iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 4 **Issue:** VIII **Month of publication:** August 2016

DOI:

Volume 4 Issue VIII, August 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Anticipatory Data Replication Strategy with Dynamic Distributed Model for Cloud Computing

Umesh Lilhore[1], Dr Santosh Kumar[2]
[1]*PhD Scholar,* [2]*PhD Supervisor*
MUIT Lucknow, Department of Computer Science & Engineering

Abstract: Cloud Computing can process intensive applications and various scientific functions across heterogeneous environments. In performance of cloud computing correct load balancing is always desirable and challenging, these methods minimize the total access time and also increase the availability. Through load balancing, nodes will be controlled and prevented from overloading and as a result system throughput will be enhanced. In this research paper we are presenting an "Anticipatory Data Replication Strategy with Dynamic Distributed Model for Cloud Computing (ADRS-DDMC)". This paper introduces a novel dynamic data replication method that is functioning based on anticipations to create the pre-replicas for future needs of the sites. Proposed ADRS-DDMC method optimizes load balancing by increasing the data availability among the existing sites. The results are compared with LRU (Least recently used) and LFU (Least frequently used).Proposed and existing method are implemented over cloud sim simulator, experimental results clearly shows that proposed method have better performance as compared to existing methods.
Keywords- Cloud Computing, Performance of Cloud computing, ADRS-DDMC, LRU and LFU

I. INTRODUCTION

Cloud computing is an entirely internet-based approach where all the applications and files are hosted on a cloud which consists of thousands of computers interlinked together in a complex manner[1,5]. Cloud computing incorporates concepts of parallel and distributed computing to provide shared resources such as various hardware, software and information to computers or other devices on demand[8]. These are emerging distributed systems which follows a "pay as you use" model. The customer need not buy the software or computation platforms. With internet facility, the customer can use the computation power or software resources by paying money only for the duration user has used the resource. This new logic forces the conventional software licensing policies to change, upgrades and avoids spending of money for the facilities the customer does not use in a software package [2,3].

II. RELATED WORKS

Load balancing is removing tasks from over loaded VMs and assigning them to under loaded VMs. Load balancing in cloud computing can affect the overall performance of a system executing an application[1,7]. A new idea of Dynamic Load Balancing Strategy for Grid Computing technique is presented addressing the problem of load balancing in Grid computing. As in this paper also proposes a load balancing model based on a tree representation of a Grid computing [5, 6]. This load balancing strategy has two main objectives-

A. Reduction of the mean response time of tasks submitted to a Grid; and,
B. Reduction of the communication costs during task transferring.

III. PROBLEM FORMULATION

Replication techniques have a dramatic impact on the performance of the systems. However, it could be costly for a computing system, if proper replication methodology is not selected. Therefore it is challenging to understand when replication is necessary, which files should be selected for replication, where the replicas should be stored and how the replicas should be synched with the original files. Least Recently Used method (LRU) and Least Frequently used (LFU) method are popular heuristic examples of data replication methods that have been applied in various Cloud computing systems [6,8].

Pre-replicating the required data, the algorithm should effectively minimize the job execution time and enhance the effective network usage so as a result load balancing would be improved. Given the circumstances, in this research a novel algorithm has been designed that can predict the future needs of the existing sites. Based on data access catalogue, the algorithm is able to anticipate the data with high access probability that could be needed in future [9,11].

Volume 4 Issue VIII, August 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

IV. PROPOSED METHOD

Proposed ADRS-DDMC method uses dynamic data replication for increasing the reliability and availability of cloud data.

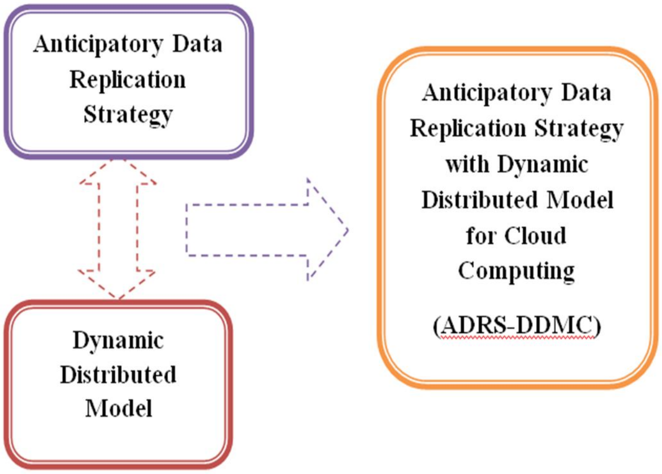


Figure 4 Architecture of ADRS-DDMC

Pre-replication can increase the data availability and robustness of the Cloud systems and hence requested jobs can be completed with minimum execution time and high network usage output.

A. Performance parameters for proposed ADRS-DDMC Method

In Proposed ADRS-DDMC method the reliable VM's are identified based on-

1) Time consumption (milliseconds)
2) Memory availability
3) Previous history

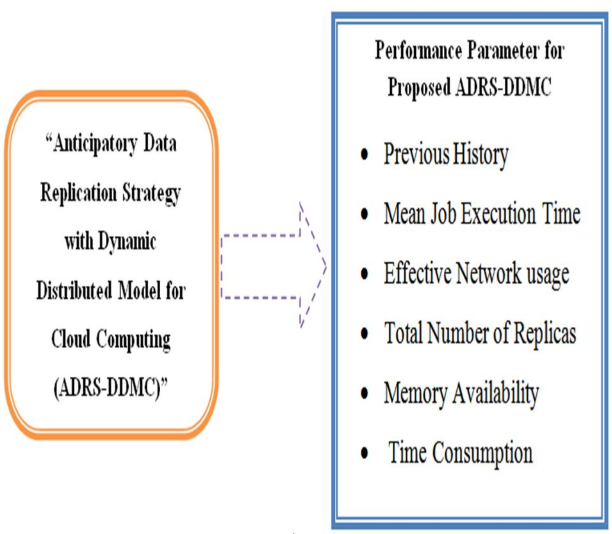


Figure 4.1 Parameters in ADRS-DDMC

B. Algorithm for ADRS-DDMC Algorithm for data replication

Following steps are used in proposed method-

Step 1- GRMS Store the access patterns as -

("Requested file name, total number of accessing file and total requested")

Step 2- User request for a file or data

Step 3-Search the history catalogue

Volume 4 Issue VIII, August 2016
ISSN: 2321-9653

3.1 If the requested file exist in local server
3.2 retrieve the file and exit algorithm
Else
3.3 A request will be sent to GRMS.
Step 4- (Call A* algorithm () for search)
4.1 A* search will be initiated in catalogue history to find the physical location of the file between all the available files,
4.2 Calculate the communication cost and select the file with minimum value
Communication cost = Size of the replica / Bandwidth between servers
Step 5- For the selected replica
5.1 Check if it is beneficial to pre-replicate its adjacent files
5.2 Check if the file has hierarchy then
5.3 If there exist only one child then
5.4 replicate and exist, Else
5.5 For 3 tiers after the replica
5.5.1 Select the child with maximum access number
5.5 retrieve selected replica & it's adjacent

V. SIMULATION & RESULT ANALYSIS

To evaluate the performance of our proposed algorithm Java programming has been used to extend the Cloudsim simulation tool.

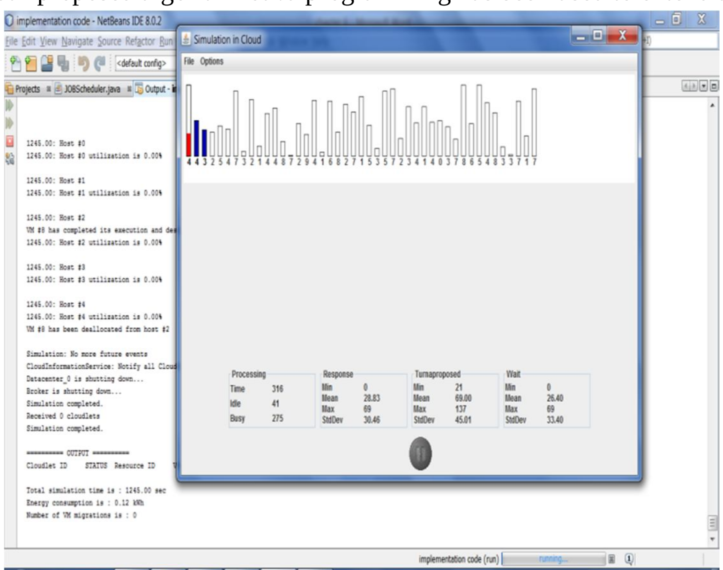


Figure 5-Simulation for LRU, MFU and Proposed method

A. Mean Job Execution Time

Mean job execution time is one of the important evaluation factors. Total execution of all jobs in milliseconds divided by a number of the jobs would highlight the mean job execution.

Access Pattern	Mean job Execution Time (in Seconds)		
	LRU	LFU	ADRS-DDMC
Sequential Access	1500	1500	1200
Random Access	1200	1100	850
Random Walk Unitary access	700	500	500
Random walk Gaussian access	1300	1200	800

Volume 4 Issue VIII, August 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Table 5.1 Mean Job Execution Time

Figure 5.1 Mean Job Execution Time

B. Effective Network Usage

The metric indicates the ratio of the files that were transferred to the requested site. Lower network usage by a method shows better performances.

ENU = (N file remote access + N file replication) / (N remote file access + N local file access)

Access Pattern	Effective Network Usage %		
	LRU	LFU	ADRS-DDMC
Sequential Access	45%	37.50%	32.50%
Random Access	30%	25%	20%
Random Walk Unitary access	20%	14.50%	10%
Random walk Gaussian access	40%	35%	27.50%

Table 5.2 Effective Network Usage

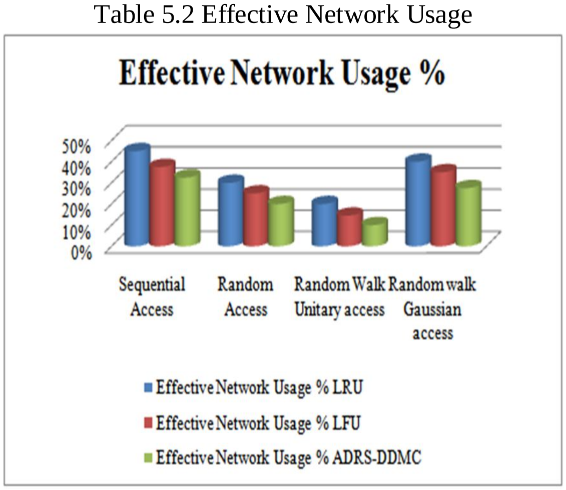


Figure 5.2 Effective Network Usages

C. Total Number of the Replicas-Greater values of the replication numbers indicates that the files were not stored locally and replication procedures were needed to make the files available.

Volume 4 Issue VIII, August 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Access Pattern	Total number of replications		
	LRU	LFU	ADRS-DDMC
Sequential Access	580	500	280
Random Access	410	390	210
Random Walk Unitary access	175	185	125
Random walk Gaussian access	385	400	200

Table 5.3 Total Number of the Replicas

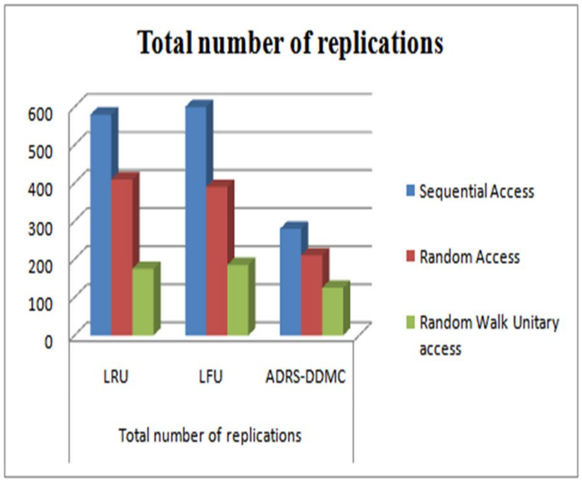


Figure 5.3 Total Numbers of the Replicas

D. Result Analysis

The above results clearly shows that the proposed dynamic replication algorithm performs outstanding over existing methods.

VI. CONCLUSION & FUTURE WORKS

In this work our proposed algorithm uses the pre-replicating technique for replicating the adjacent files of the requested jobs from different sites. The algorithm consists of three main phases which anticipate the future needs of the sites and pre-replicates the files that no requests have been submitted for them yet. By pre-replication, sites will have files locally stored, so at the time of need, sites can access the files locally with minimum response time. Simulation results clearly shows that proposed method performs outstanding in terms of access latency, number of replicas and network usage.

REFERENCES

[1] Bitar.N, Gringeri, S.Xia, T.J, "Technologies and protocols for data center and cloud are networking", IEEE Transactions Pages: 24-31, September 2013.Alicherry.M. and Lakshman, T.V, "Optimizing data access latencies in cloud systems by intelligent virtual machine placement", IEEE Communications- Pages: 647- 655-April 2013.

[2] Amit Kumar Das, Tamal Adhikary, Md. Abdur Razza que, Chong Seon Hong, "An intelligent approach for virtual machine and QoS provisioning in cloud computing" PP 462-467-International Conference on Information Networking (ICOIN), 2013.

[3] Brototi Mondala, Kousik Dasgupta and Paramartha Duttab,"Load Balancing in Cloud Computing using Stochastic Hill Climbing-A Soft Computing Approach" Science Direct, Pages: 783 -789, May 2012.

[4] Chao-Tung Yang, Wei- Li Chou, Ching-Hsien Hsu and Cuzzocrea, A., "On Improvement of Cloud Virtual Machine Availability with Virtualization Fault Tolerance Mechanism", IEEE Cloud Computing, Pages: 122-129, Dec. 1st 2011.

[5] Dong Oh Son, Hong Jun Choi, Jae Hyung Park and Cheol Hong Kim ," Analysis of Memory Management Policies for Heterogeneous Cloud Computing",IEEE Transactions, Pages:1- 3-June 2013.

[6] Faragardi, H.R., Shojaee, R., Tabani, H. and Rajabi, A.," An analytical model to evaluate reliability of cloud computing systems in the presence of QoS requirements", IEEE Transactions, Pages 315-321, June 2013.

[7] Huankai Chen, Wang, F., Helian,N. and Akanmu, G."User- priority guided Min- Min scheduling algorithm for load balancing in cloud computing", IEEE Parallel Computing Technologies (PARCOMPTECH-13), Pages: 1- 8, Feb. 2013.

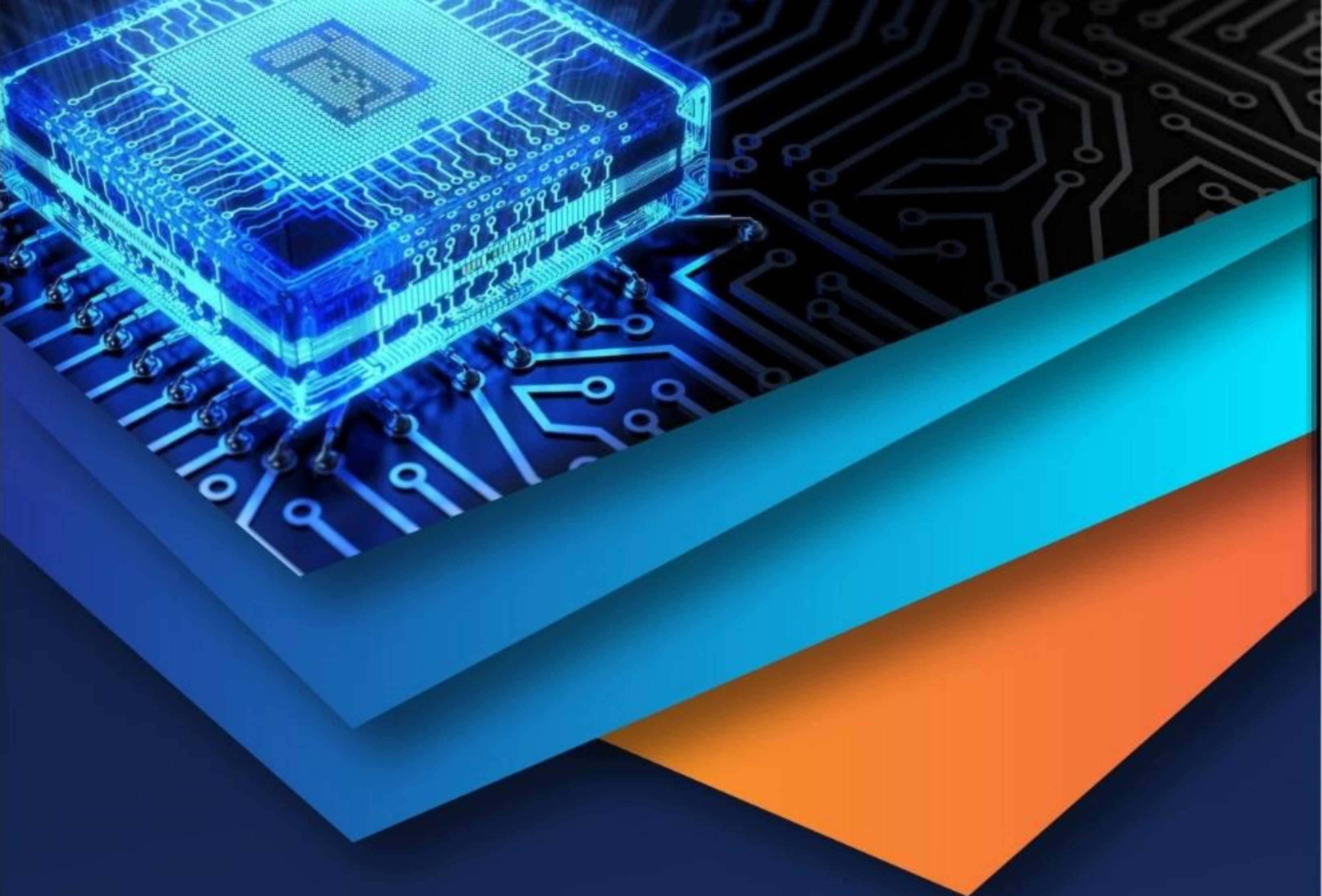
Volume 4 Issue VIII, August 2016
ISSN: 2321-9653

[8] Myunghoon Jeon, Kwang Ho-Lim, Hyun Ahn and Byoung-Dai Lee," Dynamic Data Replication Scheme in the Cloud Computing Environment " ,IEEE Transactions, Pages 40-47, Dec 2012.

[9] Nawsher Khan, A. Noraziah and Tutut Herawan,"A Cloud Architecture with an Efficient Scheduling Technique", Springer pp 381-388, 2012.

[10] Umesh lilhore and Santosh Kumar, "Analysis of performance factors for cloud computing", International Journal of Information Technology and Management (IJIMTM ignited journal), Vol IX, Issue No. XIV, November 2015, ISSN 2249-4510, PP 305-310.

[11] Umesh lilhore and Santosh Kumar, "Survey on Load balancing in Cloud Computing", International Journal of Information Technology and Management (IJIMTM ignited journal), Vol X, Issue No. XV, May 2016, , ISSN 2249-4510, PP 224-230.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY