Survey on Virtual Machine Placement Techniques in Cloud Computing Environment

Rajeev Kumar Gupta and R. K. Pateriya

Department of Computer Science & Engineering, MANIT, Bhopal, India

ABSTRACT

In traditional data center numbers of services are run onto the dedicated physical servers. Most of the time, these data centers are not used their full capacity in term of resources. Virtualization allows the movement of VM from one host to the another host ,which is called virtual machine migration, so these data centers can consolidate their services onto lesser number of physical servers than originally required. Virtual machine placement is the part of the VM migration. To map the virtual machines to the physical machines is called the VM placement. In other word, VM placement is the process to select the appropriate host for the given VM. For the efficient utilization of the physical resources, VM should be placed on to the suitable host. So many virtual machine placement algorithms have been proposed by different researchers that run under cloud computing environment. Most of the VM placement algorithms try to achieve some goal. This goal can either saving energy by shutting down some severs or it can be maximizing the resources utilization. Four steps are involved in the VM machine migration process. First step is to select the PM which is overload or undreloaded, next step is to select one or more VM, and then select the PM where selected VM can be placed and last step is to transfer the VM. Selecting the suitable host is one of the challenging task in the migration process, because wrong selection of host can increased the number of migration, resource wastage and energy consumption. This paper only focuses to the third step that is selecting a suitable PM that can host the VM. It shows an analysis of different existing Virtual Machine's placement algorithms with their anomalies.

KEYWORDS

Virtual machine, data centers, virtualization, migration, power saving

1. INTRODUCTION

During the last several decades, cloud computing is the new emerging technology in the field of computer science. It became so famous because of their cheap, on-demand and pay as use services [1]. Cloud computing provide three type of services that is Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) [2] and it can be deploy in three different way that is private, public and protected [3][4]. Private cloud can only access within an organization, public cloud can be access anywhere in the word and the hybrid cloud is the combination of private and public cloud. Private cloud offer more security than public cloud.

DOI : 10.5121/ijccsa.2014.4401

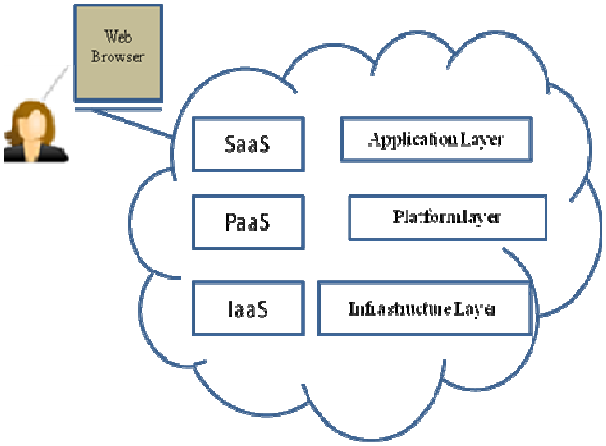

Fig. 1 cloud computing model

Virtualization [5, 6] is the key technology behind the cloud computing. It increased the resource utilization by sharing the physical resources among multiple users. Virtual machine monitor (Hypervisor) is responsible for all management related to the VM i.e. VM creation, destruction and scheduling. It is a small software program which reside between the hardware and the host OS must be installed onto the each host of the data center.

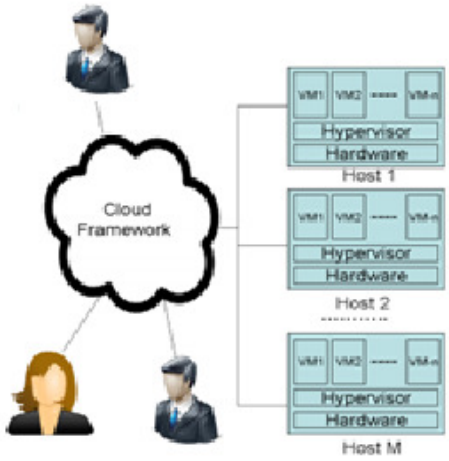

Fig.2 Cloud framework

Hypervisor create the VM for each user according to the user's requirement. These VMs are independent. One or more VM can assign to the user and each VM can run multiple applications. Furthermore, numbers of virtual machines are hosted on a single host. In the cloud environment resource required by the VM are changed dynamically, sometimes resource required by the VM are not fulfil by the PM in which VM are currently running. This problem can be solved either by adding some resources or by migrating VM. VM migration [7, 8,] is the process of moving VM from one physical machine to another physical machine. VM migration solves the problem such as fault tolerance, load balancing and resource consolidation etc. VM migration process consists of four steps. First step is to select the PM from which the VM is migrated, second step is to select the VM for the migration, next step is to select the PM, where the selected VM will be

placed and last step is to transfer the VM. Out of these four step third step i.e. VM placement is the more challenging task, because it directly affect the system performance. VM placement can be defined as a mapping between the VM and the PM. In the cloud environment there are number of host and each host run number of VMs. If the number of physical and virtual machines are less then static VM placement can be possible, but if the number of virtual machine and physical machine are more, dynamic VM placement methods is required. If *n* is the total number of virtual machine and *m* is the total number of host then the number of possible mapping can be calculate by the following equation [9]

Number of possible mapping = m^n

This indicates as the number of virtual and physical machine are increased, virtual machine placement becomes a challenging task.

2. CLASSIFICATION OF VM PLACEMENT ALGORITHMS

Goal of the VM placement can either saving energy by shutting down some severs or it can be maximizing the resources utilization. Based on these goal placement algorithm are classified into two type.

i. Power Based approach
ii. Application QoS based approach

Main aim of the power based approach is to save the energy. In these approach VM map to the physical machines in such a `way, that each servers can utilized their maximum efficiency and the other servers can be shut down depending on load conditions. While in the Application QoS based approach a VM map to the PM with the aim of maximizing the QoS delivered by the service provider.

3. LITERATURE SURVEY

VM placement problem is a non deterministic problem. Number of virtual machine placement algorithms have been proposed [10, 11, 12] that run under cloud computing environment. This section explains some of the exiting virtual machine placement approaches and their anomalies.

3.1. Heuristic based approach

3.1.1. First Fit

It is a greedy approach. In this approach scheduler visits the PM sequentially one by one and placed VM to first PM that has enough resources. Each time when the new VM arrived, scheduler starts searching the PM sequentially in the data center till it finds the first PM that has enough resources. If none of the physical machine satisfied the resource requirement of the VM, then new PM is activated and assigns VM to the newly activated PM. Main problem with this approach is that load on the system can imbalanced.

3.1.2. Single Dimensional Best Fit

This methods use the single dimension (CPU, memory, bandwidth etc.) for placing a VM. When VM arrived, scheduler visit the Physical Machines in the decreasing order of their capacity used in a single dimension and place the VM to the first PM that has the enough resources. That means

VM place to the PM which used the maximum capacity along with the given dimension. Problem with approach is that it can increase the resource imbalancing because resource in the cloud is multi-dimension (CPU, memory, bandwidth etc.). So there may be a situation where a host utilize their full CPU capacity while other resources such as memory and bandwidth are underutilized.

N. Rodrigo et al. [17], proposed a heuristic for the mapping between VM and PM. Main aim of this approach is to balance the CPU utilization on each PM. Only CPU utilization is consider as a load metric, so after the mapping only amount of available CPU is check instead of whole PM. Objective function of this method is to minimize the standard deviation of residual CPU in each PM.

If there is *n* host in the data centers and m is the number of VM in each host then objective function can be define as

$$Min \left(\sqrt{\frac{\sum_{i=1}^{n} \left(rcpu_i^{PM} - rcpu_{mean}^{PM} \right)^2}{n}} \right)$$

$$rcpu_i^{PM} = cpu_i^{PM} - \sum_{j=1}^{m} cpu_j^{VM}$$

$$rcpu_{mean}^{PM} = \frac{\sum_{i=1}^{n} rcpu_i^{PM}}{n}$$

Where $\boldsymbol{rcpu_i^{PM}}$ is the reaming CPU capacity of the i[th] PM, $\boldsymbol{cpu_i^{PM}}$ is the total CPU capacity of the i[th] PM and $\sum_{j=1}^{m} cpu_j^{VM}$ is the CPU used by the j[th] VM.

Problem with this approach is that they only consider the CPU capacity for mapping between the PM and VM. So other resource can be imbalance.

3.1.3. Volume Based Best Fit

This heuristic used the volume of the VM for placing a VM. This approach visits the Physical Machines in a decreasing order of their volume and place VM to the first PM that has the enough resources. That means Physical Machine which has the maximum volume will be considered first.

Sandpipier [13, 14] is a Xen based automated system, which is used for detecting and mitigating the hot spot. It use the sand-volume to place the virtual machine, which is given by the formula

$$\text{Sand-volume} = \frac{1}{1-cpu} * \frac{1}{1-net} * \frac{1}{1-mem}$$

Where cpu, net and mem are the normalized utilization of the cpu, network and memory respectively. In this method Physical Machines are visited in the decreasing order of their sand-volume and place VM to the first PM that has the enough resources. So the Physical Machine which has the maximum sand-volume will be considered first.

This approach may select the wrong PM because they are not considered the shape of the resource utilization. Main reason to select the wrong PM is to convert 3D (CPU, network, memory) resource information into the 1D that is sand-volume. So if two PM having the same sand-volume then both physical machines are equally suitable for placing VM, this may not be possible.

3.1.4. Dot Product Based Fit

In this heuristic resource requirement of virtual machine and the total capacity of the physical machine along the specified dimensions are expressed as vectors. Dot product of these two vectors is calculated and then PMs are arranged into the decreasing order of their dot product.

VectorDot [7] method use the dot product of resource utilization vector ($\overrightarrow{RUV}$) of physical machine and resource requirement vector of virtual machine ($\overrightarrow{RRV}$) to select physical machine and choose the physical machine which having lowest dot product.

For the proper utilization of the resources it is necessary that the virtual machine which required more CPU and less memory should be placed on the physical machine which has low CPU and more memory utilization. This method seems good, but it can choose the wrong PM, because they the not using the length of virtual machine ($\overrightarrow{RRV}$) and the remaining capacity of the physical machine.

3.2 Constraint based approach

Constraint based approaches are use in the combinatorial search problems [15]. In these approach some constraint are apply and these constraint must be fulfil during VM placement. These constraints are

Capacity Constraints: For all dimension (CPU, memory and bandwidth) of a given physical machine, sum of the resource utilize by all VMs running in that host should be less than or equal to the total available capacity of that physical machine

$$\sum_{j=1}^{m_i} u^i_{cpu} < C^i_{cpu} \qquad \forall j \in m_i$$
$$\sum_{j=1}^{m_i} u^i_{mem} < C^i_{mem} \qquad \forall j \in m_i$$
$$\sum_{j=1}^{m_i} u^i_{bw} < C^i_{bw} \qquad \forall j \in m_i$$

Where C^i_{cpu} , C^i_{mem} and C^i_{bw} is the total cpu, memory and bandwidth capacity of the i^{th} host respectively and $u^i_{cpu,}$ $u^i_{mem,}$ and u^i_{bw} are the total CPU, Memory and bandwidth used by the all VM in the i^{th} host respectively.

Placement Constraints: All virtual machines must be placed on to the available host.

SLA constraint: VM should be placed to the PM where it fulfils the SLA.

Quality of services (QoS) constraint: Some quality of services constraint such as throughput, availability etc. must be considered during the VM placement.

Constraint Programming is useful where the input data is already known. That means we know the demands of the VMs, before calculating the cost function.

3.3 Bin packing problem

Bin packing problem [16] is a NP hard problem. If PM and the VM are consider as a three dimension object, then VM placement problem is similar to the bin packing problem where item represent the VM and container represent the PM. In the bin packing problem number of item (VM) are placed inside a large container (PM). The aim is to places a number of items into a single container as possible. So that the number of container required to packing the item is minimized.

Bin packing problem is different from the VM placement problem. In the bin packing problem bins can be placed side by side or one on top of the other. But in the case of VM placement, placing VMs side by side or one on top of the other is not a valid operation. This is because the resource cannot be reused by any other VM, once a resource is utilized by a VM.

3.4 Stochastic Integer Programming

Stochastic Integer Programming is used to optimize the problems, which involve some unknown data. VM placement problem can be consider as a Stochastic Integer Programming because resource demand of the VM are known or it can be estimated and the objective is to find the suitable host which consume less energy and minimize the resources wastages.

Stochastic Integer Programming can be use where the future demands of the resources are not known, but their probability distributions are known or it can be calculated.

3.5 Genetic Algorithm

Genetic algorithm is a global search heuristics. It is useful where objective functions changed dynamically. This approach is inspired by the evolutionary biology such as inheritance.

Genetic Algorithm can be use to solve the bin packing problem with some constraints. It is useful for the static VM placements, where the resource demands do not vary during a predefine period of time.

4. CONCLUSION

Efficient placement of the VM can improved the overall performance of the system. Virtual machine placement is a technique which maps the VM to the appropriate PM. Since the size of the data center is large in the cloud computing environment, so selecting a proper host for placing the VM is a very challenging task during the virtual machine migration. In this paper number of VM placement approach are explained with their anomalies. Each placement algorithm performs well under some specific conditions. So it is a critical task to choose a technique that is suitable for both the cloud user and cloud provider.

REFERENCES

[1] R. Buyya et al., "Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility", Future Generation Computer Systems, vol. 25, no. 6,june 2011, pp 599-616

[2] Mell, P. et al., "The NIST Definition of Cloud Computing". NIST Special Publication, 2011.

[3] Sosinsk et al. "Cloud Computing Bible", 1st ed., USA : Wiley Publishing Inc.,2012

[4] Kim et al., "Experimental Study to Improve Resource Utilization and Performance of Cloud Systems based on Grid Middleware" proceeding in first International Conference on Internet pp. 25-30, December 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] Mayank Mishra et al., "On Theory of VM Placement: Anomalies in Existing Methodologies and Their Mitigation Using a Novel Vector Based Approach", IEEE/ACM 4th international conference on cloud computing, pp 275-282, July 2011.

[8] Y. Wu et al. "Performance modeling of Virtual Machine Live Migration," IEEE/ACM 4th International conference on cloud computing, pp. 492-499, July 2011.

[9] Chris Hyser et al.," Autonomic virtual machine placement in the data center", Technical report, HP Labs, Feb. 2008.

[10] A. Beloglazov et al., "Energy efficient allocation of virtual machines in cloud data centers", proceeding in 10th IEEE/ACM Intl. Symp. on Cluster, Cloud and Grid Computing, pp. 577-578, may 2010.

[11] Lei Xu,Wenzhi et al., "Smart-DRS:A Strategy of Dynamic Resource Scheduling in Cloud Data Center", IEEE International conference on Cluster Computing Workshops, pp. 120-127, Sept..2012.

[12] N. Bobro et al., "Dynamic placement of virtual machines for managing SLA violations", proc. in 10th IEEE international symposium on integrated network management pp. 119-128, May 2007.

[13] C. Clark et al., "Live migration of virtual machine", proceeding of the 2nd conference on symposium on network system design and implementation, vol. 2, pp. 273-286, 2007.

[14] 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

[15] Constraint programming. http://en.wikipedia.org/wiki/Constraint_programming.

[16] http://en.wikipedia.org/wiki/Bin_packing_problem.

[17] N. Rodrigo et al. "A Heuristic for Mapping Virtual Machines and Links in Emulation Testbeds" in the proceeding of 9th IEEE international conference on parallel computing, pp. 518-525, 2009