Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

International Conference of Trends in Information, Management, Engineering and Sciences (ICTIMES)
Volume 5, Special Issue 02, Feb.-2018 (UGC Approved)

Evaluation Based Load Balancing for Cloud Computing

[1]Dr.Sunil Tekale, [2]Mr.M.Amarnath

[1]*Professor Department of CSE MRCE Hyderabad, Telangana*
[2]*Assistant,Department of CSE MRCE Hyderabad, Telangana*

Abstract- *Cloud computing is the best technology today for all those people who wants to go with minimum investment on infrastructure and wants to outsource the burden of handling technical issues to third party by paying the charges for the services utilized .Today there is huge amount of demand from the clients to make use of cloud technology as it provides multiple features and take off the load of maintaining infrastructure. This has created a huge amount of load on servers . So it is must to handle issues related to load balancing. This is basically to see that the load on a particular server is kept maximum to its threshold level. So that it can handle the task and also can complete it in a faster manner. It minimizes the cost and time involved in the major computational models and helps to improve proper utilization of resources and system performance. Many algorithms are recommended by various researchers from all over the world to solve the problem of load balancing. In this paper, we present a new algorithm named as combo algorithm to address the issue of load balancing in a cloud environment.*

Keywords *- Cloud Computing optimization Load Balancing Network*

1. Introduction

Cloud computing is a newly progressing technique which offers online computing resources, storage and permits users to organize applications with enhanced scalability, availability and fault tolerance. Cloud computing is about storing the stuff on remote servers instead of on own computers or other devices.

This information can be retrieved using the internet with any device, everywhere in the world as long as that device can support cloud computing systems. The cloud computing system is comprised of a front-end, which is the client side and a back-end which is a collection of the servers and computers owned by a third party which stores the data. A central server which is a fragment of the back-end follows protocols and uses middleware to communicate between networked computers. Cloud computing accumulates all the computing resources and manages them automatically Its characteristics

describe a cloud computing system: on-need self-service, pooling of resources, access to the internet, the elasticity of service availability and measurement of services utilized by individual users. Cloud computing is everywhere with tools like Google Drives replacing Microsoft Office, Amazon Web Services replacing traditional enterprise data storage, banking websites replacing branch offices and Dropbox storing all our data and files. The cloud even provides different deployment models and service models.
The four deployment models present in cloud computing are: [2]

1. Public cloud: In the public cloud, the cloud provider provides resources for free to the public. Any user can make use of the resources; it is unrestricted. The public cloud is connected to the public internet for anyone to leverage.

2. Private cloud: In a private cloud, the planning and provisioning of the cloud are operated and owned by the organization or the third party. Here the hosted services are provided to a restricted number of people or group of inividuals.

3. Community cloud: These type of cloud infrastructures exists for special use by a group of users. These are a group of users who share a common mission or have specific regulatory requirements, and it may be managed by the third party or organizations.

4. **Hybrid Cloud:** Hybrid Cloud provides the best of above worlds. It is created by combining the benefit of different types of cloud (private cloud & public cloud). In these clouds, some of the resources are provided and managed by public cloud and others as a private cloud.

The three different service models present in cloud computing are:

1.Infrastructure as a Service (IaaS): IaaS model provides just the hardware and the network. It allows users to develop and install their operating system, software and run any application as per their needs on cloud hardware of their own choice.

1. Platform as a Service (PaaS): In PaaS model, an operating system, hardware, and network are provided to the user. It enables users to build their applications on cloud making use of supplier specific tools and languages

2. **Software as a Service (SaaS):):** In SaaS model, a pre-built application together with any needed software, hardware, operating system and the network is provided to the user.

2. Load Balancing

Load balancing is a serious concern in cloud computing. With the increase in attractiveness of cloud computing among users, the load on the servers and the quantity of processing done is surging drastically. There are multiple nodes in the cloud, and due to the random allocation of a request made by the client to any node, the nodes become unevenly loaded.

So to avoid the condition where some nodes are either severely loaded or under loaded, the load balancer will evenly divide the workload among all the nodes [3]. Thus load balancing will equally distribute the workload among the nodes, and it can help in minimizing delays in communication, maximizing the throughput, minimizing execution time and maximizing resource utilization [3].

2.1 Goals of load balancing:

Some of the key purposes of a load balancing algorithm as pointed out by are:

1. It should possess fault tolerance.
2. It should be capable of modifying itself according to any change or expansion in the distributed system configuration [3].
3. Regarding system performance, it should give greater overall improvement at a minimal cost.
4. Regardless of the origin of job it must treat all jobs in the system equally.
4. It should also maintain system stability.

2.2 Issues of Load Balancing

The issues of load balancing are described below4.

1. Load balancing becomes critical because, in the middle of execution, the processes may shift amongst nodes to ensure equal workload on the system [5].

2. For a load balancing scheme to be good it should be scalable, general and stable and should add minimal overhead to the system. These requirements are interdependent [6].

3. One of the critical aspects of the scheduling problem is load balancing [7]. The challenge for a scheduling algorithm is to avoid the conflict between prerequisites: fairness and data locality.

4. Algorithms for load balancing have to be dependent on the hypothesis that the on hand information at each node is accurate to avoid processes from being continuously circulated the system without any progress[5].

5. How to accomplish a balance in load distribution amongst processors such that the computation can be done in the minimum possible time is one of the important problems to resolve.

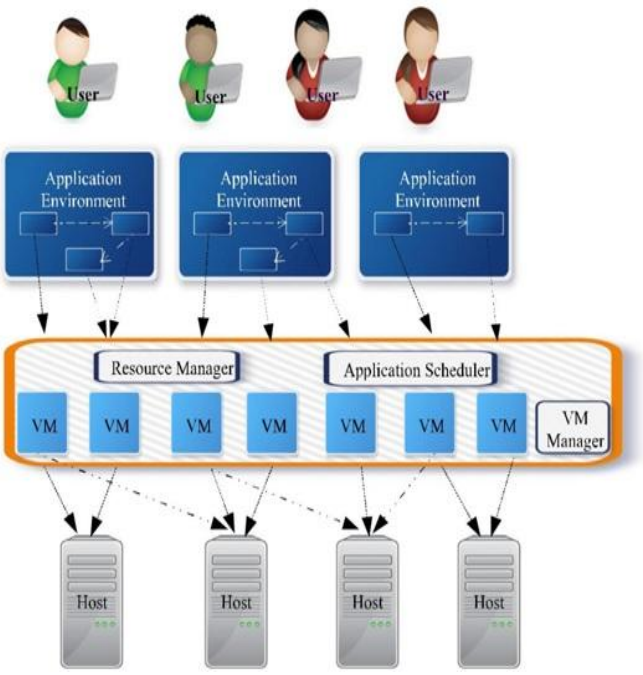
6. Load balancing and task scheduling in distributed operating systems is a vital factor in gross system efficiency because t he distributed system is not pre- emptive and non-uniform, that is, the processors may be different [7].

2.3 Components of Load Balancing Algorithms:

A load balancing algorithm has five major components [8]

1. Transfer Policy: The portion of the load balancing algorithm that picks a job for moving from a local node to a remote node is stated as Transfer policy or Transfer strategy.
2. Selection Policy: In this policy, it specifies the processors involved in the load exchange (processor matching) so that the overall response time and throughput may be improved.
3. Location Policy: The portion of the load balancing algorithm that is responsible for choosing a destination node for a task to transfer is stated as location policy or Location strategy.
4. Information Policy: The part of the dynamic load balancing algorithm that is in charge of gathering information about the nodes present in the system is started to as Information Policy or Information strategy.
5. Load Estimation Policy: In this policy, it determines the total workload of a node in a system.

Fig:1 shows the VM,Application,Host relationship in DC



Benefits of Cloud Load Balancing

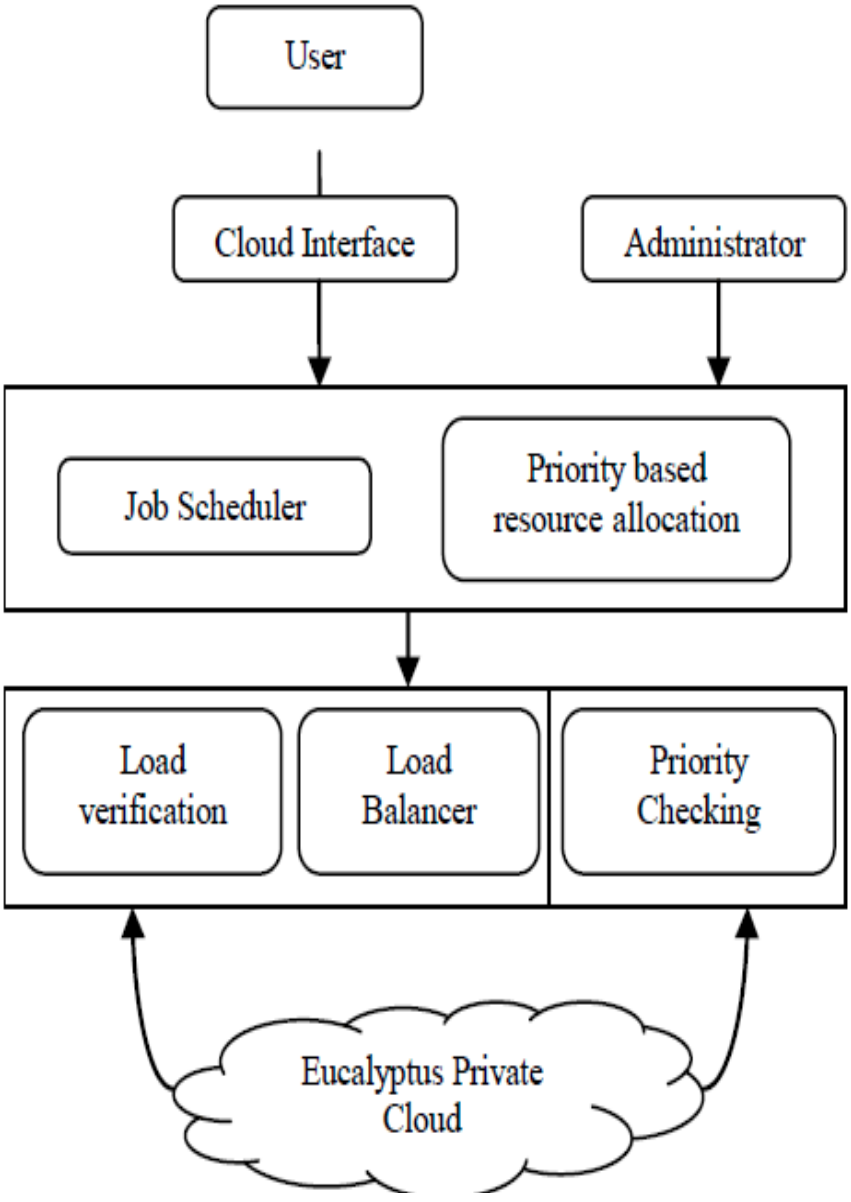
The benefits of cloud load balancing in particular arise from the scalable and global character of the cloud itself. The ease and speed of scaling in the cloud means that companies can handle traffic spikes (like those on Cyber Monday) without degraded performance by placing a cloud load balancer in front of a group of application instances, which can quickly autoscale in reaction to the level of demand. The ability to host an application at multiple cloud hubs around the world can boost reliability.

If a power outage hits the northeastern U.S. after a snowstorm, for example, the cloud load balancer can direct traffic away from cloud resources hosted there to resources hosted in other parts of the country.

4. Proposed Algorithm.

The Evaluation based load balancing can be done in a much more easier and simplest manner with high efficiency. In this we have to just calculate the capacity of each server in terms of its memory, computing speed and so on. Based on the evaluation of this we can give a score in terms of marks on a scale of 100. So all data centers will have server and each server will have a score based on the evaluation of its capacity.

Fig:2 Represents the proposed load balancing.



The centralized control system will allocate the task that comes to the system by evaluating the amount of resources required in a tentative manner and based on that its allocates the server with either high or low scoring. The maximum the resources are required the server with highest score will be allocated and where ever the low resources are required as server with less score will be assigned or allocated. With this an advantage is the work or the task that comes will be given

to appropriate server and there by the task completes faster and efficiently. The server then will be ready to take up the next task.

All server will also have threshold level and this is also taken into account for allocation of task. The threshold level will help us to know the amount of load that needs to be assigned to that server, once the server reaches that level the other server with next level of score will be given the task.

The sum of loads of all virtual machines is defined as

$$L = \sum_{i=1}^{k} l_i,$$

where i represents the number of VMs in a data center.

The load per unit capacity is defined as

$$\text{LPC} = \frac{L}{\sum_{i=1}^{m} c_i}$$

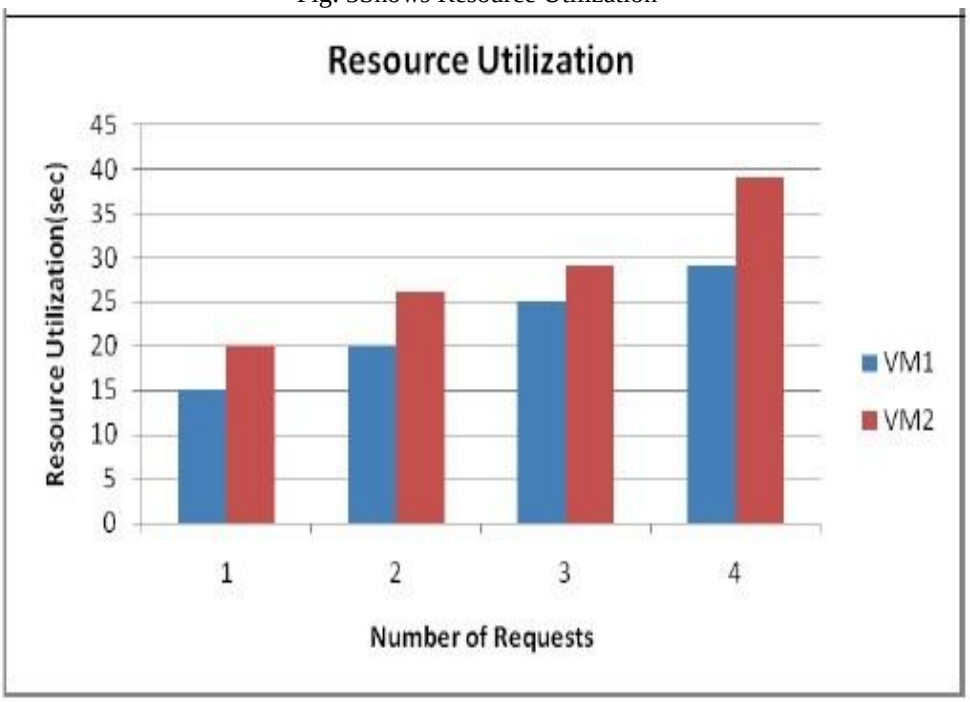
$$\text{Threshold } T_i = \text{LPC} * c_i,$$

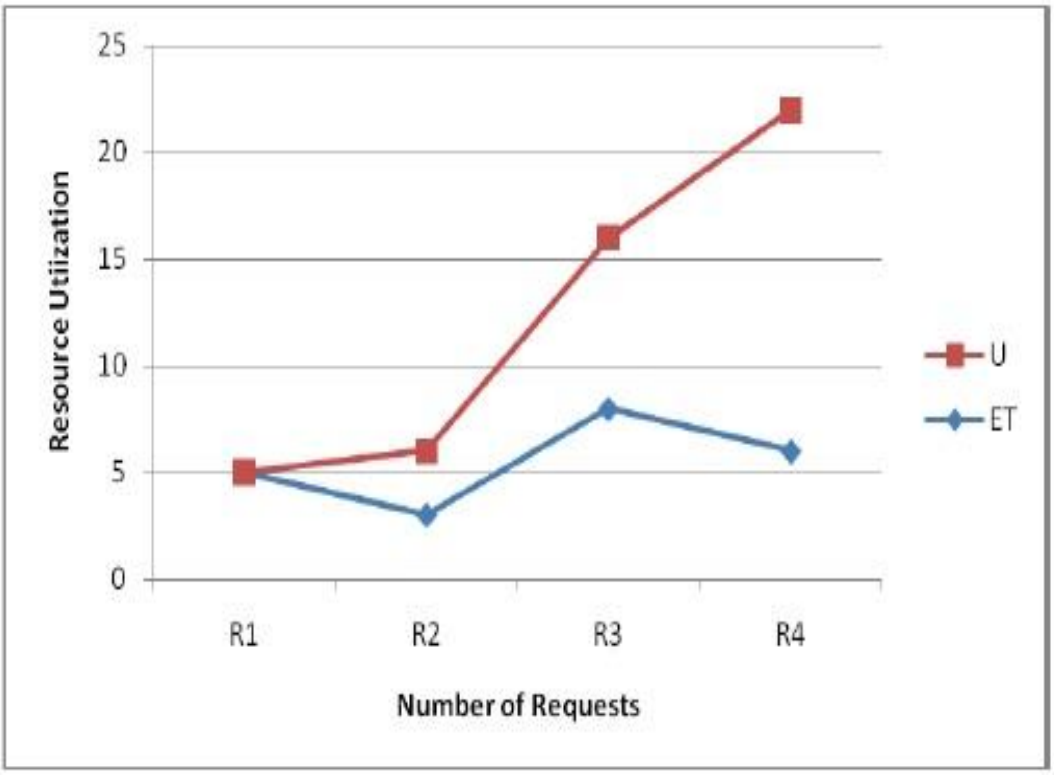
where c_i is the capacity of the node.

The load imbalance factor of a particular virtual machine is given by

$$\text{If VM} \begin{cases} < \left| T_i - \sum_{v=1}^{k} L_v \right|, & \text{Underloaded,} \\ > \left| T_i - \sum_{v=1}^{k} L_v \right|, & \text{Overloaded,} \\ = \left| T_i - \sum_{i=1}^{k} l_i \right|, & \text{Balanced.} \end{cases}$$

Fig: 3Shows Resource Utilization





Simulation toolkits
Thinking of unpredicted network environment and laboratory resource scale (like servers), sometimes it is helpful to and more convenient for developing and running simulation tools to simulate large-scale experiments. The research on dynamic and large-scale distributed environment can be fulfilled by constructing data center simulation system, which offers visualized modeling and simulation for large-scale applications in cloud infrastructure.9

The data center simulation system can describe the application workload statement, which includes user information, data center position, the amount of users and data centers, and the amount of resources in each data center.10 Under the simulated data centers, load balancing algorithms can be easily implemented and evaluated.

CloudSim: CloudSim is an event-driven simulator implemented in Java. Because of its object-oriented programming feature, CloudSim allows extensions and definition of policies in all the components of the software stack, thereby making it a suitable research tool that can mimic the complexities arising from the environments.11

CloudSched: CloudSched enables users to compare different resource scheduling algorithms in Infrastructure as a Service (IaaS) regarding both hosts and workloads. It can also help the developer identify and explore appropriate solutions considering different resource scheduling algorithms.9

Conclusion:

Load Balancing is a necessary task in Cloud Computing environment to attain maximum use of resources. In this paper, we talk about Evaluation based load balancing method which helps in providing maximum efficiency in terms of execution and allocation of various tasks. The advantage of this is algorithm is most sought after resources task will get the maximum scored server to perform computation and least resource required task gets the least scored server. As high rated server is performing high rated task and low rated server is performing low rated task the resource utilization will happen in most appropriate manner.,

References

1 D. Saranya et.al, "Load Balancing Algorithms in Cloud Computing: A Review," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 5, Issue 7, July 2015.
2 S. Sethi et.al, "Efficient Load Balancing in Cloud Computing using Fuzzy Logic," IOSR Journal of Engineering (IOSRJEN) ISSN: 2250-3021 vol. 2, pp. 65-71, July 2012.
3 T. Desai et.al, "A Survey of Various Load Balancing Techniques and Challenges in Cloud Computing," International Journal of Scientific & Technology Research, vol. 2, Issue 11, November 2013.
4 S. Rajoriya et.al, "Load Balancing Techniques in Cloud Computing: An Overview," International Journal of Science and Research (IJSR), vol. 3, Issue 7, July 2014
5 Sharma S. et.al, "Performance Analysis of Load Balancing Algorithms," World Academy of Science, Engineering and Technology, 38, 2008.
6 Gross D. et.al, "Non cooperative load balancing in distributed systems", Elsevier, Journal of Parallel and Distributed Computing, No. 65, pp. 1022-1034, 2005.
7 Nikravan M. et.al, "A Genetic Algorithm for Process Scheduling in Distributed Operating Systems Considering Load Balancing", Proceedings 21st European Conference on Modelling and Simulation (ECMS), 2007.
8 M. Amar et.al, "SLA Driven Load Balancing for Web Applications in Cloud Computing Environment",
9.Tian W, Xu M, Chen A, et al. Open-source simulators for cloud com-puting: Comparative study and challenging issues. Simul Modell Pract Theory. 2015;58:239–254.
10.Calheiros RN, Ranjan R, Beloglazov A, De Rose CA, Buyya R. Cloudsim: A toolkit for modeling and simulation of cloud computing environ-ments and evaluation of resource provisioning algorithms. Software: Pract Exper. 2011;41(1):23–50.
11.Tian W, Zhao Y, Xu M, Zhong Y, Sun X. A toolkit for modeling and simu-lation of real-time virtual machine allocation in a cloud data center. IEEE Trans Autom Sci Eng. 2015;12(1):153–161.