International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 06 | June-2016 www.irjet.net p-ISSN: 2395-0072

A Comparative Study on Various Existing VM Selection and VM Placement Approach in the Cloud Environment

Amit Kumar Dehariya, Yogendra P.S. Maravi, Sanjeev Sharma
[1]*Amit Kumar Dehariya (amit.dehariya101@gmail.com)*
[2]*Yogendra P.S. Maravi (yogendra.rgpv@gmail.com)*
[3]*Dr. Sanjeev Sharma,(sanjeev@rgtu.net) Dept. of SOIT, RGPV college, MP, India*

Abstract - *Load balancing is one of the critical issues in cloud due to the change in user requirement at run time. Cloud provider allots resources to the user with the help of virtualization which allows dividing the physical resources in the form of virtual machine (VM). User services are running on these VM which is hosted inside the physical machine (PM). If the VM is not distributed properly then it will degrade the performance of the physical and virtual machine. Hence load balancing is the core management function of the cloud provider. Three steps are involved in the migration process i.e., source PM selection, VM selection and the last step is target PM selection. The study of previous work on the VM migration says that VM selection and VM placement are the two challenging task in the cloud environment and the performance of the load balancing approach is totally dependent on the VM selection and placement. Further performance of the load balancing approach can be controlled by selecting the suitable physical and virtual machine. Plenty of work on the load balancing in cloud computing environment are presented in the last few decade and mostly they are differ in the VM selection and VM placement policies. This paper presents various existing VM selection and placement approaches with their anomalies.*

***Key Words*: VM selection, VM placement, SLA violation, VM migration.**

1. INTRODUCTION

In the cloud computing, the computing resources are provided to the client through virtualization, via the Internet. The large scale computing infrastructure is established by cloud providers to make availability of online computing services in flexible manner so that the user find easiness to use the computing services [1]. According to National Institute of Standards and Technology (NIST) cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to shared pool of configurable computing resources. The computing resources include networks, servers, storage, applications, and services. In cloud computing, the shared pool of computing resources can be rapidly provisioned and released [2].
In the cloud computing there are three types of service delivery model [3] as software as a service (SAAS), platform as a service (PAAS), and infrastructure as a service (also known as hardware as a service). It can be deployed by four different model i.e., private, public, hybrid and community cloud. As shown in figure 1.

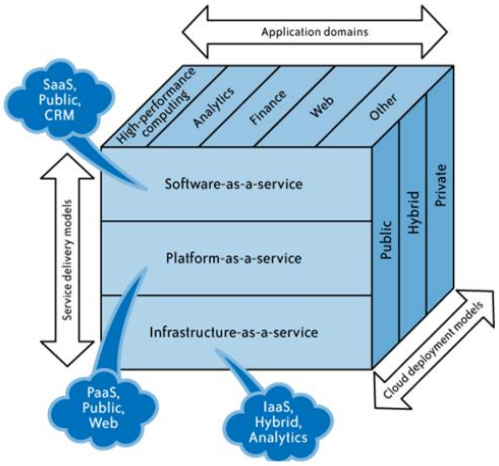

Figure 1: Cloud Computing Models in 3D

In Software as a Service (SaaS) delivery model only software is provided on demand to the client. There is no need to install software on the client side. In Platform as a Service (PaaS) delivery model a complete platform which is required to design new application is provided to the client. It is mainly used by the application developer. In infrastructure as a Service (IaaS) delivery model computing environment i.e., hardware, software, network etc. are provided to the user.
There are four types of cloud computing deployment model [4] in the cloud computing known as public, private, community and hybrid cloud.
In private cloud environment all computers are connected locally. Services running in private cloud environment cannot be accessed from outside the network. It is more secured and less scalable as compared to the other cloud computing. Public cloud is a model of cloud computing where all users are allowed to access the services using the Internet. The user needs only the Internet connection and web browser to access with pay per use scheme. All the services with infrastructure of cloud computing provider are available on the internet. User need to subscribe the application and make enable to use it. Community cloud includes number of organizations to share their services. The

aim is 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 on the basis of their advantages. The hybrid cloud environment is a good solution for business oriented strategy because many modern businesses have a wide range of concerns to support users demand.

In cloud computing, several PMs are connected to each other in the form of cluster. Virtualization [4, 5, 6] is the enabling technology in the cloud computing, which divide the physical resources into the multiple part via VM. When any user needs resources, scheduler assigns the resources of these PM to the user through the VM. Each user has its own VM and the resource requirement of the VM can be changed dynamically at run time. Due to this reason load balancing in the cloud computing is the crucial task. Since physical resources are shared by the multiple users, so there might be a situation where some nodes (PM) are over utilized whereas some nodes are underutilized. In order to balance the PM, VM migration approach [8, 9, 10] is used which transfer the VM from one PM to another. In cloud computing, VM migration consists of three steps. In the first step of the migration process we have to find the source PM which is overloaded or under loaded. For this purpose lower and upper limit of the physical resource is set and based on these values source PM is chosen. In the second step selection of the VM for the migration has to be done. The last step involves selection of the target PM to place the selected VM. Figure 2 shows the process involved in the VM migration where 5th VM is migrated from PM-2 to PM-3.

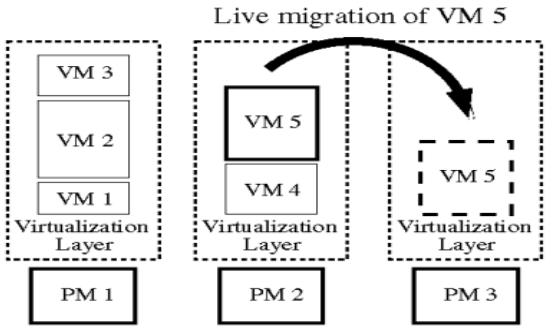

Figure 2: VM Migration

In this paper, the overview of cloud computing with their basic components and deployment models are discussed. The goal of this paper is to provide a complete study of same existing VM selection and placement techniques with their comparative analysis. Section 2 covers the background study of VM selection and VM placement with their framework. Section 3 concludes the paper with the focus on the future possibilities.

2. RELATED WORKS

Mohammad h et al. [9], proposed energy aware VM migration for the cloud computing. This approach uses four thresholds namely power off threshold, wake up threshold, load in threshold and target server threshold. The VM migration approach is invoked either when the VM complete its execution or when existing VM dynamically resizes itself. When load on the pm is below the power off threshold, then all VM running on that pm is migrated to the other pm and switch off the pm to save the power. Similarly when the load on the pm is more than the wake up threshold for a predefined time then they migrates the largest utilized VM. To select the target server threshold is used. First they find the list of pm whose threshold is less than the target server threshold then use first fit approach to place the VM. Main problem with this approach is they select the largest VM for the migration in the case of overloaded which increase the total migration time. In addition they use first fit approach to place the VM which may lead to the situation where some PMs are underloaded and others are overloaded.

G.shobana et al. [10], proposed load balancing approach for the cloud based on the preemptive task scheduling. This approach uses c.p.u. and bandwidth as a decision metrics for calculating load on the virtual machine. They say that load on the pm is the summation of all VM load running on the pm. In this approach VM are grouped into three types named overloaded VMs (OVM), underloaded VMs (UVM) and balanced VMs (BVM) based on their load. Following equation is used to calculate the load on the pm.

$$C = \sum_{i=1}^{m} c_i$$

Where,
C is the capacity of a single VM is given by
$c_j = VM_{mipsj} + VM_{bwj}$
J is the VM
VM_{mipsj} is the millions instruction per second (mips) of the jth VM
VM_{mipsj} is the bandwidth of the jth pm
M is the number of running VM

When the pm is overloaded all VM in OVM (overloaded virtual machine) group are required to be arranged in descending order and UVM (underloaded virtual machine) set by ascending order. One task is to remove from the OVM, VM is and place to the any VM in the UVM group. This approach migrate the task from one VM to another VM to balance the load. But this approach is not effective for the cloud because task is move from one VM to another VM. In this case load on the pm is remains same hence pm is still overloaded.

A. Rabiatul et al. [11], introduced a load balancing approach based on the VM migration. This approach set the value of lower and upper limit for the resource utilization of the pm which is 10 and 90 respectively. When the load is above the upper limit, larger VM is selected from the overloaded pm and placed to the pm where the resource utilization of the pm is less than 50. This approach seems good but may increase the number of migration due to setting higher value of the upper threshold.

G. Xu et al. [12], presented load balancing approach which places the VM according to the partition. This approach first divide the data center into the partition according to the distance and then assign the VM according to the near partition. In each partition pm is divided into three categories named idle, normal and overload. To find in which categories host belongs load_ degree is used which is given by following equation

$$\text{Load\_degree(n)} = \sum_{i=1}^{m} \propto_i * F_i$$

$$load\_degree_{avg} = \sum_{i=1}^{n} \frac{load\_degree(N_i)}{n}$$

Where

$\propto_i$ is the waiting coefficient

N represents the current pm.

M is the different type of resources

N is the number of VM

Based on the value of $load\_degree_{avg}$ status of the pm will be determine. This approach may increase the number of active server due to partition of the datacenters.

Y. Fang et al. [13], proposed task scheduling model for the VM in cloud environment. They proposed the two layer architecture for the VM placement. First layer gives the description of the VM and second layer assign the resource to the VM. The VM is assigned to the smallest pm. When the hot spot or load unbalancing situation occurs they select the smallest VM for the migration and place it to the lightest loaded pm. This approach selects the small VM for the migration which may increase the number of migration. In addition they are also not focused on the server consolidation.

Previous study says that we can enhance the performance of any load balancing technique by appropriately performing the VM selection and VM placement tasks. Larger VM selection may increase the total migration time and down time whereas smaller VM selection may increase the number of VM migration which lead to more SLA (service level agreement) violation. Hence the proper VM selection is the prime requirement of any load balancing approach.

All the above discussed approaches are summarized in table 1. This table shows type of VM utilization basis VM is selected for the migration and where it is scheduled. Y letter in the table represent that the approach considered the corresponding metrics during the load balancing whereas n letter in the table represent that the approach does not considered the corresponding metrics.

Table 1: Comparisons of Various Existing Load Balancing Approach

Paper	Type	VM Selection	VM Placement	Objective		
				Reduce Energy Consumption	Support for Migration	If simulation time is considered
Mohammad et al. [9]	VM	Largest Utilize	First Fit	N	Y	Y
G.Shobana et al. [10]	Task	Larger task from Overloaded PM	Place any Underloaded VM	Y	Y	Y
A. Rabiatul et al. [11]	VM	Least Utilize	PM load < 50	N	Y	N
G. Xu et al., [12]	VM	No Migration	Neighbor PM	N	N	Y
Y. Fang et al., [13]	VM	Small	Largest Utilize	N	Y	Y

3. CONCLUSIONS

Virtualization is the key technology in the cloud. Without virtualization cloud computing cannot be imagine. One of the important feature of the cloud computing is the migration because it allows to move the VM from one PM to another. But successive migration, total migration time and down time are three parameters that define the quality of the load balancing approach. This paper is an effort to highlight some of the existing approaches for the VM selection and VM placement. These are the two critical issues that are involved in the performance enhancement of the load balancing approach. In the last few decades a lot of work has been done on the load balancing in cloud environment and the only thing they differ in is the VM selection and VM placement polices. The sole purpose of this paper is to provide the brief overviews of some existing approaches and to show a comparative study among them. It has been observed that larger VM selection may increase the total migration time and down time whereas smaller VM selection may increase the number of VM migration which lead in more SLA violation. Hence the proper VM selection is the prime requirement of any load balancing approach.

REFERENCES

[1] R. Buyya, CS Yeo, S. Venugopal, J. Broberg, I. Brandic, "Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility, Future Generation Computer Systems, 2011

[2] R.K. Gupta and R.K. Pateriya, "A Complete Theoretical Review on Virtual Machine Migration in Cloud Environment", International Journal of Cloud Computing and Services Science (IJ-CLOSER), Vol.3, No.3, pp. 172-178, June 2014.

[3] R.K. Gupta and R.K. Pateriya," Survey on Virtual Machine Placement Techniques in Cloud Computing Environment", International Journal on Cloud Computing: Services and Architecture (IJCCSA) ,Vol. 4, No. 4, pp. 1 -7, August 2014.

[4R. P. Goldberg et al.," Architecture of virtual machines", Proceedings of the workshop on virtual computer systems (ACM), pp. 74-112, 1972..

[5] R. Santhosh and T. Ravichandran, "A Survey on Cloud-Based Scheduling Algorithms", International Journal of Communications and Engineering (IJCE), Volume 01– No.1, Issue: 01 May2013.

[6] S. K. Mandal and P. M. Khilar, "Efficient Virtual Machine Placement for On-Demand Access to Infrastructure Resources in Cloud Computing", International Journal of Computer Applications (IJCA), Vol. 68, No.12, April 2013.

[7] "Xen, virtual machine manager in Cloud computing," [online] available: http://www.xen.org, April 2013.

[8] A Jain et al., "A Threshold Band Based Model For Automatic Load Balancing in Cloud Environment", in proc. of IEEE International Conference on Cloud Computing in Emerging Markets, pp 1-7, 2013.

[9] Mohammad H et al.,"An Energy-aware Virtual Machine Migration Algorithm" IEEE International Conference on Advances in Computing and Communications, pp. 242-246, 2012.

[10] G.Shobana et al., "Nature Inspired Preemptive Task Scheduling for Load Balancing in Cloud Datacenter", International Conference on Information Communication and Embedded Systems (ICICES), pp. 1-6, 2014.

[11] R. Addawiyah et al., "Virtual Machine Migration Implementation in Load Balancing for Cloud Computing", six IEEE international conference, 2014.

[12] G. Xu et al. "A Load Balancing Model Based on Cloud Partitioning for the Public Cloud" in the proceeding of IEEE conference on Tsinghua Science and Technology, pp. 34-39, 2013.

[13] Y. Fang, F. Wang, and J. Ge, "A Task Scheduling Algorithm Based on Load Balancing in Cloud Computing", Web Information Systems and Mining, Lecture Notes in Computer Science, Vol.63, pages 271-277, 2010