Volume 5, Issue 1, January 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

# Survey on Load Balancing Approaches in Cloud Computing Environment

**Rajyashree*, Vineet Richhariya**
Lakshmi Narain College of Technology,
Bhopal, India

***Abstract— Cloud computing is a new emerging technology in the era of IT industries. It is a business oriented model which provides the computing resources to the client. It has become famous in short time because of their attractive services like easy to use, pay as use and accessibility of their services throughout the world etc. Resource utilization is the main challenging task in the cloud. Virtualization technology is used to increase the resource utilization, which increase the resource utilization through the sharing of the physical resources. Through the virtualization multiple users can share the same physical resources. Although virtualization increased the resource utilization, but add new issue that is load balancing. Since resource required by the VM can be changed dynamically, so load balancing is a challenging task in the cloud as compare to the cluster computing. Lot of work have been done in the load balancing. In this paper we compare some exiting load balancing approach with their anomalies.***

***Keywords— Distributed Computing, On Demand Resources, Cloud Computing, Virtualization, Server Consolidation, Load Balancing***

## I. INTRODUCTION

Cloud computing is emerging as a new paradigm for next generation computing in the field of computer science and information technology because of their attractive services such as adaptive, online, value added and pay as use scheme. Cloud can be defined in a number of ways. It is a business model, which provides the on demand hardware and software as services to the client through the internet [1]. According to NIST cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources. Theses computing resources include networks, servers, storage, applications, and services. This cloud model is basically composed of five essential characteristics, three types of service models, and four deployment models.

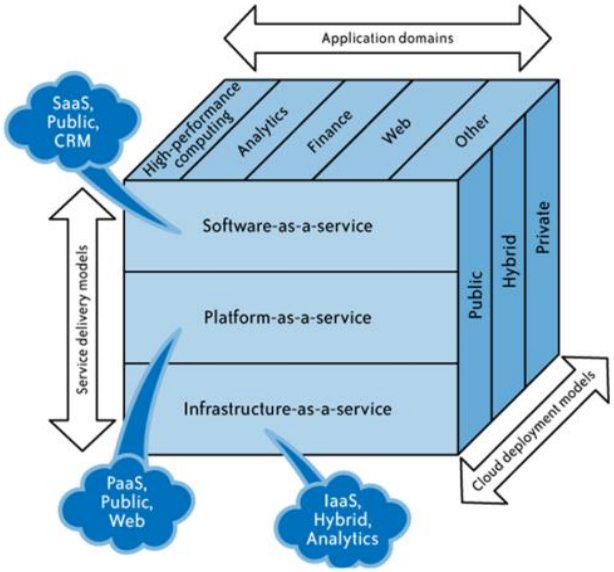


Fig. 2 Cloud Computing Models in 3D

Cloud computing provide three type of services i.e. software as a service (SAAS), platform as a service (PAAS), and infrastructure as a service (also known as hardware as a service) [2,3].
Software as a services mainly deliver the online software application to the client of cloud computing. It is responsible to provide capability to use cloud applications in a cloud infrastructure which is supplied by the cloud service provider. The applications are accessible from various client using computing devices like a thin or thick clients interface such as a web browser. The users do not have permission to manage or control the underlying cloud infrastructure including hardware resources or platform infrastructure etc. Gmail and Facebook is one of the most famous cloud applications.

 

Platform as a Service (PaaS) gives the capability to create application services as on their desire. It allows users to develop their software using programming languages and tools supported by the provider. The user does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage. The user has control only over the deployed applications and application hosting environment configurations.

Infrastructure as a Service (IaaS) provides the capability to have control over complete cloud infrastructure with CPU processing, storage, networks, and other computing resources. The cloud user is able to deploy and run their software, which can include operating systems and other software applications as website.

There are four types of cloud deployment model [4] in the cloud computing known as public, private, community and hybrid cloud.

Private Cloud is a model of cloud computing whose framework is allowed to use with a particular organization. All the resources and services are dedicated to a limited number of peoples. The server and data center is also setup within organization. Sometimes infrastructure is setup by third party but it is in full control of organization. The private clouds are good to privacy and security.

Public cloud is model of cloud where all users are allowed to access the services using internet. The user need only internet connection and web browser to access with pay per use scheme. All the services with infrastructure of cloud provider are available on the internet. User need to subscribe the application and make enable to use it.

Community cloud includes number of organization to share their services to increase resource utilization of cloud infrastructure. The cloud infrastructure is not limited to only one organization.

Hybrid cloud combines both public and private cloud with their advantages. Hybrid cloud offers the benefits of both the public and private cloud. The hybrid cloud is the good solution for purely business oriented concept because many modern businesses have a wide range of concerns to support users demand.

Virtualization [5, 6] is the core technology in the cloud, which allows the sharing of the physical resources. With the help of virtualization single physical device can be share by the multiple users. When any user demands for the resources hypervisor or virtual machine monitor (VMM) create a VM and bind the requested resources with the VM. Virtualization can be classified in two type's i.e. Full virtualization and paravirtualization. Full virtualization is a technique in which a complete installation of one machine is run on another machine. In full virtualization, the entire system is emulated (BIOS, drive, and so on), but in paravirtualization, its management module operates with an operating system that has been adjusted to work in a virtual machine. Paravirtualization typically runs better than the full virtualization model, simply because in a fully virtualized deployment, all elements must be emulated. Virtualization can allow the creation of VM. So number of VM can be created in single host and each VM behave like a physical machine.

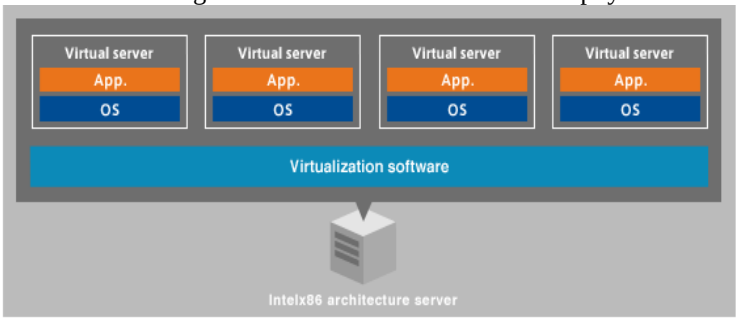


Figure 2- Virtualization

Virtualization increased the resource utilization, but performance may be decreased due to the sharing of the resources. So load balancing is the very important task in the cloud. Overall system performance depends on the efficient load balancing. An efficient load balancing approach can minimize the number of migration as well as energy consumption.

This paper, we discuss the overview of cloud computing. The goal of this paper is to provide a basic to the cloud with some load balancing approach. In section 2, we discuss the background knowledge of the load balancing approach with their anomalies.. Section 3 gives the conclusion of different exiting load balancing approach. Section 4 concludes the paper with the focus on the future work.

## II. RELATED WORK

Lots of works have been done in the field of load balancing. Generally lower and upper threshold are used to define the underloaded and overloaded host respectively. VM migration [2, 7, 8] is use to deal with the overloaded and underloaded condition. VM migration is the technique which move the VM from one host to the another host. When the load on the VM is below the lower threshold all VM running on that PM are move to the other PM. This situation is known as server consolidation.

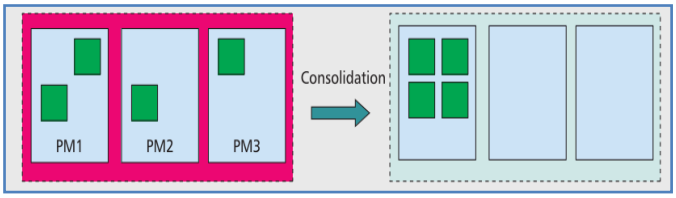


Figure 2: Server consolidation

 

In the case of overloaded some VM are move from the overloaded PM to the other PM. This process known as load balancing

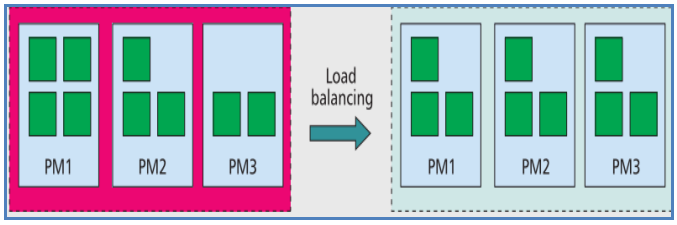


Figure 2: Load Balancing

Load balancing approach can be categorized in five category [9] i.e. static, dynamic, distributed, centralized and hierarchically.

***A. Static Load Balancing Approach***- In the static approach fixed threshold are used that can not changed with time to define a percentage of resources that can be used. Although static environment is easy to implement, but it is not effective for the cloud.

Round Robin algorithm [10] used the static threshold for the placement. In this approach resources are provided to the process on the basis of first come first serve (FCFS). Processes share the resources on the time sharing basis. Eucalyptus uses greedy (first-fit) with round-robin for VM mapping.

B. Radojevic et al. [11], proposed an improved round robin load balancing approach. It uses the basis of round robin but it also measures the duration of connection between client and server by calculating overall execution time of task on given cloud resource.

A. Beloglazov et al. [12], proposed an energy efficient load balancing approach. They argue that average power consumed by an idle server is 70% of power consumed by fully utilized server. So power consumed by the data center can be controlled by the proper load balancing approach. They used fixed lower and upper threshold with the difference of 40 between lower and upper threshold. So if lower threshold is 30 than upper threshold is 70. This approach reduced the number of migration but main problem with this approach is that they used the fixed value of lower and upper threshold.

***B. Dynamic Load Balancing Approach***- In these approaches threshold can be changed with time. Static load balancing approach is not suitable for the cloud, where user request can change with time.

T. Wood et al. [13], proposed an approach for the hot spot mitigation know as sandpiper. Sandpiper use black and gray box approach to monitor the host. They use the Xen hypervisor. The monitoring engine is responsible for tracking the processor, network and memory usage of each virtual server. It also tracks the total resource usage on each physical server by aggregating the usages of resident VMs. Problem with this approach is that they only consider the cpu load to calculate the load on host.

S. Wang et al. [14] proposed min-min load balance algorithm, which uses three level frameworks for resource allocation in dynamic environment. It uses OLB (opportunistic load balancing) algorithm as its basis. Since cloud is massively scalable and autonomous, dynamic scheduling is better choice over static scheduling.

***C. Centralized Load Balancing Approach***- In centralized load balancing technique single node is responsible for taking the decision regarding to the placement and all other task. This center node has all the knowledge about the other node and then apply static or dynamic approach for the placement. This approach is suitable for the small network, not for the cloud where the resources are distributed. This technique reduces the time required to analyze different cloud resources but creates a great overhead on the centralized node. Also the network is no longer fault tolerant in this scenario as failure intensity of the overloaded centralized node is high and recovery might not be easy in case of node failure.

***D. Distributed Load Balancing Approach***- In distributed load balancing technique, multiple domains monitor the network to make accurate load balancing decision. Every node in the cluster keep the knowledge of all other node in the cluster. In centralized load balancing approach single node is responsible for making decision. If this node is fail, its effect all the system. This problem can be resolved by the distributed load balancing approach.

M. Randles et al. [15], proposed an distributed load balancing in approach called Honeybee foraging. This approach use the same concept which is use by the Honeybee to find the food.

***E. Hierarchical Load Balancing Approach***- In these types of approach tree data structure are used to making the decision regarding to the VM placement, wherein every node in the tree is balanced under the supervision of its parent node. Load balancing decision are taken by the parents nods base on the information gathered by the parent node.

## III. COMPARISON OF DIFFERENT LOAD BALANCING APPROACH

| *Type of Approach* | *Based On* | *Addressing Issues* | *Drawback* |
|---|---|---|---|

 

| | | | |
|---|---|---|---|
| Static | Fixed values are used, so previous knowledge is required. | Response time<br>Resource utilization<br>Scalability<br>Power consumption and Energy Utilization<br>Makespan<br>Throughput/Performance | Not Scalable<br>User can not changed demands at run time |
| Dynamic | Decisions are made at run time. Run time statics are required to take the decision | Load estimation.<br>Minimizing the number of migrations.<br>Throughput | Complex<br>Time Consuming |
| Centralized | All load balancing polices are operated in single node. | Threshold policies<br>Throughput<br>Failure Intensity<br>Communication between central server and processors in network. | Not fault tolerant<br>Overloaded central decision making node |
| Distributed | Load balancing polices are operated in multiple node | Migration time<br>Inter processor communication<br>Information exchange criteria<br>Throughput<br>Fault tolerance | Algorithm complexity<br>Communication overhead |
| Hierarchical | Tree data structures are used to making the decision regarding to the VM placement, wherein every node in the tree is balanced under the supervision of its parent node. | Threshold policies<br>Information exchange criteria<br>Selection of nodes at different levels of network<br>Failure intensity | Less fault tolerant<br>Complex |

## IV. CONCLUSIONS

Load Balancing is an essential task in Cloud Computing environment to achieve maximum utilization of resources. Resource in the cloud is distributed geographically and resource required by the VM can changed dynamically. Therefore load balancing in the cloud is a more challenging task as compare to the cluster and grid computing. This paper explain the different type of exiting load balancing approaches, each having some pros and cons. Static and centralized load balancing approaches are not suitable for the cloud, due to the large network infrastructure. So dynamic and hierarchical are the two approaches that can be used in the cloud.

## REFERENCES

[1] Peter Mell,Timothy Grance, "Cloud Computing" by National Institute of Standards and Technology - Computer Security Resource Center-www.csrc.nist.gov.

[2] RK Gupta et al., "A Complete Theoretical Review on Virtual Machine Migration in Cloud Environment", IJ-Closer, vol. 3, pp. 172-178, 2014.

[3] Wenke Ji, Jiangbo Ma "A Reference Model of Cloud Operating and Open Source Software Implementation Mapping" in 18th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises, 2009.

[4] Miyuki sato"creating next generation cloud computing based network services and the contribution of social cloud operation support system (OSS) to society" 18th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises, 2009

[5] Borja Sotomayor et al., "Enabling cost-effective resource leases with virtual machines," Research Gate article may 2014..

[6] L. Cherkasova et al. "When virtual is harder than real: Resource allocation challenges in virtual machine based it environments" in proc. 10th conference on hot topic in operating system, Vol. 10, pp.20-20, 2005.

[7] H. Jin et al., "Live migration of virtual machine based on full system trace and replay", proceeding of the 18th ACM 2009.

[8] Paul, M. et al. "Task-scheduling in cloud computing using credit based assignment problem",.International journal of Comput. Sci. Eng., pp. 26-30, 2011

 

[9] M. Katyal et al., "A Comparative Study of Load Balancing Algorithms in Cloud Computing Environment", International Journal of Distributed and Cloud Computing, vol. 1, pp. 1-14, 2013.

[10] Sotomayor, B., Montero, R. S., Llorente, I. M. & Foster, I. (2009). Virtual infrastructure management in private and hybrid clouds. IEEE Internet Computing, 13(5), 14-22.

[11] Radojevic, B. & Zagar, M, " Analysis of issues with load balancing algorithms in hosted (cloud) environments", In proceedings of 34thInternational Convention on MIPRO, IEEE, 2011.

[12] A. Beloglazov et al. "Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing", Elsevier journal of Future Generation Computer Systems, pp. 755–768, 2012

[13] T. Wood et al., "Black-Box and Gray-Box strategies for virtual machine migration", NSDI'07 Proceedings of the 4th USENIX conference on Networked systems design & implementation, pp. 7-17, 2007

[14] Wang, S. C., Yan, K. Q., Liao, W. P. & Wang, S., "Towards a load balancing in a three evel cloud computing network. Proceedings of 3rd International Conference on Computer Science and nformation Technology (ICCSIT), IEEE, July, (2010).

[15] Randles, M., Bendiab, A. T. & Lamb, D., "Cross layer dynamics in self-organising service oriented architectures", IWSOS, Lecture Notes in Computer Science, 5343, pp. 293-298, Springer, 2010.

 