}^{(i)} - \{\mathrm{O}_{OMAX}\}$ /*Reamove OMAX object from the list of potentials to be replicated*/

FOR all sites in LS update the relevant $\mathrm{SN}_{OMAX}^{(i)}$ field. /*Update "nearest sites"*/

$\mathrm{b}^{(i)} = \mathrm{b}^{(i)} - \mathrm{O}_k$ /*New remaining capacity*/

IF $\mathrm{L}^{(i)} = \varnothing$ THEN LS=LS-{|$\mathrm{S}^{(i)}$|} /*Remove $\mathrm{S}^{(i)}$ if there are no other candidates (objects) to be replicated*/

ENDWHILE

B. Adaptive replication Algorithm

If *R* is the old scheme of replication for all objects with R' being the newly defined scheme. If D_R and $D_R{}'$, indicate the total NTC that is created by the *R* and R' both will be computable. Further, the X'' be an *M*×*N* (0, 1) this matrix having an element X'_{ik} that is 1 in case the O_k has been replicated at the point $S^{(i)}$ under a scheme of R' and otherwise is 0. While realizing that some R' replicas will have to be deleted, there is a need for others to be created. If it is assumed that the replica creation is made by means of transferring one copy of this object from its respective primary site, the replica deletion has no cost and the NTC $I_{RR'}$, realizing that the R' has been given as per equation (2):

$$I_{RR'} = \sum_{i=1}^{N} \sum_{k=1}^{M} X'_{ik}(1 - X_{ik}) O_k C(i, SP_K) \quad (2)$$

In which X_{ik} is the variable of allocation of the replication scheme *R*. The benefit, $V_{RR'}$, to move it from R' scheme to that of the R' has been given as per equation (3):

$$V_{RR'} = D_R - (D_R{}' + I_{RR'}) \quad (3)$$

This Adaptive Data Replication Problem (ADRP) is defined as: Given this *X* matrix identifies values of the X' which will maximize $V_{RR'}$, being subject to that of the constraints of storage capacity. This scope of the ADRP is found to be different to that of the DRP. The decisions of allocation have been represented using the DRP during nighttime by means of that of a monitor site by means of which the statistics are duly gathered on the requests of the decisions and the objects that are accordingly taken (with an assumption that is indirect making the actual cost for realizing the scheme of replication in the nighttime that is significant and can be omitted). At the same time, this situation has been described using the ADRP while the replication schemes have been realized during nighttime and will not function well during the day owing to the frequencies of request exhibited differing to a great extent from that of the used estimation. Owing to this as opposed to the currently existing techniques there is a new one that has been defined from a scratch which is needed for this algorithm to be fine-tuned inside limits that are reasonable

C. Proposed Harmonized Replication Algorithm (HRA)

- The harmony method's inspiration has been made through that of the working principles of the improvisation of harmony [21]. This has been ensured by means of the memory considering as well as the pitch ensuring for which there are some good local solutions which have been retained for the purpose of randomization as well as harmony memory while consider-ing the exploring of the global search space in an effective manner. There is also a subtlety which is the presence of a controlled diversification around a good solution (for the good harmonics and the pitches) and this will act as a factor of in-tensification. The basic HS technique has been shown in a flowchart in Fig 2 and the principle steps involved in this are as below:
- Step 1. Initializing of the HS Memory (HM). An initial HM will contain a particular number of solutions that are generated randomly to the problems of optimization that are considered. For that of an n-dimension problem, the HM will have a size of *N*
- Step 2. Improvising of a new solution from that of the HM. Every component will be obtained on the basis of the Harmo-ny Memory Considering Rate (HMCR). This HMCR has been defined as that probability of the choice of a component from its HM members and the 1-HMCR will therefore be the probability of its random generation.
- Step 3. Updating of the HM. This new solution from the Step 2 will be evaluated. In case it yields better fitness than the worst member within the HM it replaces that one or else is eliminated.
- Step 4. Repeating of Step 2 to Step 3 till such time a preset termination criterion like a maximal iteration number is met

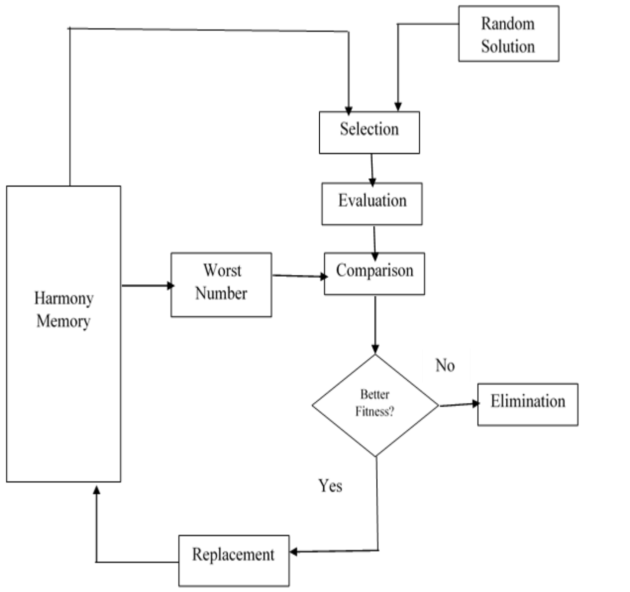

Fig 2 Flowchart for Proposed Harmonized Replication Algorithm (HRA)

In this proposed harmonized replication algorithm, every memory will be a bit strength having a length *M* that indicates the assignments of the *Ok*. The proposed method has been taken as the input and also the new patterns which have been exhibited for the *Ok* and along with a set of some schemes of near optimal replication for one specific object by not considering the capacity of storage for the sites. After that for the initial HRA solutions, the *Rks* are incorporated, and violation of constraint of storage is repaired by the de-allocating of least beneficial objects.

IV. Results And Discussion

For the experiment static, the adaptive the harmonized replication algorithms will be used. The tables 1 to 3 and the Fig 3 to 5 show its result table along with the graph of the Number of the Cloud datacenter, the Number of Data that has to be replicated and the Capacity % for the Static, the adaptive and also the Proposed Harmonized Replication Algorithm.

Table 1 Number of Cloud datacenter for Proposed Harmonized Replication Algorithm.

Number of Cloud datacenter	Static Replication Algorithm	Adaptive Replication Algorithm	Proposed Harmonized Replication
5	30	32	34
10	35	38	41
15	41	44	47
20	48	51	55
25	56	60	64

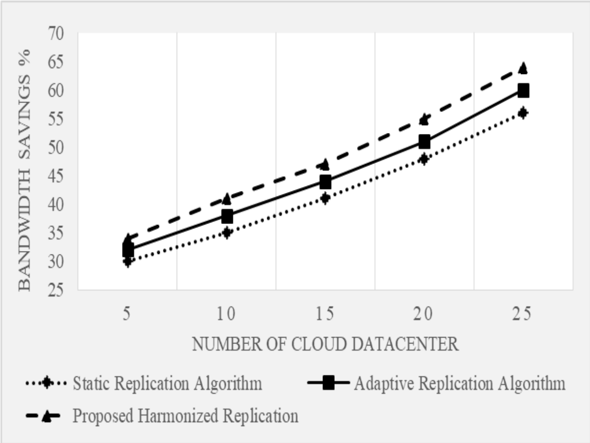

Fig 3 Number of Cloud datacenter for Proposed Harmonized Replication Algorithm

It is observed from table 1 and Fig 3 is that the proposed harmonized replication performs better by produce bandwidth saving by 12.5% and by 6.1% than static replication algorithm and adaptive replication algorithm for 5 number of cloud data centers. Again the proposed harmonized replication performs better by produce bandwidth saving by 13.64% and by 6.6% than static replication algorithm and adaptive replication algorithm for 15 number of cloud data centers. Again the proposed harmonized replication performs better by produce bandwidth saving by 13.33% and by 6.5% than static replication algorithm and adaptive replication algorithm for 25 number of cloud data centers.

Table 2 Number of Data to be replicated for Proposed Harmonized Replication Algorithm

Number of Data to be replicated	Static Replication Algorithm	Adaptive Replication Algorithm	Proposed Harmonized Replication
50	32	34	36
100	37	40	44
150	43	46	49
200	51	54	59
250	59	63	67

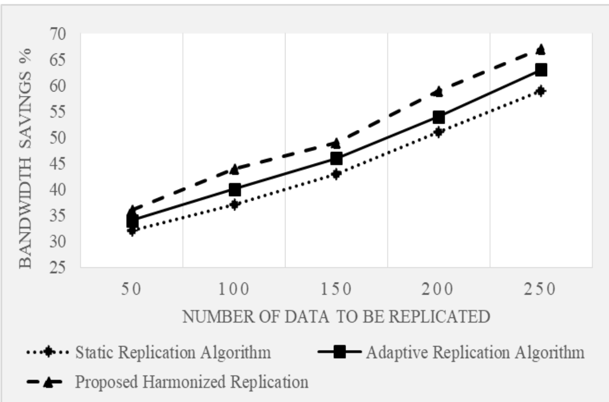

Fig 4 Number of Cloud datacenter for Proposed Harmonized Replication Algorithm

It is observed from table 2 and Fig 4 is that the proposed harmonized replication performs better by produce bandwidth saving by 11.76% and by 5.71% than static replication algorithm and adaptive replication algorithm for 50 number of data to be replicated. Again the proposed harmonized replication performs better by produce bandwidth saving by 13.04% and by 6.32% than static replication algorithm and adaptive replication algorithm for 150 number of data to be replicated. Again the proposed harmonized replication performs better by produce bandwidth saving by 12.7% and by 6.2% than static replication algorithm and adaptive replication algorithm for 250 number of data to be replicated.

Table 3 Capacity % for Proposed Harmonized Replication Algorithm

Capacity %	Static Replication Algorithm	Adaptive Replication Algorithm	Proposed Harmonized Replication
10	41	45	48
30	51	56	60
50	56	60	65
70	70	75	81
90	79	85	92

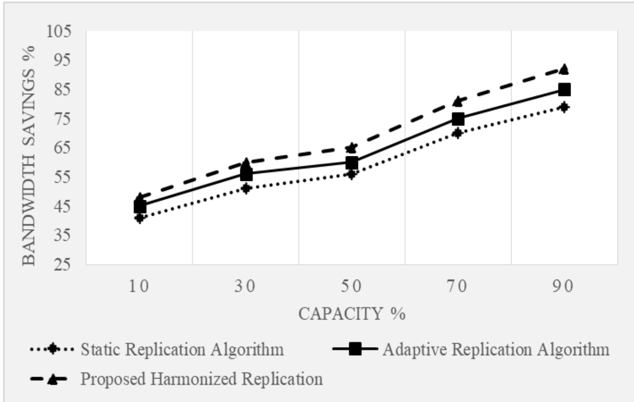

Fig 5 Capacity % for Proposed Harmonized Replication Algorithm

It is observed from table 2 and Fig 5 is that the proposed harmonized replication performs better by produce bandwidth saving by 15.73% and by 6.45% than static replication algorithm and adaptive replication algorithm for 5 capacity%. Again the proposed harmonized replication performs better by produce bandwidth saving by 14.9% and by 8% than static replication algorithm and adaptive replication algorithm for 15 capacity%. Again the proposed harmonized replication performs better by produce bandwidth saving by 15.2% and by 7.91% than static replication algorithm and adaptive replication algorithm for 25 capacity%.

V. CONCLUSION

The Cloud computing has recently emerged as one popular model of busing for the utility of the computer system. The concept of the Cloud will be the provision of computing resources being a utility of service on demand to the customers in the Internet. The efficiency of storage has been improved by means of data deduplication and for the experiments, the static, the adaptive and the methods of harmonized replication have been used. According to the results the harmonized replication performed better in terms of produce bandwidth and saving by about 12.5% and further by 6.1% than that of the static replication algorithm and the adaptive replication algorithm for the 5 number of the cloud data centers. Also the proposed harmonized replication had performed better in terms of the produce bandwidth saving by about 13.64% and further by about 6.6% than that of the static replication algorithm and also the adaptive replication algorithm for the 15 number of such cloud data centers. Further, this proposed harmonized replication had performed better by the produce bandwidth saving by about 13.33% and further by about 6.5% than that of the static replication algorithm along with an adaptive replication algorithm for the 25 number of such cloud data centers. The future work can be extended by proposing a better optimization techniques and the existing results can be improved in an effective way.

REFERENCES

1. M Armbrust, A Fox , R Grifth, A D Joseph,R Katz, A Konwinski ,G Lee ,D Patterson,A Rabkin,I Stoica, M Zaharia. A view of cloud computing. Communications of the ACM, 2010, 53(4): 50-58.
2. P Mell, T Grance. The NIST definition of cloud computing. Communications of the ACM, 2010, 53(6): 50.
3. D W Sun,G R Chang, S Gao,L Z Jin, &X W Wang, Modeling a dynamic data replication strategy to increase system availability in cloud computing environments. Journal of computer science and technology, 27(2), 256-272,2012.
4. Long Sai-Qin, Zhao Yue-Long, Chen Wei. MORM: A Multi-objective Optimized Replication Management strategy for cloud storage cluster. J Syst Arch 2014;60 (2):234–44,2014.
5. Berl Andreas, Gelenbe Erol, Di Girolamo, Marco Giuliani, Giovanni De Meer, Hermann Dang, Minh Quan, Pentikousis Kostas. Energy-efficient cloud computing. Comput J, 53; 1045–51.
6. Shvachko Konstantin, Kuang Hairong, Radia Sanjay, Chansler Robert. The Hadoop distributed file system. Paper presented at the 2010 IEEE 26th symposium on Mass Storage Systems and Technologies (MSST); 2010.
7. B A Milani &N J Navimipour , A comprehensive review of the data replication techniques in the cloud environments: Major trends and future directions. Journal of Network and Computer Applications, 64, 229-238, 2016.
8. Bai Xiaohu, Jin Hai, Liao Xiaofei, Shi Xuanhua, Shao Zhiyuan. RTRM: a response time-based replica management strategy for cloud storage system. In: Proceedings of the grid and pervasive computing. Springer; 124–33.
9. J W Lin, C H Chen, &J M Chang,(2013). QoS-aware data replication for data-intensive applications in cloud computing systems. IEEE Transactions on Cloud Computing, 1(1), 101-115.
10. D.S Jayalakshmi, &R Srinivasan, R. (2015, September). Dynamic Data Replication Strategy in Cloud Environments. In Advances in Computing and Communications (ICACC), 2015 Fifth International Conference on (pp. 102-105). IEEE.
11. U Tos, R Mokadem, A Hameurlain, T Ayav & S Bora, A performance and profit oriented data replication strategy for cloud systems. In 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), 2016 Intl IEEE Conferences (pp. 780-787).July 2016.
12. Rajalakshmi, A., Vijayakumar, D., & Srinivasagan, K. G. (2014, April). An improved dynamic data replica selection and placement in cloud. In Recent Trends in Information Technology (ICRTIT), 2014 International Conference on (pp. 1-6). IEEE.
13. Leesakul, W., Townend, P., & Xu, J. (2014, April). Dynamic data deduplication in cloud storage. In Service Oriented System Engineering (SOSE), 2014 IEEE 8th International Symposium on (pp. 320-325). IEEE.
14. Shorfuzzaman, M. (2014, December). Access-efficient QoS-aware data replication to maximize user satisfaction in cloud computing environments. In Parallel and Distributed Computing, Applications and Technologies (PDCAT), 2014 15th International Conference on (pp. 13-20). IEEE.
15. H Zhang, B Lin, Z Liu&W Guo, Data replication placement strategy based on bidding mode for cloud storage cluster. In Web Information System and Application Conference (WISA), 2014 11th (pp. 207-212). IEEE,2014.
16. K Shalini &Y Surekha, Effective file replication and consistency maintenance mechanism in P2P systems. Global Journal of Computer Science and Technology,2012.
17. B Liao,J Yu,H Sun, & M Nian, A QoS-aware dynamic data replica deletion strategy for distributed storage systems under cloud computing environments. In Cloud and Green Computing (CGC), 2012 Second International Conference on (pp. 219-225). IEEE,2012.
18. D Boru, D Kliazovich, F Granelli,P Bouvry &A Y Zomaya. Energy-efficient data replication in cloud computing datacenters. Cluster computing, 18(1), 385-402,2015.
19. M Jeon, K H Lim,H Ahn,&B D Lee, Dynamic data replication scheme in the cloud computing environment. In Network Cloud Computing and Applications (NCCA), 2012 Second Symposium on IEEE, pp. 40-47,2012.
20. T Loukopoulos &I Ahmad, Static and adaptive distributed data replication using genetic algorithms. Journal of Parallel and Distributed Computing, 64(11), 1270-1285,2004.
21. X Z Gao,V Govindasamy,H Xu,X Wang & K Zenger, . Harmony search method: theory and applications. Computational intelli

AUTHORS PROFILE

Chandrakala H L. Received B.E in CSE from Malnad College of Engineering ,Hassan, M.TECH IN CSE from VTU study center ,BMS College Of Engineering ,Bangalore. PhD in CSE from from VTU-RRC. He is now working as Associate professor of HKBK College of Engineering, Bangalore. His Academic experience is 23 years. He published his research work in 2 journals and attended Seminars and workshops.

Dr. Loganathan R. Received B.E in CSE from Bharathiar University, Coimbatore, M.TECH IN CSE CSE] from Visvesvaraya Technological University, Karnataka. PhD in CSE from from Sathyabama University, Chennai. He is now working as Professor and HOD/CSE of HKBK College of Engineering, Bangalore. His Academic experience is 20 years. He published his research work in several journals and conferences.