International Journal of Current Trends in Engineering & Technology
ISSN: 2395-3152
Volume: 01, Issue: 07 (November-December 2015)

A Load Balancing Approach using VM Migration for Increasing the Resource Utilization in Cloud

Md. Ahamad Ansari[1], Prof. K.K. Tiwari[2]
ahamad.ansari88@gmail.com[1],krishna19it@gmail.com[2]
Department of Computer Science & Engineering
Surbhi College of Engineering & Technology, Bhopal, India

Abstract:- Cloud computing evolve as a new technology in the field of IT and growing so much faster due to attractive feature like easy to use, dynamic allocation and reallocation of the resources, less costly etc. It provides on demand resources to the client on the rent basis. Cloud support for the utility model, so user has to pay only for the use resources. Since it provide resource to the users and demand for the resources increasing very fast in the past few decades. So load balancing is the main requirement of the cloud system. But load balancing in cloud is more difficult as compare to other technology because it is so large and user requirement can be change dynamically. It helps in optimizing the resource utilization, hence enhancing the system performance. The prime goal of any load balancing approach is to maximize the resource utilization and reducing the number of active server which will further reduce energy consumption and carbon emission. During the past decades several load balancing approach have been proposed. Main objective of these approaches is to increase the system performance by reducing the number of migration. But these approaches are not focusing on the resource wastage. This paper proposed a load balance approach that reduce the number of migration and increase the resource utilization. It is implemented in CloudSim simulator. Experiment result shows that proposed approach gives better result as compare to the existing load balancing approach.

Keywords: *Cloud computing, Physical machine, Virtual machine, Cloud service model, Utility computing, Load balancing, Energy efficiency, Green computing.*

1. INRODUCTION

Cloud computing is an on demand service in which shared resources, information, software and other devices are provided according to the client requirement at specific time [1]. It can be deploy in four different way (private, public, community and hybrid) and provide three type of services (software as service, platform as service and infrastructure as service) [2, 3, 4] as shows in figure 1. Private cloud allows users to access cloud services within the network. User can't access cloud services from outside the network. It is suitable for the small organization. User in public cloud can access cloud services from anywhere in the world. It is larger than

International Journal of Current Trends in Engineering & Technology
ISSN: 2395-3152
Volume: 01, Issue: 07 (November-December 2015)

the private cloud and provides larger number of services. Community cloud is a cloud which is share by the multiple organizations. Hybrid cloud is a combination of one or more cloud like combination of private and public cloud, private and community cloud etc. Cloud provide services to the users and provide three type of service named software as service (SaaS), platform as service (PaaS) and infrastructure as service (IaaS). Software as a service mainly delivers the online software application to the client on-demand. Users use these software and application without any installation. Platform as a Service (PaaS) allows developer or gives the capability to create application as a service according to their desire. It allows users to develop their software using programming languages and tools supported by the provider. Infrastructure as a Service (IaaS) provides the capability to have control over complete cloud infrastructure with CPU processing, storage, networks, and other computing resources.

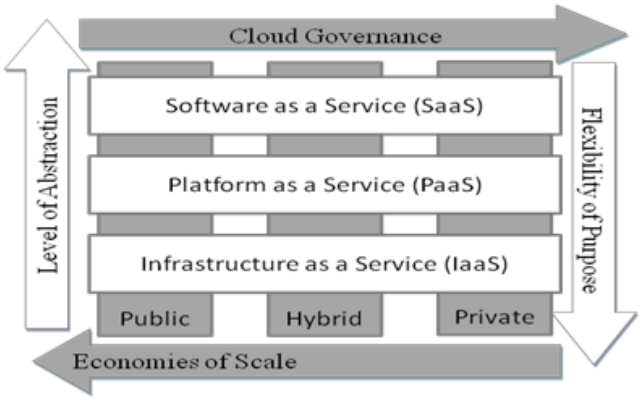

Figure 1: Cloud Computing Model

Virtualization [5, 6] is the core technology in cloud computing. It allows the dividing of the physical resources. When user request for the computing resources like CPU, memory, bandwidth etc., provider create the virtual machine (VM) and assign to the physical machine (PM). Resources are gives to the user in the form of VM. Hypervisor is the software which creates the VM and divides the single physical resources into the multiple types. Virtualization allows the provider to transfer VM from one PM to another PM which is known as VM migration. VM migration [7, 8] is the important features of the virtualization which allow balancing the PM. Since in cloud computing resources are share by the multiple users, so there is a requirement of a load balancing approach that distributes the load on the physical machine equally. Load balancing approach can be static or dynamic. But static approach is more suitable for the cloud as compare to the dynamic approach. The advantages of appropriate migration of virtual machines include effective load balancing, server consolidation, online maintenance and proactive fault tolerance. In this given scope, the main objective of our research is to improve the load balancing process. In the long run, improvement of the load balancing process will improve the overall process of resource management on Cloud and thus will help in promoting the concept of green cloud computing.

2. LITERATURE SURVEY

Load balancing in cloud is a complicated task due to change in user requirement at run time. Numbers of load balancing approaches have been proposed in the last decades. These all approaches use lower and

upper threshold to define the overloaded and under loaded situation. To balance the PM these approach migrate the VM from one PM to another PM. R. Addawiyah et al. [9], proposed a load balancing approach for the cloud which is based on the VM migration. Main objective of this approach is to minimize the power consumption for this purpose VM placed only on the basis of CPU utilization. This approach set the value of lower threshold is 10 and the value of upper threshold is 90. That means when the value of CPU utilization is more than 90 then the PM is overloaded. Similarly when the value of CPU utilization is less than 10 then the PM is underloaded.

A. Beloglazov et al. [10], proposed threshold based an energy efficient load balancing approach. According to this paper average power consumed by an idle server is 70% of power consumed by fully utilized server. Hence, appropriate load balancing approach can controlled the power consumed by the data center. This approach used static lower and upper threshold with the difference of 40 between lower and upper threshold. After the experiment they set 30 as a lower threshold and 70 as an upper threshold. This approach reduced the number of migration but main problem with this approach is that they are not working on resource balancing.

M. Mishra et al. [11], proposed a VM placement approach for the cloud environment. In this approach all resource related information is explained in the vector form. This approach considers three resources named CPU, memory and bandwidth. This vector is identified as Total Capacity Vector (TCV). Resource Utilization Vector (RUV) represents the current utilization of resources of a PM. RUV is the vector addition of normalized utilization vectors of each resource type. Note that the normalization of utilization of individual resource type happens with respect to the total capacity of that resource type. The vector difference between TCV and RUV represents the Remaining Capacity Vector (RCV), which essentially captures how much capacity is left in the PM. The resource requirement of a VM is represented by Resource Requirement Vector (RRV) which is the vector addition of normalized resource requirement vectors of each resource type. Note that the resource requirement is normalized with respect to the total capacity of the target PM. Main concept of this approach is assign VM to the PM where resource requirement of the VM and resource usage of the PM is same but in the opposite direction. This is the first approach which minimized the resource wastage. But main problem with approach is that only theory of the model is given.

A Jain et al. [12] proposed a threshold based load balancing approach for the cloud. Only single threshold i.e. upper threshold is use for the load balancing. They use the dynamic threshold which changes with the load. Load balancing done on the basis of this dynamic threshold. VM allocated to the host till the upper threshold. A host is called to be overloaded when the load on the host is greater than the upper threshold. Host CPU utilization are used

International Journal of Current Trends in Engineering & Technology
ISSN: 2395-3152
Volume: 01, Issue: 07 (November-December 2015)

to calculate the upper threshold. For calculating the threshold following formula are use

$$\theta = \frac{\alpha_1 + \alpha_2 + \alpha_3 - - - - \alpha_n}{n}$$

Where α_i is the CPU utilization of the i[th] host and n is the total number of host in the cluster. This approach increase the resource utilization, but not support to the sever consolidation. So it will increase the number of active server.

3. PROPOSED APPROACH

After reviewing the theory of load balancing it is fond that lot of work has done in the field of load balancing but main objective of these approaches are to minimize the power consumption and minimize the number of migrations. Some of them focus on the resource wastage. As we know a PM consist of multiple resources named CPU, RAM, hard disk, network, bandwidth etc., similarly VM is also consist of same type of resource. Hence, there is a possibility where one type of resources is over utilized while other type resource is underutilized. So if all VM which required more one type of resource (say CPU) are assign to the same PM then CPU of this PM is overutilize whereas other resources are underutilize. This situation decreases the utilization of the PM and increases the number of running PM. Most of the work done previously [13, 14], consider three resources as a metrics named CPU, memory and bandwidth for the load balancing. Based on these theories we also consider these three resources. Load on the PM is depends on the VM and normally enhance with the number of VM. When the PM is overloaded its performance is degraded. Similarly when the PM is underloaed it resource is wastage. Hence load on the PM is distributed properly. For this purpose proposed approach use lower and upper thresholds. When the value of the CPU utilization is lower than its lower threshold then PM is under loaded and resources are wastage and if value of the CPU utilization is larger than its upper threshold then PM is over loaded and the performance of the PM is degraded. The value of lower and upper threshold is set to 20 and 80 respectively [14].

3.1 Load on the Physical and Virtual Machine

To calculate the load on the physical and virtual machine following equation is used:

$$VM_{cpu} = \frac{Requested\ CPU}{Toatl\ CPU\ available}$$

$$VM_{mem} = \frac{Requested\ memory}{Toatl\ available\ memory}$$

$$VM_{bw} = \frac{Requested\ BW}{Toatl\ available\ BW}$$

Each PM can run several virtual machines. Hence total load of the PM is defined as a load of all virtual machine running on the physical machine. If m VM are running on the n[th] PM then average load on the n[th] PM can be given by the following equation

International Journal of Current Trends in Engineering & Technology
ISSN: 2395-3152
Volume: 01, Issue: 07 (November-December 2015)

$$PM^{n}_{cpu} = \frac{\sum VM_{cpu}}{Total\ CPU\ of\ the\ PM}$$

$$PM^{n}_{mem} = \frac{\sum VM_{mem}}{Total\ memory\ of\ the\ PM}$$

$$PM^{n}_{bw} = \frac{\sum VM_{bw}}{Total\ BW of\ the\ PM}$$

Above equation measured the total CPU, memory and bandwidth load on the PM.

Where

{ PM_{cpu} is the CPU load of the n^{th} PM.
PM_{mem} is the memory load of the n^{th} PM.
PM_{bw} is the bandwidth load of the n^{th} PM.
VM_{cpu} is the CPU load of the VM.
VM_{mem} is the memory load of the VM.
VM_{bw} is the bandwidth load of the VM. }

As discussed earlier, to balance the PM our approach use VM migration approach which migrate some VM from one PM to another PM. VM migration is the unique solution for balancing and optimizing the performance of the PM. Three steps are involved in the VM migration.

(A) Source PM Selection

For selecting the source PM we are using lower and upper threshold. The lower thresholds shows that PM is underutilize whereas upper threshold shows that PM is overutilized. In the first case i.e., PM is underloaded, all VM running on that PM is migrated to the other PM and underutilize PM is put to the power saving mode. In the earlier case i.e., PM is overloaded, we continuously migrate the VM from the overloaded PM till the PM is balance.

(B) Virtual Machine Selection

Each host can have number of VM. So after the source PM selection, next step is to select the VM which has to be migrated. In our approach when the PM is underloaded all running VM are migrated to the other active host. But in case of load balancing, size of the VM plays an important role because small size VM selection may increase the number of migrations. In our proposed VM selection approach we select the largest VM from the overloaded PM to balance the PM. Since the value of upper threshold in our approach is 80, so when the load on the PM is greater than its upper threshold then select the largest utilize VM for the migration.

Algorithm for the VM selection

1. Input: hostList Output: migrationList
2. **for** each h in hostList **do**
3. vmList ← h.getVmList()
4. vmList.sortDecreasingUtilization()
5. hostUtil ←host.Util()
6. **while** hostUtil > 80 **do**
7. bestVm ← { First VM from the vmList}
8. bestVmUtill ← vm.getUtil()
9. hostUtil ← (hostUtil ─ bestVmUtill)
10. migrationList.add(bestVm)
11. vmList.remove(bestVm)
12. **end while**
13. **end for**

Algorithm for the Consolidation

International Journal of Current Trends in Engineering & Technology
ISSN: 2395-3152
Volume: 01, Issue: 07 (November-December 2015)

1. Input: hostList Output: migrationList
2. **while** (hostUtil < 30) **do**
3. migrationList.add(h.getvmList())
4. vmList.remove(h.getvmList())
5. return migrationList()
6. **end while**

(C) Target PM Selection for the VM Placement

In cloud, each data center have multiple PM, so after selecting the VM next step is place this VM to the suitable PM. VM placement is one of the most challenging task in the load balancing. If we place the VM to the wrong PM, then it will increase the number of active server which will increase the energy consumption because energy consumption is depends on the number of active servers. Moreover it also increases the number of migrations. Main objective of our approach is to minimize the resource wastage which can be control effectively by the proper placement of the VM. Computer system consist of multiple resources like CPU, memory and bandwidth etc., so if all VM which need more one type of resources (say CPU) is place to the PM then CPU of this PM is utilize completely whereas other resources (memory and bandwidth) are underutilize. This situation will increase the resource because remaining resources can't be allocated to any VM. Our proposed VM placement approach avoids this situation. For placing the VM our approach divides the physical and virtual machine into six types. Since our approach use three resource i.e., CPU, memory and bandwidth, so there are six possible combinations (3!= 6). These combinations are CMB, CBM, MBC, MCB, BCM and BMC. In case of VM CMB stands for the VM which needs more CPU than memory and RAM. Hence, VM required resource in the following order (CPU > Memory > Bandwidth). Similarly CBM type VM required resource in the following order (CPU > Bandwidth > Memory). Whereas in the case of PM, CMB stands for the PM which use more CPU than memory and bandwidth. Hence, PM use resources in the following order (CPU > Memory > Bandwidth). Similarly CBM type PM use resources in the following order (CPU > Bandwidth > Memory). To reduce the resource wastage our approach places the VM to the PM, where resource requirement of the VM and resource utilization of the PM is opposite to each other. Hence, VM which required resources in the following order (CPU > Memory > Bandwidth) is places to the PM which use the resources in the following order (Bandwidth > Memory > CPU). That means CMB type VM is place to the BMC type PM and vice versa.

Algorithm for the VM Placement

1. Input: hostList, vmList Output: allocation of VMs
2. **for each** vm in vmList **do**
3. **if** VM= new **then**
4. **for each** host in hostList **do**
5. if host has enough resource for vm then
6. power ← estimatePower(host, vm)
7. Assign vm to the host where power difference between before and after is minimum.
8. **end for**

International Journal of Current Trends in Engineering & Technology
ISSN: 2395-3152
Volume: 01, Issue: 07 (November-December 2015)

9. **else**
10. **foreach** host in hostList **do**
11. Arrange host into the corresponding type (CMB, CBM, MBC, MCB, BCM and BMC)
12. **end for**
13. vmType ← Find type of the VM (CMB, CBM, MBC, MCB, BCM and BMC)
14. find PM into the opposite VM type
15. if multiple host has enough resource for vm then
16. power ← estimatePower(host, vm)
17. Assign vm to the host where power difference between before and after is minimum
18. else
19. Active new host and assign VM.
20. **end for**

4. EXPERIMENTAL SETUP AND RESULT

To measure the performance of the proposed approach it is compare with the other existing load balancing approach [46]. Both approaches are implemented in CloudSim simulator. Each PM in the data center has 2000 MIPS, 10000MB of RAM and bandwidth of 100000 bit/s. VM created on these PM are having 250, 500, 750, 1000 MIPS, 2048, 512, 128 MB of RAM and 2500, 7500, 12500 bandwidth. First experiment is performed to check the number of migrations because it affects the system performance. Figure 2 shows the number of migrations for the different number of VM in base and proposed approach. Prime objective of our approach is to increase the resource utilization. Next experiment is performed to measure the resource utilization for these approaches.

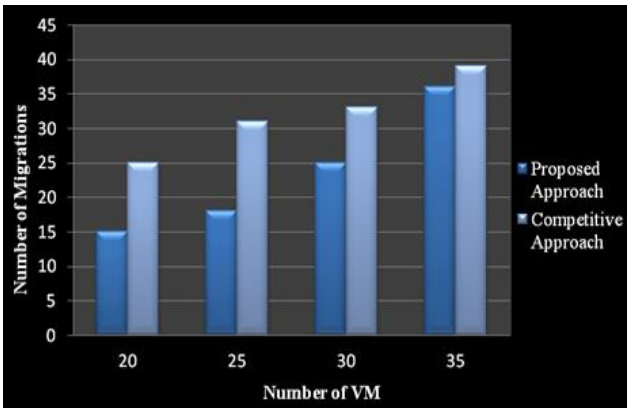

Figure 2: Number of migrations

Figure 3 and 4 shows the resource utilization of the proposed and base approach for different number of VM. Proposed approach utilized resources effectively as compare to the base approach.

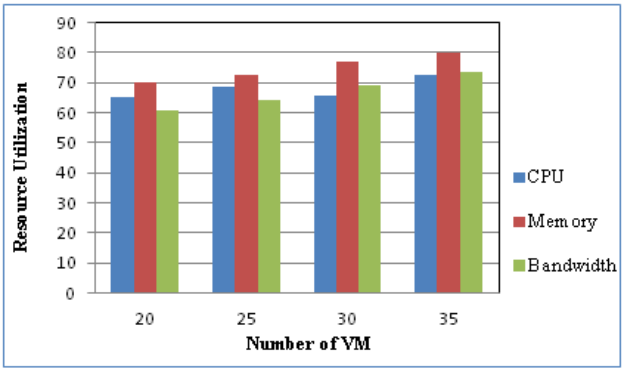

Figure 3: Resource Utilization for Proposed Approach

This is because base approach did not focus on the resource balancing whereas our approach place VM to the PM where resource requirement of the VM and resource utilization of the PM are opposite to each other. So CBM type VM is place to the MBC PM. Last experiment is to perform to measure the power consume by the data center on the basic of above results we can say that our method minimize the number of migrations and increase the resource utilization by utilizing the resource effectively. Since

International Journal of Current Trends in Engineering & Technology
ISSN: 2395-3152
Volume: 01, Issue: 07 (November-December 2015)

energy consumption is depends on the number of active server and host utilization. So this approach also reduces the energy consumption by shutting down the ideal server.

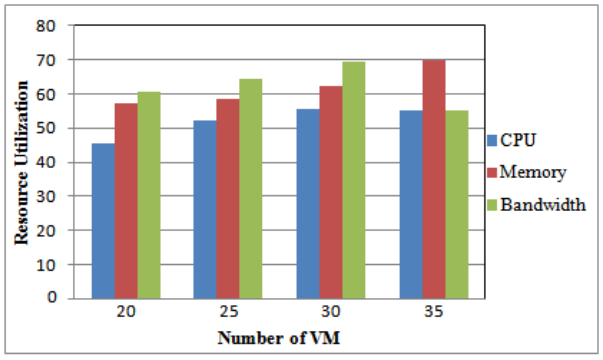

Figure 4: Resource Utilization for Base Approach

Due to reducing the number of active servers it consume the minimum amount of energy as compare to the other approach. Hence increase the resource utilization.

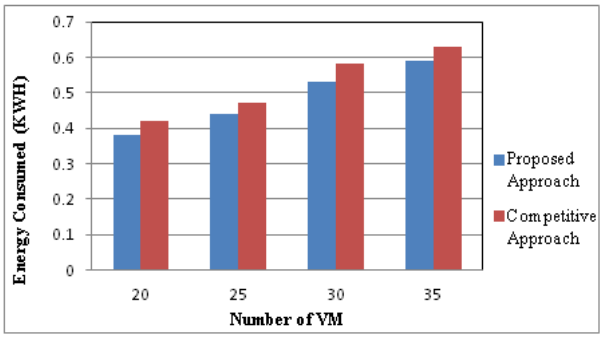

Figure 5: Energy Consumed by 10 PM for Different VM

5. CONCLUSION

Load balancing in the cloud is the complicated task due to the change in user requirement at rum time. Numbers of load balancing approaches have been proposed in the last decades. These all approaches use lower and upper threshold to define the overloaded and under loaded situation. To balance the PM these approach migrate the VM from one PM to another PM. Main problem with the existing load balancing approach is the resource wastage. Resource wastage reduce the resource utilization which increase the number of running server. These running server increases the power consumption which result increase the running cost for the provider. So from cloud provider point of view this resource wastage must be minimized. This paper proposed a load balancing approach which minimize the resource wastage. Proposed approach is implemented in CloudSim simulator. To measure the effectiveness of the proposed approach it compare with other approach and experiment result explain that proposed gives good better result.

6. REFERENCE

[1] M. D. Dikaiakos, G. Pallis, D. Katsa, P. Mehra, and A. Vakali, "Cloud Computing: Distributed Internet Computing for IT and Scientific Research", IEEE Journal of Internet Computing, Vol. 13, No. 5, September/October 2009, pages 10-13.

[2] R. K. Gupta et al., "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, June 2014, pp. 172-178.

[3] G. Pallis, "Cloud Computing: The New Frontier of Internet Computing", IEEE Journal of Internet Computing, Vol. 14, No. 5, September/October 2010, pages 70-73.

[4] B. P. Rima, E. Choi, and I. Lumb, "A Taxonomy and Survey of Cloud Computing Systems", Proceedings of 5th IEEE International Joint Conference on INC, IMS and IDC, Seoul, Korea, August 2009, pages 44-51.

[5] D.Perez Botero, "A brief tutorial on live migration of virtual machine from a security perspective", 2011.

[6] C. Clark, K. Fraser, s. Hand and J.C. Warfield, "Live migration of virtual machine", proceeding of the 2nd conference on symposium on network system design and implementation, 2007.

[7] T. Wood et al., "Black-Box and Gray-Box strategies for virtual machine migration", proc. 4th conference on symposium on network system design and implementation, 2007.

[8] Gunjan Khanna et al., "Application performance management in virtualized server environments", Network Operations and Management Symposium NOMS 10th IEEE/IFIP, pp. 373-381, 2006.

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

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

[11] Mayank Mishra and Anirudha Sahoo, "On Theory of VM Placement: Anomalies in Existing Methodologies and Their Mitigation Using a Novel Vector Based Approach", IEEE 4th international conference on cloud computing in 2011.

[12] 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.

[13] M. Mishra et al., "On Theory of VM Placement: Anomalies in Existing Methodologies and Their Mitigation Using a Novel Vector Based Approach", IEEE 4th international conference on cloud computing in 2011.

[14] L. Xu at el.,"Smart DRS: A Strategy of Dynamic Resource Scheduling in Cloud Data Center", proceeding of the IEEE International Conference on cluster Computing Workshop, pp. 120 -127, 2012.