International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 03 Issue: 06 | June -2019 ISSN: 2590-1892

Virtual Machines in Cloud Environment

Rishabh Narayan Parasar

Roorkee College of Engineering

Uttarakhand Technical University, Dehradun

Abstract-The proposed VMP-LR has three main components, namely, Resource Request Handling Component, Placement Component and Load Monitoring Component. VMP-LR considers three resources, namely, CPU, RAM and Bandwidth as relevant server resources in the context of VM placement, load balancing and rebalancing. The Request handling component groups resource requests into three queues, namely, high, low and medium resource requests queues. A simple rule-based algorithm is used to perform this categorization. As cloud is a dynamic environment, when new requests arrive after queue formation, they are joined to appropriate queue using Join Shortest Queue concept. The placement component is designed to be both traffic aware and load aware. During non-rush hours, as the number of requests is minimum, the VMP-LR uses a simple enhanced round robin method for placing VMs to PMs. The requests in high request queue are placed using an enhanced Max-Min, Ant Colony Optimization and Artificial Bee Colony. As the consequence of more and more virtual machines is packed onto a physical machine, the load imbalance factor increases, leading to the degradation of the performance of the cloud system. In order to solve this issue, the load monitoring component is used. The load monitoring component uses a load rebalancing algorithm to efficiently maintain load among VMs and PMs after placement. For this purpose, a hybrid algorithm that combines ant colony optimization with artificial bee colony algorithm is used. The proposed algorithms are implemented using Clouds Simulator and evaluated using seven performance metrics. They are throughput, response time, SLA violation rate, resource utilization rate, power usage, load imbalance rate and migration rate.

Keywords-Cloud Computing, Virtual Machines,PhysicalMachines.

Introduction

This expansion has particularly been accelerated by the twin developments of desktop computing and the World Wide Web . In these environments, the tasks are executed by applications hosted on dedicated servers. Increase in the availability of economical and sophisticated hardware/software often led to the spreading of these servers across datacentres. However, this has caused an increase in costs in terms of overhead and maintenance .

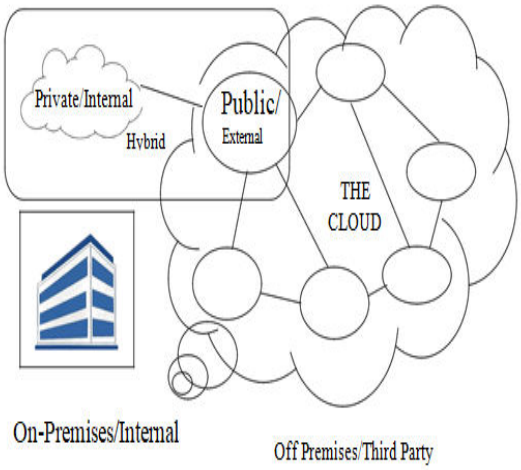
In this phenomenon, resource wastages or resource underutilization, have been envisaged. For example, CPU-extensive applications waste I/O resources and I/O-bound applications do not utilize CPU-resources fully, thus leading to under-utilization of resources. Cloud computing in the broad sense means the computing performed on web or internet. It has emerged as a favored computing model by several multinational organizations.

Examples include Google, Yahoo, Microsoft, Amazon, a nd IBM, all of which have built cloud platforms that host large data centers of networked machines available for the users to rent. From this definition, cloud computing can be considered as network accesses to

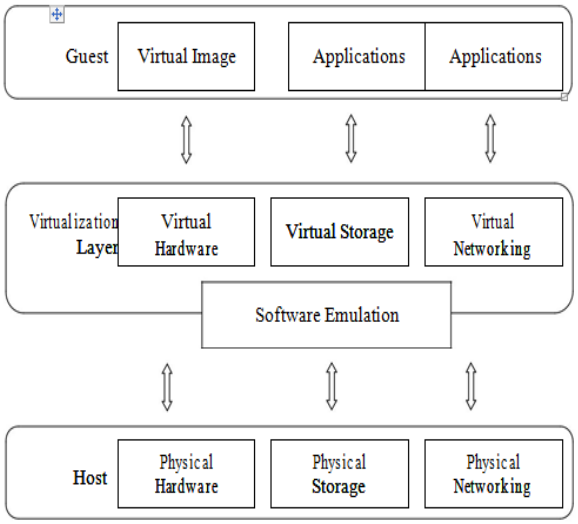
potential resources from a shared pool, which are provided in an on-demand basis in a manner that requires less response time, minimal effort from the management,minimal interaction from the service provider.

2. ***Service and Deployment Models***

Cloud computing technology allows developers and IT professionals with the ability to focus on significant matters and frees them from works like maintenance, procurement, and capacity planning. Deployment models refer to the placement and management of the cloud infrastructure, while service models consist of varieties of services that the user can access a cloud computing platform. Each type of cloud service and deployment method provides different levels of control, flexibility, and management. All these three services can be deployed in four different ways, namely, private, public, community and hybrid cloud.



3. ***Virtualization-***Virtualization is the ability to run multiple operating systems on a single physical system and share the underlying hardware resources . In a virtualized environment and , there are three main components, namely, host, virtualization layer and guest, as shown in Figure 1.6. The virtualization layer component recreates the environment where the guest will operate. Virtualization technology is considered to be one of the most important factors behind the scalability characteristic of cloud computing . One of its attractive features is the ability to utilize computing power more efficiently.Specifically, virtualization provides an opportunity to consolidate multiple VM instances running on under-utilized computers into fewer hosts, enabling many of the computers to be turned off, and thereby resulting in substantial energy savings. Here, the resources of one PM are partitioned into a pool of logical resources and rearranged into VMs. This shows a significant increase in the utilization of a single PM by running heterogeneous application stacks on one and the same machine. This results in huge time and effort saving along with scalability and reusability.



Without virtualization, all machines require the same power, emit same heat and need the same physical space. Moreover, the setup cost, maintenance overhead, support overhead, cost per hardware, etc. are also reduced but are directly proportional to the number of

machines. Thus, the multiple advantages obtained through the use of virtualization can be summarized as follows

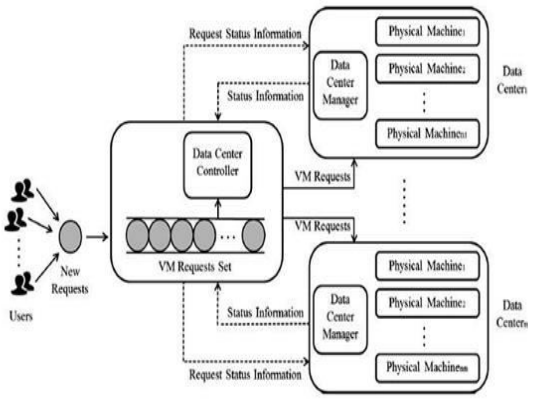
- Increases resource utilization by sharing physical resources among multiple users and applications.
- Sharing of resources helps cost reduction
- Provides several desirable characteristics like isolation, encapsulation, hardware independence and portability
- Improve IT throughput and costs by using physical resources as a pool from which virtual resources can be allocated.

4. ***Virtual Machine Placement-*** Virtual machine placement is the process of mapping virtual machines to PMs. In other words, virtual machine placement is the process of selecting the most suitable host for the virtual machine. It is the decision to place a particular VM to a particular host. The autonomic virtual machine placement algorithms are designed keeping in mind the above goals.

   A large number of PMs are deployed in DCs. A cloud service provider can have multiple DCs. Thus, the VM requests should be first distributed optimally over DCs and then the requests in each DC are distributed over PMs.

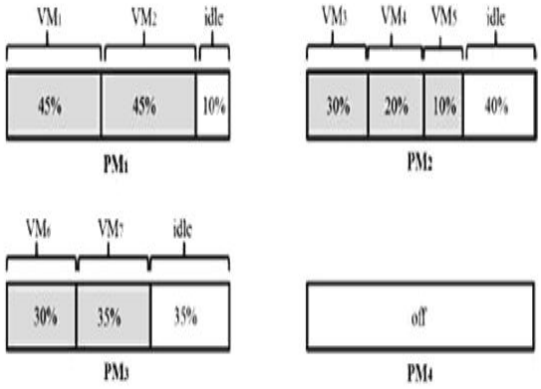
5. ***Categories of VM placement-*** The VM placement can be static virtual machine placement or dynamic virtual machine placement

   Static virtual machine placement- Static placement of VMs is performed either during system startup or in offline mode. This is the initial placement of VMs in the cloud computing environment, where no prior mapping of VMs is made.

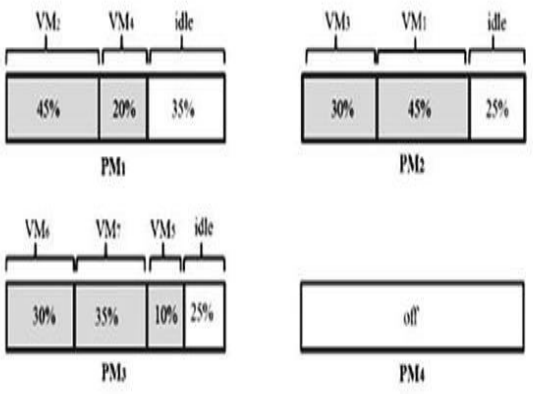


Cloud users use the services provided by the cloud service provider and issue requests. These requests are in the form of VM requests since every request will be completed on a VM on top of a PM. The VM requests comprise the VM Requests Set. The VM requests contain information of the resources (CPU, RAM, Network etc.) needed in order to complete the request. The DC controller optimally distributes VM requests to the DC managers following a heuristic algorithm. This approach to distribute VM requests uses a centralized approach, where the decision to schedule requests depends on the central controller.

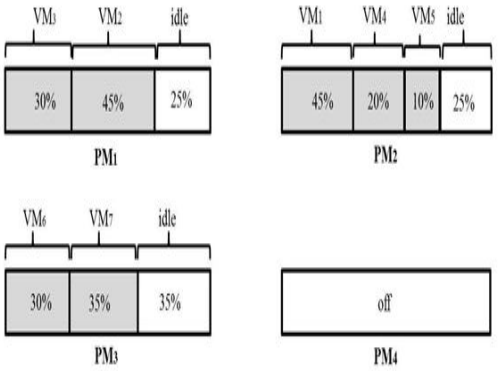
Dynamic Virtual Machine Placement- Dynamic placement of VMs places VMs based on an existing mapping which is in contrast to the static placement of VMs which starts with no mapping. This would not shut down or stop the already running VMs, thus the placement solution should provide the list of live migrations to be executed in order to reach the optimal state from the existing state. The whole process is in contrast to the static placement of VMs where VMs can be stopped and restarted which increases the energy consumption and thus degrades the complete system performance.



(a) Initial Solution 'i'



(b) Target Solution 's_1'



(b) Target Solution 's_2'

in order to reach solution „s1" from „i", live migrations need to be performed. One example list of migrations might be VM5 from PM2 to PM3, migration of VM1 from PM1 to PM2 and finally, VM4 from PM2 to PM1. Since there isn"t much space in PM1 and PM2 for the direct interchange of VM1 and VM3, there needs to be an involvement of PM3. Thus the migrations needed for solution s2 to be reached are the extra migration of VM3 from PM2 to PM3 and VM1"s migration from PM1 to PM2 and migrating VM3 from PM3 to PM1. But reaching s2 from i will result in more cost than to reach solution s1 from the initial placement i because of the migrations involving larger VMs. VM1 and VM3 could have been stopped and restarted in order to facilitate direct interchange involving more cost. The involvement of PM4 for the interchange of VM1 and VM3 also results in higher costs since it is considered to be the cost of state change for PM4 later.

6. ***Conclusion-***Recent years have cloud computing technology emerging as a commodious resource of computation power. Virtual machine placement is the process of mapping resource requests in the form of VMs to PMs in an optimal manner that improves resource utilization and reduces time involved in providing the service desired. This work proposes a system called Virtual Machine Placement and Load Rebalancing Based on Multi-Dimensional Resource Characteristics in Cloud Computing Systems (VMP-LR) to improve the VM placement process. They are, Resource Request Handling Component, Placement Component and Load Monitoring Component. VMP-LR is designed using a two phase methodology, where the first phase handles the tasks involved with Resource Request Handling and Placement Components, while Phase II handles the tasks of Load Monitoring Component. The resource request handling component in Phase I of the research work, manages the multi-dimensional resource requests from users in the form of VM requests. VMP-LR considers three resources, namely, CPU, RAM and Bandwidth as relevant server resources in the context of VM placement, load balancing and rebalancing and the requests are structured as 3-dimensional cubic vector. This component interacts with the resource repository and VM manager of the cloud system along with the other two components of the VMP-LR to aid in improving the process of VM placement. The main aim of the request handling component is to group the request into three queues, namely, high resource request queue, medium resource request queue and low resource request queue.

7. ***References-***

Xu, G., Pang, J. and Fu, X. (2013) A Load Balancing Model Based on Cloud Partitioning for the Public Load, Tsinghua Science and Technology, Vol. 18, No. 1, Pp. 34-39.

Xu, J. and Fortes, J. (2010) Multi-objective virtual machine placement in virtualized data center environments, Proceedings of the IEEE/ACM International Conference on Green Computing and Communications and International Conference on Cyber, Physical and Social Computing, Pp.1-8.

Xue, S.J. and Wu, W. (2012) Scheduling Workflow in Cloud Computing Based on Hybrid Particle Swarm Algorithm, TELKOMNIKA Indonesian Journal of Electrical Engineering, Vol.10, No.7, Pp. 1560-1566.

Yadav, R.K., Mishra, A.K., Prakash, N. and Sharma, H. (2010) An Improved Round Robin Scheduling Algorithm for CPU scheduling,

International Journal on Computer Science and Engineering, Vol. 2, No. 4, Pp. 1064-1066 .

Yang Y., Zhou Y., Sun Z. and Cruickshank H. (2013) Heuristic scheduling algorithms for allocation of virtualized network and computing resources, Journal of Software Engineering and Applications, Vol. 6, No. 1, Pp. 1-13.

Yao, J. and He, J.H. (2012) Load Balancing Strategy of Cloud Computing based on Artificial Bee Algorithm, 8th International Conference on Computing Technology and Information Management, Vol. 1, Pp. 185-189.

Zahedani, D.S. and Dastghaibyfard, G. (2014) A hybrid batch job scheduling algorithm for grid environment. 4th IEEE International e-Conference on Computer and Knowledge Engineering, Pp. 763-768.

Zhan, S. and Uo, H.H. (2012) Improved PSO-based Task Scheduling Algorithm in Cloud Computing, Journal of Information and Computational Science, Pp.3821-3829.

Zhang, L., Chen, Y., Sun, R., Jing, S. and Yang, B. (2008) A task scheduling algorithm based on PSO for grid computing, International Journal of Computational Intelligence Research, Vol.4, No.1, Pp. 37-43.

Zhang, L., Zhuang, Y. and Zhu, W. (2013) Constraint Programming based Virtual Cloud Resources Allocation Model, International Journal of Hybrid Information Technology, Vol. 6, No.6, Pp. 333-344.

Zhang, Q., Zhani, M.F., Zhang, S., Zhu, Q., Boutaba, R. and Hellerstein, J.L. (2012) Dynamic energy-aware capacity provisioning for cloud computing environments, ACM Proceedings of the 9th International Conference on Autonomic Computing, Pp. 145-154.