

# **A New Load Balancing Technique for Virtual Machine Cloud Computing Environment**

**Divya Chaudhary**

M.Tech Scholar  
Department of Computer Science and Applications,  
M.D. University, Rohtak, Haryana, India

**Rajender Singh Chhillar**

Head  
Department of Computer Science and Applications,  
M.D. University, Rohtak, Haryana, India

## **ABSTRACT**

Cloud computing is one of the latest and upcoming paradigm that offers huge benefits such as reduced time to market, unlimited computing power and flexible computing capabilities. It is a model that provides an on-demand network access to a shared pool of computing resources. It comprises a large number of concepts primarily Load Balancing, Scheduling, etc. This paper discusses load balancing as a mechanism to distribute the workload evenly to all nodes in the system to achieve a higher resource utilization and user satisfaction. It helps in allocation and de-allocation of instances of applications without failure. This paper reports a new load balancing technique and its comparison with round robin providing encouraging results.

**Keywords:** Load Balancing, Cloud Computing, Scheduling, Round Robin, CloudSim.

## **1. INTRODUCTION**

Cloud Computing is intended to enable the computing across largely and diverse resources and dynamic stability. It is a subsidised form of cluster computing and utility computing. It has the potential of making the 'computing as a utility' a reality in future. It is used to describe both are type of application and a type of application as well as a platform. It describes a large number of applications that are accessible through the internet and for this purpose large data servers to host web applications are used [1]. It provides an on-demand computing that highlights subscription based services where one can obtain huge number of computer resources and the storage spaces. The clouds depict virtualized data centres. Cloud Computing provides scalable, pay-as-you-go, just in time allocation of resources. Cloud Computing provides a new deployment model of resources as accessible using public or private networks. NIST states the cloud computing definition as:-"Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." A Cloud is a conceptually an elastic and distributed system where the resources and storage spaces are distributed through the networks. The Clouds offers a large number of characteristics likely:-

- a. On-Demand Self Service
- b. Measured Service
- c. Rapid Elasticity
- d. Resource Pooling
- e. Broad Network Access

The clouds help in achieving better utilization of resources and obtaining performance enhancements for applications capable of offering dynamic cloud based respectively Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). The full set of resources help a system to implement a client which requires intercommunication between various components of the system [2]. The CloudSim is an example of such an approach, where a computing platform is provided. The main objective is the provision of the resources through management and governance. The separation permits the implantation of management techniques such as Scheduling, Load Balancing, etc. The governance and scalability are managed as a part of differentiated coordination framework. The rest of this paper is organized as follows. The load balancing fundamentals are discussed in Section II. The existing load balancing policies are presented in Section III. The proposed work is specified in Section IV. Section V describes the experimental setup for simulation environment, CloudSim. Finally, the conclusions and the future works are discussed in Section VI.

## **2. LOAD BALANCING AND ALGORITHMS**

Load balancing is a computer networking method to distribute workload across multiple computers or a computer cluster, network links [2], central processing units, disk drives, or other resources, to achieve optimal resource utilization, maximize throughput, minimize response time, and avoid overloading of the system. Using multiple components with load balancing, instead of a single component, may increase reliability through redundancy. The load balancing service is usually provided by dedicated software or hardware, such as a Domain Name System server. It is an essential part of distributed computing system. The load distribution among server in a network needs to be performed. The CPU passes the information from the heavily loaded servers to the underutilized servers [3]. The main task is to efficiently and effectively utilize the infrastructure available in cloud for processing the user's requests. It helps in the design of the power efficient systems leading to Green Computing.

It is a process of reassigning the total load to the individual nodes of the collective system to make resource utilization effective and to improve the response time of the job, simultaneously removing a condition in which some of the nodes are over loaded while some others are under loaded [4]. To distribute the load on different we generally use a large number of traditional and predefined algorithms in the system. But some of the traditional algorithms do not provide proper results in cases of proper utilization and the allocation of the resources in the system. The virtual machines evenly distribute the load in the system [5]. Load balancing handles various issues:-

- a. Efficient distribution of user processes on virtual machines.
- b. Efficient distribution of virtual machines on the physical servers.

Load Balancing helps in:

- a. Improving the performance substantially.
- b. Having a Reverse up plan in case the system fails even partially.
- c. Maintenance of system stability.
- d. Accommodation of future modification.
- e. Efficient load distribution.
- f. Cost effectiveness.

## **2.1. Load Balancing Algorithms**

The load balancing algorithms are designed to balance the load in the system for the fulfilment of the goals and producing the optimized results as a whole [6]. The important things to consider while developing such algorithm are: estimation and comparison of load, stability of different system, performance of system, interaction between the nodes, nature of work to be transferred, selecting of nodes and many other ones. The Challenges faced by Load Balancing Algorithms are [7]:

- a) Virtual Machines Migration: How to distribute the data among various machines
- b) Development of small data centres for cloud computing: Helpful in creation of small and highly beneficial data centres leading to geo-diversity computing.
- c) Energy Management: Helpful in obtaining the economy of the scale.
- d) Management of stored data: Distribution of data for the optimum utilization of space for the storage of data.
- e) Automated service provisioning: Increasing elasticity and provisioning of resources automatically.

### **a) Round Robin:**

It is a type of Static and Decentralized algorithm in nature. In the following algorithm, the processes are divided between all processors. Each process is assigned to the processor in a round robin order using a particular time value. The process allocation order is maintained locally independent of the allocations from remote processors. Though the work load distributions between processors are equal but the job processing time for different processes are not same. So at any point of time some nodes may

be heavily loaded and others remain idle. [8] This algorithm is mostly used in web servers. This algorithm is also implemented as Weighed Round Robin Algorithm. In this we assign the weights to the system in a manner for better resource allocation and utilization.

### **b) Throttled Load Balancer:**

This algorithm is a dynamic load balancing algorithm. It is used for load balancing in the case of the virtual machines to be used. Here we first check the index values of all the virtual machine in the system. The request is sent where load balancer parses a table for the allocation of the resources in the system. It assigns the request to a particular load balancer which passes or responds reverse the request to the requester and updates the allocation policy [9]. After the successful allocation of the system the whole process for the de-allocation of the system also starts. This mechanism provides a greater a higher amount of resource sharing and allocation on a whole in the system resulting in the higher performance and utilization. The throttling threshold maintained generally is 1. It could be modified easily to make the threshold a configurable value.

### **c) Active Monitoring Load Balancer:**

This load balancing policy attempts to maintain equal workloads on all the available VMs. The algorithm used is quite similar to the throttled case as explained above but with faster and timely checking as well as the accessibility of the resources in the system. The ids for the allocation and de-allocation are specified. The value of the count changes with a new request. It gives the maximum utilization and performance of resources and machines respectively. It is a dynamic allocation algorithm. [10] The important things to consider while developing such algorithm are: estimation and comparison of load, stability of different system, performance of system, interaction between the nodes, nature of work to be transferred, selecting of nodes and many other ones.

After studying various load balancing algorithms, we observe where they are improving and lacking as well. As a consequence, simulating of these algorithms might be supportive in order to accomplish this research. Here, a new algorithm has been proposed by modifying the round robin load balancing algorithm in the virtual machine environment of cloud computing to achieve better response time, processing time and cost.

## **3. EXISTING ROUND ROBIN LOAD BALANCING ALGORITHM**

The datacentre controller uses a load balancer called round robin using the simple Round robin algorithm to allocate virtual machines to cloudlets. This algorithm works as follows:

1. RoundRobinVmLoadBalancer makes an index table of Virtual Machines and the state of the VM (BUSY/AVAILABLE). At the start all VM's are available.
2. DataCenterController receives a new request.
3. DataCenterController queries the RoundRobinVmLoadBalancer for the next allocation.
4. RoundRobinVmLoadBalancer parses the allocation table from top until the first available and allocates VMs on the

basis of the time quantum for the equal sharing of the resources.VM is found.

If found:

- a. The RoundRobinVmLoadBalancer returns the VM id to the DataCenterController
- b. The DataCenterController sends the request to the VM identified by that id.
- c. DataCenterController notifies the RoundRobinVm LoadBalancer of the new allocation
- d. RoundRobinVmLoadBalancer updates the allocation table accordingly

If not found:

- e. The RoundRobinVmLoadBalancer returns -1.
- f. The DataCenterController queues the request
5. When the VM finishes processing the request, and the DataCenterController receives the response cloudlet, it notifies the RoundVmLoadBalancer of the VM de-allocation.
6. The DataCenterController checks if there are any waiting requests in the queue. If there are, it continues from step 3.
7. Continue from step 2.

#### 4. REVERSE FILLING LOAD BALANCING ALGORITHM

The proposed naïve algorithm named as Reverse Filling Load Balancing Algorithm. This is a modification of the existing Round Robin in which the after the creation of the virtual machines at the datacenter's, they are stored in the reverse order in the queue. The allocation policy used here also allocates the hosts in the reverse order of their creation in the cloud. Thus, the cloudlets are assigned to their respective virtual machines in the Last in First out Order and giving lesser debt, cost, waiting time and faster response time. The algorithm is as follows:

1. ReverseFillingVmLoadBalancer makes an index table of Virtual Machines and the state of the VM (BUSY/AVAILABLE). At the start all VM's are available.
2. DataCenterController receives a new request.
3. DataCenterController queries the ReverseFillingVmLoadBalancer for the next allocation.
4. The ReverseFillingVmLoadBalancer parses the table in the reverse order and creates a new allocation table containing the index table of the virtual machines.
5. ReverseFillingVmLoadBalancer allocates VMs on the basis of the time quantum for the equal sharing of the resources.VM is found.

If found:

- a. The ReverseFillingVmLoadBalancer returns the VM id to the DataCenterController
- b. The DataCenterController sends the request to the VM identified by that id.
- c. DataCenterController notifies the ReverseFillingVm LoadBalancer of the new allocation
- d. ReverseFillingVmLoadBalancer updates the allocation table accordingly.

If not found:

- e. The ReverseFillingVmLoadBalancer returns -1.
- f. The DataCenterController queues the request
6. When the VM finishes processing the request, and the DataCenterController receives the response cloudlet, it notifies the ReverseFillingVmLoadBalancer of the VM de-allocation.
7. The DataCenterController checks if there are any waiting requests in the queue. If there are, it continues from step 3.
8. Continue from step 2.

#### 5. EXPERIMENTAL SETUP

The proposed algorithm is implemented through simulation packages like CloudSim 3.0. [11] CloudSim is run on the Java Net Beans interface. The CloudSim consists of a large number of classes like Cloudlet, Scheduler, Data Center, Hosts, etc. It provides a simulated environment for performing research on virtual machines. Java language is used for implementing VM load balancing algorithm. [12]Assuming the application is deployed in two data center's having 20 virtual machines (with 1024MB of memory in each VM running on physical processors capable of speeds of 250 MIPS) and Parameter Values are as given in Table 1:

Table 1: Parameter Value

Parameter	Value
Data Center OS	Window 7
VM Memory	10000 MB
Data Center Architecture	X86
Service Broker Policy	Optimize Cost & Time
VM Bandwidth	1000
Users	2

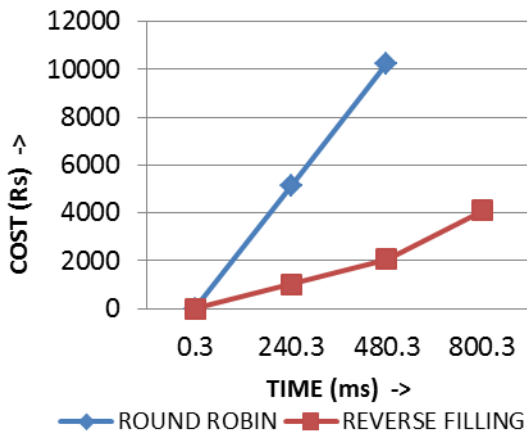
Table 2 shows the experimental results based on Reverse Filling VM Load Balancing Algorithm:

Table 2: Results of Round Robin and Reverse Filling VM Load Balancing Algorithm

Values	Round Robin	Reverse Filing Algorithm
Arrival Time	0.3 ms	0.3 ms
Finish Time	480.3 ms	800.3 ms
Cloudlets	40	40
Virtual Machines	20	20
Response Time	160.5 ms	160.5 ms
Minimum Response Time	20ms	20 ms
Processing Elements	1	1
DataCenters	2	2
File / Output Size	300	300
Cost Of Processing	3.0	3.0
Cost Per Memory	0.05	0.05
Cost Per Bandwidth	0.1	0.1
Cost Per Storage	0.1	0.1
Debt/ Total Cost	10256	4102.4

Figure 1 below highlights the graphical comparison between the existing Round Robin load balancing algorithm and proposed

Reverse Filling load balancing algorithm on the time taken for the task to complete and the cost incurred in its completion.



**Figure 1: Graphical comparison of the existing Round Robin and proposed Reverse Filling algorithm.**

Thus, it is observed that the implementation of the proposed algorithm yields lesser cost and response time than the existing approach. It also reduces the amount of the overheads incurred in the system. Thereby reducing the waiting time in the system and increasing the efficiency of the system providing more services to users or clients.

## 6. CONCLUSION AND FUTURE WORK

Load Balancing is the most important aspect of cloud computing. It helps in reducing the dynamic workload across the all nodes for the achievement of higher user and resource satisfaction. It helps in reducing the overhead, response time and increasing scalability. This paper depicts the cloud computing in detail. It basic fundamentals of the cloud and its characteristics like on-demand access, resource provisioning, etc. Secondly, it elaborates the fundamental concept of load balancing defining its goals and challenges. Thirdly, it presents the review of the some of the most prominent load balancing algorithms in the system. Then, it provides the Existing Work as well as the proposed work on the simulated environment. Lastly the experimental setup used for the creation of the virtual system and the resources used are specified in the system along with the graphical and tabular results of the algorithms. Thus, the proposed algorithm reduces the waiting time of the machines along with the overheads in the system resulting in lesser cost than existing approach. The future work includes modifications in the system leading to greater resource utilization and performance.

## 7. REFERENCES

[1] Bhatiya Wickremasinghe, Rodrigo N. Calheiros, and Raj Kumar Buyya, "CloudAnalyst: A CloudSim-based Visual Modeller for Analysing Cloud Computing Environments and Applications", in 2010 24th IEEE

International Conference on Advanced Information Networking and Applications (2010)

[2] Divya Chaudhary, Prof. Rajender Singh Chhillar, "Strategic Evaluation of Load Scheduling Techniques in Cloud Computing: A Review", Volume-II, NCACT-2013, DCSA, M.D University, Rohtak, 548-552.

[3] F el. M a, Feng L1' U and Zhen L1' U, "Distributed Load Balancing Allocation of Virtual Machine in Cloud Data Centre", 978-1-4673-2008-5/12/\$31.00 ©2012 IEEE

[4] Shu-Ching Wang, Kuo-Qin Yan, Wen-Pin Liao, Shun-Sheng Wang, "Towards a Load Balancing in a Three-level Cloud Computing Network", 2010 IEEE, pp. 108-113.

[5] Jinhua Hu, Jianhua Gu, Guofei Sun, Tianhai Zhao, "A Scheduling Strategy on Load Balancing of Virtual Machine Resources in Cloud Computing Environment", 3rd International Symposium on Parallel Architectures, Algorithms and Programming

[6] A.KHIYAITA, EL BAKKALI, M.ZBAKH, Dafir EL KETTANI, "Load Balancing Cloud Computing: State of Art", IEEE Transactions on Software Engineering.

[7] Rashmi. K. S, Suma. V, Vaidehi. M, "Enhanced Load Balancing Approach to Avoid Deadlocks in Cloud", Special Issue of International Journal of Computer Applications (0975 – 8887) on Advanced Computing and Communication Technologies for HPC Applications - ACCTHPCA, June 2012

[8] Martin Randles, David Lamb, A. Taleb-Bendiab, "A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing", 2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops, pp. 551-556.

[9] Soumya Ray and Ajanta De Sarkar, "EXECUTION ANALYSIS OF LOAD BALANCING ALGORITHMS IN CLOUD COMPUTING ENVIRONMENT", International Journal on Cloud Computing: Services and Architecture (IJCCSA), Vol.2, No.5, October 2012 DOI : 10.5121/ijccsa.2012.2501 1

[10] Bhaskar. R, Deepu. S.R and Dr. B.S. Shylaja, "DYNAMIC ALLOCATION METHOD FOR EFFICIENT LOAD BALANCING IN VIRTUAL MACHINES FOR CLOUD COMPUTING ENVIRONMENT", Advanced Computing: An International Journal ( ACIJ ), Vol.3, No.5, September 2012 DOI : 10.5121/acij.2012.3506 53

[11] Calheiros Rodrigo N., Rajiv Ranjan, César A. F. De Rose, Raj Kumar Buyya (2009): CloudSim: A Novel Framework for Modeling and Simulation of Cloud Computing Infrastructures and Services CoRR abs/0903.2525: (2009)

[12] CloudSim: A Framework for Modeling and Simulation of Cloud Computing Infrastructures and Services, The Cloud Computing and Distributed Systems (CLOUDS) Laboratory, University of Melbourne, (2011) available from:<http://www.cloudbus.org/clouds>.