

A Comparative Study of Load Balancing Algorithms for Cloud Computing

Harsha M B^{*}, Dr. Sarojadevi H^{**}

^{*}PG Student, Department of Computer Science and Engineering, NMAMIT, Nitte

Email: harshamb21@gmail.com

^{**} Professor and Head, Department of Computer Science and Engineering, NMAMIT, Nitte

Email: sarojadevi@nitte.edu.in

Abstract

Cloud Computing is fast growing technology in both industry research and academy. User can access the cloud service and pay based on the usage of resource. Balancing the load is major task of cloud service provider with minimum response time, maximum throughput and better resource utilization. There are many load balancing algorithms proposed to assign a user request to cloud resource in efficient manner. In this paper three load balancing algorithms are simulated in Cloud Analyst and results are compared.

Keywords: Cloud Computing, Cloud Analyst, Virtual Machine, Load Balancing Policy, Data Center

I. INTRODUCTION

Cloud computing is an innovative technology, which enables the user to access the software and service over the network. User can access super computer power in cloud computing. The key concept of cloud computing is that user does not buy the hardware or software. They rent the computation power, application and pay according to use. The services provided by cloud computing are mainly divided into three basic service models as Software as a service (SaaS), Platform as a service (Paas), Infrastructure as a service (IaaS). The cloud computing domain is divided into four categories such as Private, Public, Hybrid and Community Cloud [1].

- Software as a service (SaaS) model, where user can request for particular software and have to pay based on their usage time. An example for SaaS is Google Docs.
- Platform as a service (Paas) model, where complete resource needed to develop and deploy an application into cloud is provided as service. An example for Paas is Google App Engine.
- Infrastructure as a service (IaaS) model, where infrastructure is like a virtual machine server, memory and storage being provided as service. An example for IaaS is Amazon Elastic Compute.

Figure 1 shows the three important factors of Cloud computing such as characteristics, deployment model and service model. Cloud computing is a fastest growing technology; the number of users of cloud computing service are increasing rapidly. To balance the load is the major task of cloud service provider.

The primary objective considered here is efficient utilization of resources i.e., resource should not be over utilized and underutilized.

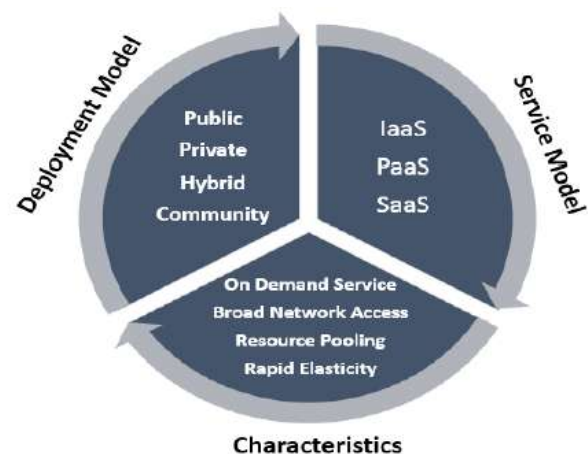


Figure 1: Cloud Computing

II. LOAD BALANCING

Load Balancing is a method to distribute workload on the multiple computers or a computer cluster through network links to achieve optimal resource utilization for maximizing throughput and minimizing overall response time. As the number of users in cloud computing technology is increasing very rapidly, it is becoming very difficult to manage and fulfill the user requirements. So load balancing has come into picture to spread the user request to all the resources (virtual machine in data center) without under or over utilization. Load balancer is not only used in cloud but it is also used in database and website. Load balancing in cloud computing is done in two

levels, which are Virtual Machine (VM) level and Host level.

VM level: It is to map between applications in uploading on cloud and virtual machine.

Host level: Mapping is done between virtual machine and host resources.

2.1 Classification of load balancing algorithm

Load balancing algorithms in cloud computing are classified based on state and behavior of the system. There are two types, named as Static and Dynamic load balancing algorithm [2].

Static Load Balancing Algorithm

In static algorithm load does not depend on the current state and behavior of the system, but requires a prior knowledge about the resource and application. It divides the traffic equally among all available server or virtual machine. It aims to minimum execution time and limits communication overhead and delay. This algorithm has drawback of low reliability and no fault tolerance. Examples of static load balancing algorithms are Round robin and Random load balancing algorithm.

Dynamic Load Balancing Algorithm

Dynamic algorithm is more flexible than static algorithm. It depends on the current state and behavior of the system, but does not require prior knowledge about the resource. It divides the traffic according to capacity of all server and virtual machine. It aims to reduce the fault tolerance and improve reliability. This algorithm has drawback of less stable and high utilization of resource. Examples of dynamic load balancing algorithms are Active monitoring and Throttle load balancing.

III. EXISTING LOAD BALANCING IN CLOUDANALYST

3.1 Round Robin Load Balancing Algorithm

Round robin is one of the most simple and straight forward scheduling algorithm and it uses the principle of time slice. In Round robin each VM is given a fixed quantum of time for execution of user request (Figure 2). If the user request is not completed in quantum of time VM will wait for next round to complete the remaining operation. Drawback of this algorithm is without checking the capacity of server or resource it allocates the request. Also the algorithm will not consider the state and behavior of previous allocation of VM.

V. Darji et al., [3] proposed an algorithm in their paper "Dynamic Load Balancing For Cloud Computing Using Heuristic Data and Load on Server". In this algorithm a dynamic strategy is used, in which weights are assigned based on continuous monitoring

of the servers and is therefore constantly changing. Thereby it achieves optimum resource utilization

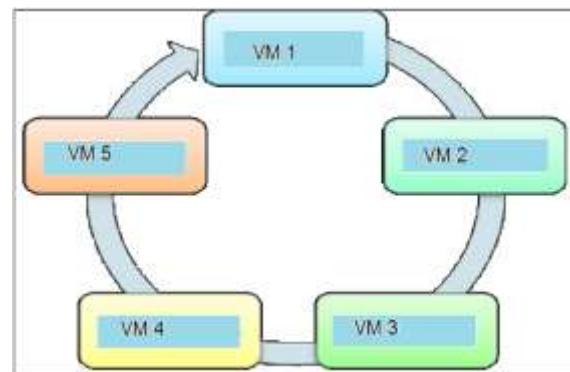


Figure 2: Round Robin Load Balancing

3.2 Active Monitoring Load Balancer

Active load balancing algorithm is a dynamic approach, shown in Figure 3. It keeps the information about each virtual machine and the number of requests allocated to it. Using this information next request is allocated to virtual machine. When request arrives it checks the information table to know which virtual machine is least loaded and then allocates the request. Drawback of this algorithm is that it does not check whether virtual machine is previously utilized or not.

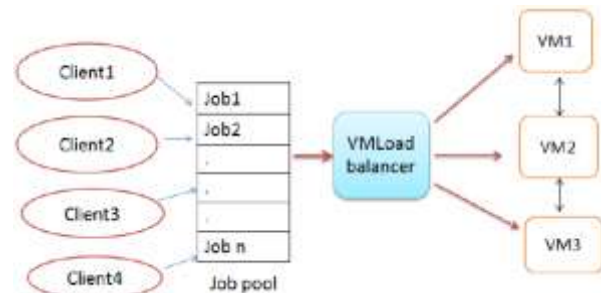


Figure 3: Active Monitoring Load Balancing

S.G. Damanal et al., [4] proposed an algorithm in their paper "Optimal Load Balancing in Cloud Computing by Efficient Utilization of Virtual Machines", which provides a solution, based on least loading principle, to improve response time and efficient utilization of available resources.

3.3 Throttled load balancing algorithm

Throttled load balancing algorithm is also dynamic in nature, in which load balancer is used for allocating the request to available virtual machine as shown in Figure 4. The Load balancer maintains index table which consists of the information about the state of virtual machine which could be Available or Busy. When the request arrives load balancer checks which virtual machine is available to execute the user request and allocates for it. Drawback of this algorithm is that it checks table from first index every

time to allocate. Virtual machines available at the bottom of the table will not be utilized if there are entries available on top.

S. G. Damanal et al. [5] proposed an algorithm “Load Balancing in Cloud Computing Using Modified Throttled Algorithm”, which focuses mainly on how incoming jobs are assigned to the available virtual machines intelligently.

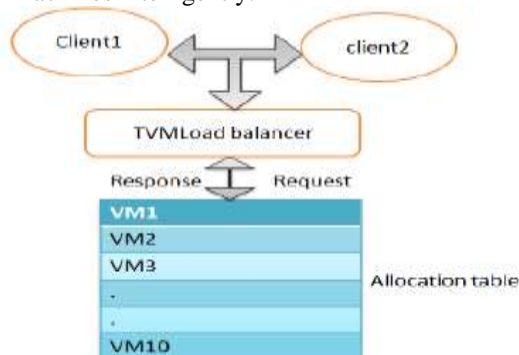


Figure 4: Throttled Load Balancing

IV. CLOUDANALYST SIMULATOR

The Cloud Analyst [6] is built on the top of the CloudSim tool kit [7], a block diagram of which is shown in Figure 5. It has support graphical user interface which enables provision for description of location of datacenter, generating traffic, number of users, and resource of each data center. In this simulator we can do high level configuration such as internet configuration of each datacenter and traffic. The simulation experiment set up exercise is explicitly different from a programming exercise. A modeler is enabled to concentrate on the simulation parameters rather than the techniques of programming. It also enables the modeler to execute simulations repeatedly with modifications to the parameters quickly and easily.

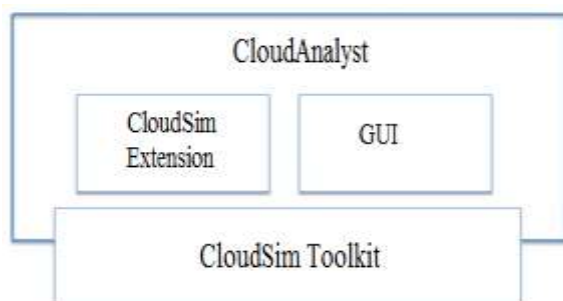


Figure 5: Cloud Analyst

Feature of Cloud analyst simulator

Features of Cloud analyst simulator are listed below.

- Ease of setting up and executing a simulation experiment.

- Ability to define a simulation with a high degree of configurability and flexibility
- Graphical output in the form of tables and charts is highly useful to summarize the potentially large amount of statistics that is collected during the simulation.
- Ability to save the simulation setup (the set of input parameters) as a file and allowed to load the file for further simulation and also be able to save the results of an experiment as a file.
- Ease of extension which means we can evolve the existing load balancing policy with minimal effort with suitable configuration.

V. SIMULATION RESULT AND COMPARISON

Simulation configuration is given in Table 1. Same configuration is used for the simulation of the three load balancing policies. Data center (DC) configuration consists of operating system, virtual machine manager, memory, storage, bandwidth, number of processes and VM policy should be defined under datacenter configuration tab. User base (traffic generator), number of virtual machines in each data center and service broker policy should be defined under configuration tab and snapshot of cloud analyst as shown in Figure 6.



Figure 6: Snapshot of Cloud Analyst

Table 1: Simulation parameter

Parameter	Value Used
Request Per User Per Hour	60
Data Size Per Request	100
Number of region use	6
Peak hour start(GMT)	3
Peak hour end (GMT)	9
DC 1– No Of VM	5
VM Image Size	10000 MB
VM Memory	512 MB
VM Bandwidth	1000 bps
DC 1– Storage Per Machine	100000000 Mb

DC 1– Available BW Per Machine	1000000
DC 1– No Of Processors Per Machine	4
DC 1– Processor Speed	10000 MIPS
DC 1– VM Policy	Time Shared
User Grouping Factor	10
Request Grouping Factor	10
Executable Instruction Length	100
Service broker policy	Closest DC

Results of three load balancing algorithms containing response time of request and data center processing time is presented in Tables 2, 3 and 4 respectively. Average response time is calculated as a ratio of total response time of requests over total number of requests. Average data center processing time is calculated as ratio of total processing time of requests in the data center over total number of requests.

Table 2: Result of Round Robin

	Avg(ms)	Min(ms)	Max(ms)
Overall Response time	309.47	218.65	402.81
Data center processing time	6.46	0.09	28.21

Table 3: Result of Active Monitoring

	Avg(ms)	Min(ms)	Max(ms)
Overall Response time	309.20	218.65	402.81
Data center processing time	6.18	0.09	27.32

Table 4: Result of Throttled Load Balancing

	Avg(ms)	Min(ms)	Max(ms)
Overall Response time	303.55	218.65	402.81
Data center processing time	3.41	0.09	16.86

VM usage of the three load balancing algorithms for execution of user requests is shown in Table 5. In round robin algorithm requests are distributed equally to virtual machines, as it does not consider the state and behavior of virtual machine. Active monitoring and Throttled load balancing policies show imbalance of load distribution, which may be due to the dynamic nature of the algorithms. VMs are over and underutilized in Active and Throttled load balancers.

Table 5: Virtual Machine usage

Sl. No	Round Robin Load balancer	Active Load Balancer	Throttled Load Balancer
VM0	254	1174	1182
VM1	254	78	76
VM2	254	10	8
VM3	254	4	2
VM4	254	2	0

VI. EXTENSION TO CLOUDANALYST

Any extension to cloud analyst pertaining to load balancing is proposed here. The classes and methods need to be changed are shown in Table 6.

Table 6: Classes and methods for extension

Package	Class	Method
cloud-sim.ext.gui.screens	ConfigureSimulationPanel	createMainTab
cloudsim.ext	Constants	-
cloud-sim.ext.datacenter	datacenter-controller	-
cloud-sim.ext.datacenter	Add new class Newloadbalancer	-
cloud-sim.ext.datacenter (for new policy)	Newloadbalancer	getnextavailableVm

VII. CONCLUSION

The main role of a load balancing algorithm in cloud computing is to ensure efficient utilization of cloud resource, minimum response time, better client satisfaction and increased performance of cloud service. In this paper we present comparative study of load balancing algorithms for cloud computing. The simulations are done in Cloud Analyst which is built on Cloudsim. We also mention the approaches to improve the performance of three load balancing policies. In this paper we bring out the importance of load balancing in cloud computing which helps in better resource utilization to gain high performance and better response time of a system.

REFERENCES

- [1] Harjit Singh, "Current Trends in Cloud Computing Survey of cloud computing system", 2009.
- [2] Reena Panwar, Bhawna mallick."A Comparative of Load Balancing Algorithm In Cloud Computing" IJCA (0975 – 8887) Volume 117 – No. 24, May 2015
- [3] Darji, Vinay, Jayna Shah, and Rutvik Mehta. "Dynamic Load Balancing For Cloud Com-

- puting Using Heuristic Data and Load on Server." (IOSR-JCE) Volume 16, Issue 4, Pages 59-69, Aug. 2014
- [4] Shridhar G.Domanal and G. Ram Mohana Reddy. "Optimal load balancing in cloud computing by efficient utilization of virtual machines" COMSNETS 2014, Page(s): 1- 4, 2014
- [5] Shridhar G.Domanal and G.Ram Mohana Reddy, "Load Balancing in Cloud Computing Using Modified Throttled Algorithm", IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), 2013
- [6] W. Bhatiya, 2009. Cloud Analyst a Cloud-Sim-based tool for modelling and analysis of large scale cloud computing environments. MEDC Project, Computing and Distributed Systems Laboratory, University of Melbourne, Australia, pp. 1-44, June 2009.
- [7] Rajkumar Buyya, Rodrigo N. Calheiros, Rajiv Ranja, Anton Beloglazov, A. F. De Rose, "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms", 2010.