RESEARCH REVIEW JOURNALS
Volume-05
Issue-01
January -2020

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com[Peer Reviewed Journal]

A Study of Virtual Machineries Technology in Cloud Computing

Omkar Ramesh Ghatage

Department of Computer Science Mansarovar Global University (India)

ARTICLE DETAILS

Article History
Published Online: 16 Jan 2020

Keywords
Virtual Machine, Technology, Cloud Computing, Information Technology.

ABSTRACT

Cloud computing, along with social, mobile and analytic technologies, has revolutionized the Information Technology (IT) industry by enabling elastic on-demand provisioning of computing resources. According to Right Scale Survey, 82% of enterprises reported a cloud usage strategy as compared to 74% in 2014, which shows a 9.76% growth in cloud usage. On the other hand, predicts that there will be an increase of 41.3% of cloud usage at the end of 2016. The statistical report of indicates that more than 82% of the companies have envisaged tremendous cost benefits after migrating to cloud environment. The same report also estimates that that the global DC traffic will grow three-fold, while global traffic growth will increase by 3.5 fold in the next year. These statistics and predictions show cloud as an essential strategy with high service quality expectations in terms of safety, easy accessibility, cost and maximum resource utilization.In cloud computing, one area of research that has attracted researchers, industrialists and academicians equally is "Resource Management". According to a data center uses four stages to perform efficient management of resources. They are, VM provisioning, resource provisioning (includes mapping and scheduling requests onto PMs), run-time management and workload modeling. In this research work, the focus in on VM provisioning, which apart from mapping VMs to appropriate PMs should alsotry to satisfy both client SLA for Qos and reduce operating costs, which is a challenging task for cloud service providers. More often, the client service providers are faced with challenges of under-provisioning (a starvation or saturation of VM resources that leads to service degradation) and over-provisioning (underutilization and subsequent waste of VM resources). Under-provisioning often leads to SLA penalty resulting into business revenue loss on the part of the cloud providers and also a poor Quality of Experience (QoE) for the cloud client's customers (unacceptable response time for time critical applications for example). On the other hand, over-provisioning can lead to excessive energy consumption, culminating in high operating cost and waste of resources.

1. Introduction

Virtualization, a key concept of cloud computing, is the ability to run multiple operating systems on a single Physical System (PM) and share the underlying hardware resources. It is the process by which one computer hosts the appearance of many computers). Usage of virtualization increases the resource utilization by sharing physical resources among multiple users and applications. Sharing of resources provides multiadvantages like cost reduction, provides isolation, encapsulation, hardware independence and portability. Virtualization in cloud computing involves three broad stages, namely, application profiling, generation of Virtual Machine (VM) configurations and VM placement. This research work is focused on the Stage 3, that is, VM Placement.VM placement is defined as the process of mapping the VM requests to the PMs, according to the availability of resources in these hosts. VMs must be distributed in an efficient way such that no system or a request starves for the response from cloud. The primary goal in VM placementtask is to maximize the usage of the available resources. However, new challenges have cropped up due to the ability to host multiple applications (VMs) onto the same PMs, and migrate them seamlessly across different servers,. These challenges include balancing load amongst all PMs, figuring out which VMs to place on which PMs and handling sudden surges in resource demands. The main objective of this research work is to address these challenges and provide solutions to solve them.

As a solution to the above situation, cloud computing and virtualization have been evolved. Cloud computing has gained popularity due to its various characteristics like cost effective and on-demand / pay as use services that are independent of time and geographical locations. It is a general term used to describe a collection/group of integrated and networked hardware, software and internet infrastructure and describes a platform that hides the complexity and details of the underlying infrastructure from users and applications by providing very simple graphical interface or API (Applications Programming Interface). Cloud computing systems consist of several elements that include virtualization, distributed computing, service oriented architectures, broadband networks, browser as a platform, servers, SAN/NAS (Storage Area Network/Network Attached Storage) and free and open source software. Each of these elements performs tasks with the main goal of offering better services to the users. Out of various elements, the focal point of this research is on the task of virtualization.

2. Literature Review

(Gahlawat et al. 2014) presented a brief survey of the main cloud federation architectures and approaches considered for

VMP problem formulation in this particular scenario. It is important to remember that cloud federation is the practice of voluntarily interconnecting cloud infrastructures of different Cloud Service Providers (CSPs), mostly to respond to workload peaks. Several authors have surveyed the existing methods by grouping the methodologies reported according to some common goals or characteristics of the algorithms.

(Li et al. 2013) broadly categorized VM placement into two methods, namely, direct placement and migration-based placement, based on the fact that the time taken to complete a job depends upon the type of VM placement. Studies have also considered VM placement process as either single objective or multi-objective problem. Static Server Allocation Problem (SSAP) algorithm, Static Server Allocation Problem with variable workload (SSAPv) algorithm, Dynamic Server Allocation Problem (DSAP) algorithm.

According to (Medhat et al. 2013), scheduling requests in cloud environment was a combinatorial optimization problem, where the aim was to look for an object from a finite set. This object was typically an integer number, a subset, a permutation or a graph structure. Recently, a more advanced kind of approximation algorithms have been prominently used, which basically tried to combine basic heuristic methods in higher level frameworks that aim at effectively exploring a search space. This class of algorithms was termed as metaheuristic swarm intelligence and evolutionary algorithms.

(Banerjee et al. 2009) used a modified ACO for scheduling and load balancing. The basic pheromone updation formula was modified to improve the load balancing problem and was designed for dynamic heterogeneous system. On the other hand, (Zhang et al. 2010) proposed the usage of ACO along with network theory to perform scheduling and load balancing for open cloud computing systems. The algorithm used small-world and scale free characteristics of the complex network to improve load balancing and was designed to work with heterogeneity systems. They proved that their algorithm provided good scalability and high fault tolerance.

(Ali et al. 2010) also proposed ACO algorithm for load balancing in distributed systems. This algorithm was designed for fully distributed systems where information is updated dynamically at each ant movement. Paradigm of multiple colonies was adopted such that each node will send a colored colony throughout the network and which were then used to prevent ants of the same nest from following the same route and hence enforcing them to be distributed all over the nodes in the system and each ant acts like a mobile agent that carries newly updated load balancing information to the next visited node.

Later, (Li et al., 2011) proposed an algorithm called Load Balancing using ACO algorithm (LBACO) to find optimal resource and task allocation for dynamic cloud environment. The objective function minimized the task completion time and was designed to work with distributed cloud environment. (Nishant et al., 2012) also modified ACO to improve the synchronization of ants and to improve load balancing process. The modified ACO resulted with the following advantages.

(Mishra et al., 2012) developed an effective load balancing algorithm using ACO to maximize or minimize different performance parameters like CPU load, memory capacity, delay or network load for the clouds of different sizes.

(Gao et al., 2013) also proposed a multi-objective ACO algorithm for the virtual machine placement problem, where the goal was to efficiently obtain a set of non-dominated solutions that simultaneously minimized the total resource wastage and power consumption.

(Suseela et al., 2014) presented another multi-objective algorithm that combined the advantages of ACO and PSO to form an hybrid model for VM placement for cloud computing systems. This system was designed to reduce resource wastages, minimize power consumption and perform optimal load balancing. The hybridization was formed by using the output of ACO algorithm as input to PSO algorithm. ACO algorithm finds VM placement solutions by considering resource wastage and power consumption in each server and PSO algorithm finds VM placement solution when considering fault tolerance through load balancing in each server. ACO algorithm and PSOalgorithm are used as local search algorithm and global search algorithm respectively in this proposed hybrid algorithm.

(Gupta et al., 2014), on the other hand, used ACO for performing load balancing in cloud environment. They proved that consideration of foraging and trailing pheromones help in optimizing the traversing process of ants. They proved that usage of ACO for load balancing can reduce the number of migrations and improve time complexity. (Khan et al., 2014) implemented an algorithm to improve the process of load balancing using ACO for cloud computing systems. (Shilpa et al., 2014) reviewed load balancing algorithms based on partitioning for cloud computing systems.

Recently, (Patel et al., 2016) presented a review of various methodologies based on ACO for scheduling in cloud computing. (Gupta et al., 2016) presented a trust and deadline aware scheduling algorithm for cloud infrastructure using ACO.

Anatomy of Cloud Computing

The important elements of a cloud computing system are shown in Figure 1.Virtualization component in a node is provided by an infrastructure layer known as Hypervisor (also called Virtual Machine Monitor or VMM). This layer provides the interface to execute multiple instances of operating system at the same time on a single PM. The hypervisor creates objects known as Virtual Machines which encapsulate operating system, configuration and applications. Device emulation is also provided to the PM either in hypervisor or as a VM. Virtual machine management takes place both locally in different PMs and globally in a Data Center (DC).

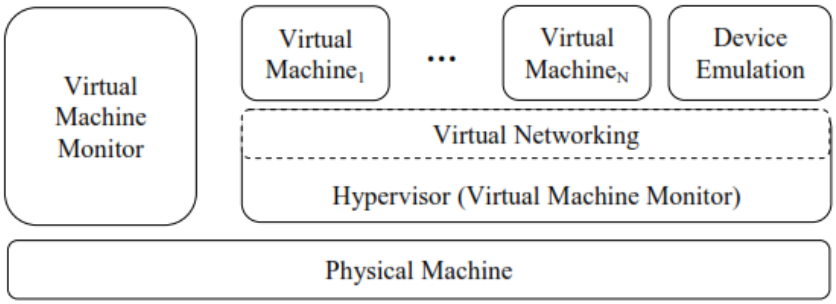

Figure 1: Elements in Cloud System

The nodes represented in Figure 1 are then multiplied on a physical network with management orchestration over the entire infrastructure to form a Data Center (DC) as shown in Figure 2.

Hypervisors: Hypervisor acts as the base level of a PM or node, manages the execution of the guest operating systems by providing them with a virtual operating platform. These are known as virtual machines. The VMs share the virtualized hardware resources of the node.

Device Emulation: Hypervisors provide platform where VMs can share the virtualized physical resources. But, the whole node needs to be virtualized, by a device emulator in order to provide full virtualization.

Virtual Networking: Networking needs of a system increase as more and more VMs consolidate on physical servers. Thus, instead of VMs communicating in the physical level, the whole network is also virtualized in order to reduce the load on the physical infrastructure. Virtual switches are introduced in order to optimally communicate between VMs.

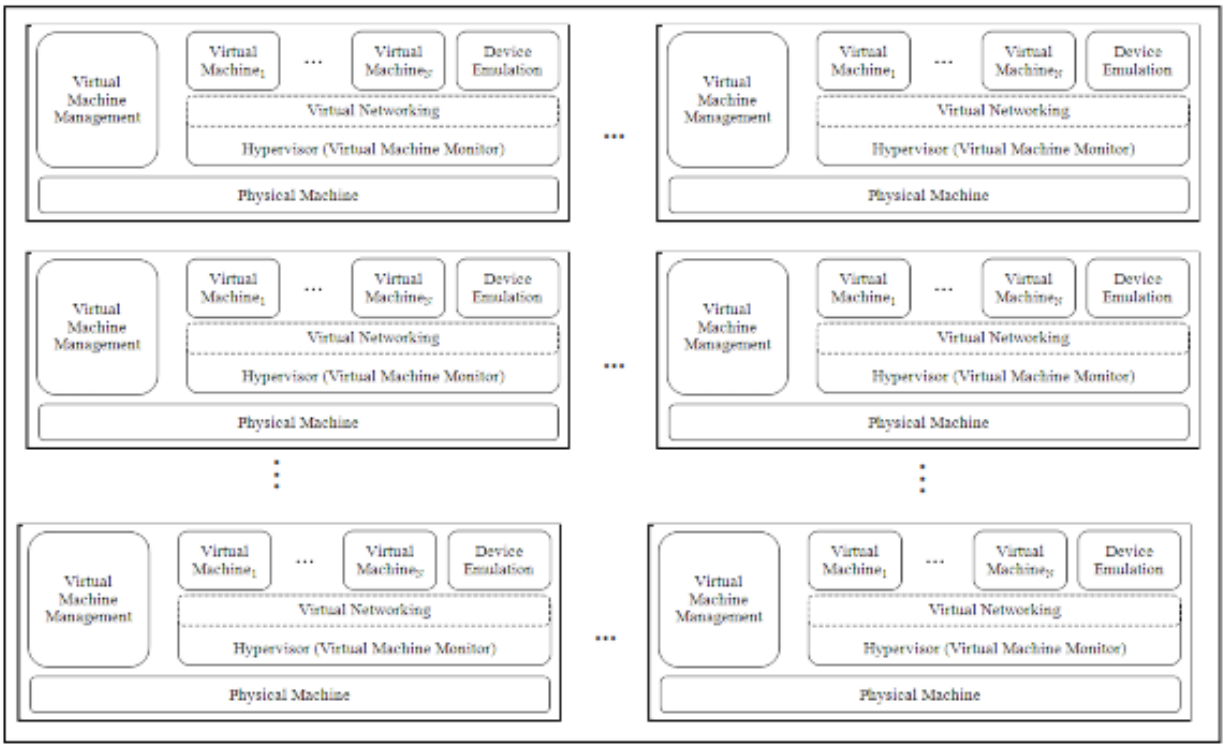

Figure 2: Representation of Data Centre

3. Cloud Computing Architecture

In a cloud computing environment, the users depend on cloud service providers to satisfy their needs and demands. Thus, some QoS (Quality of Service) parameters must be maintained by the cloud providers which are cataloged in the Service Level Agreement (SLA). In order to achieve this, market oriented architecture, as shown in Figure 3, is needed instead of the traditional resource management architecture. Users or brokers (acting on behalf of users) submit their requests to the cloud DC in order to be processed. SLA Resource Allocator acts as the interface between the DC and the users/brokers. On submission of a user request, the Service Request Examiner checks the request for QoS parameters‟ requirements to make a decision whether to accept or reject the user request. It obtains information from the VM Monitor on the availability of resources and already available workload from the Service Request Monitor, thus ensuring that no overloading occurs. Considering these parameters, the examiner assigns the user requests to VMs and decides the allotted VMs‟ resource requirements.

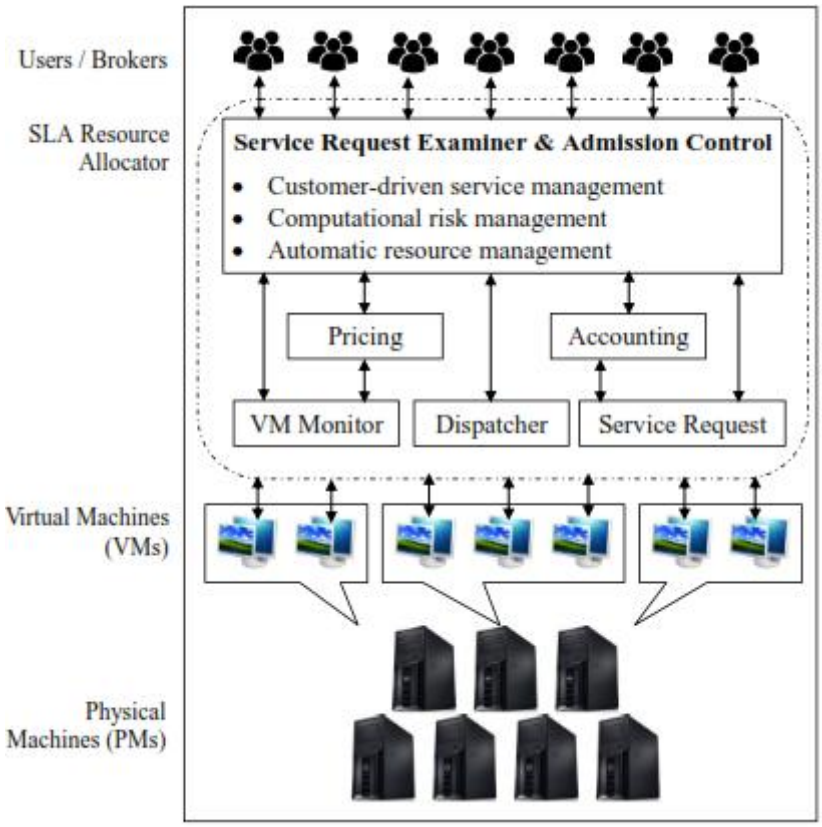

Figure 3: Cloud Computing Architecture

The Pricing operation makes decisions on the prices of service requests; that is whether to charge the request on the basis of time of submission, or resource availability or simply based on fixed rates. This mechanism aids in effective prioritization of allocation of resources in a DC. Accounting operationis used to keep a tab on the actual resource consumption, according to which the pricing mechanism can calculate the final cost to be charged from the users. It also helps in effective resource allocation decisions to be made by the service examiner by using the historical resource usage information kept by the accounting mechanism.The VM Monitor keeps a check on the resource requirements and VMs" availability. Dispatcher starts executing the accepted user requests on the allotted VMs. Service Request Monitor keeps a tab on the status of the execution of the service requests.

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. As cloud computing has grown in popularity, several different service models and deployment strategies have emerged to help meet specific needs of different users. 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 on a cloud computing platform. Each type of cloud service and deployment method provides different levels of control, flexibility and management. There are three fundamental service models in Cloud computing, namely, Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). All these three services can be deployed in four different ways, namely, private, public, community and hybrid cloud.

4. Virtualization

Virtualization is the ability to run multiple operating systems on a single physical system and share the underlying hardware resources.It has become a key technology in DCs and cloud computing and helps to provide resource sharing characteristics, scalability, dynamic architectures and also provide the ability of VM migration between PMs for load balancing. It is used as a base for provisioning services to the client. According to SNIA, virtualization is defined as, “The act of abstracting, hiding, or isolating the internal functions of a storage (sub)system or service from applications, host computers, or general network resources, for the purpose of enabling application and network - independent management of storage or data”. It is the process by which one computer hosts the appearance of many computers. In a virtualized environment and there are three main components, namely, host, virtualization layer and guest, as shown in Figure 4. The host component represents the original physical system where the guest systems are managed. The virtualization layer component recreates the environment where the guest will operate. The guest component interacts with the virtualization layer directly rather than host system. 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 compute 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

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 a huge time and effort saving along with scalability and reusability.

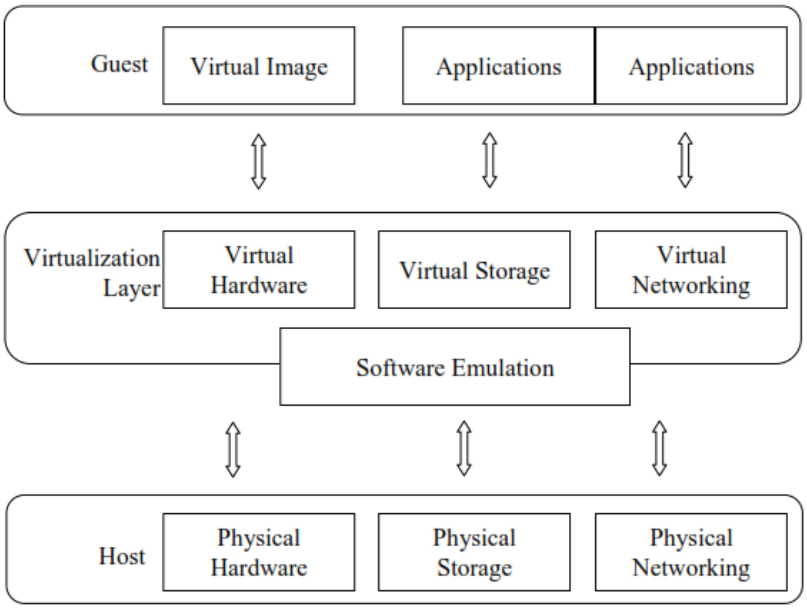

Figure 4: Virtualization Reference Model

Without virtualization, all machines require same power, emit same heat and need 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. Figure 5 shows the difference between servers with and without virtualization. Thus, the multiple advantages obtained through the usage of virtualization can be summarized as follows.

- Increases the 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.

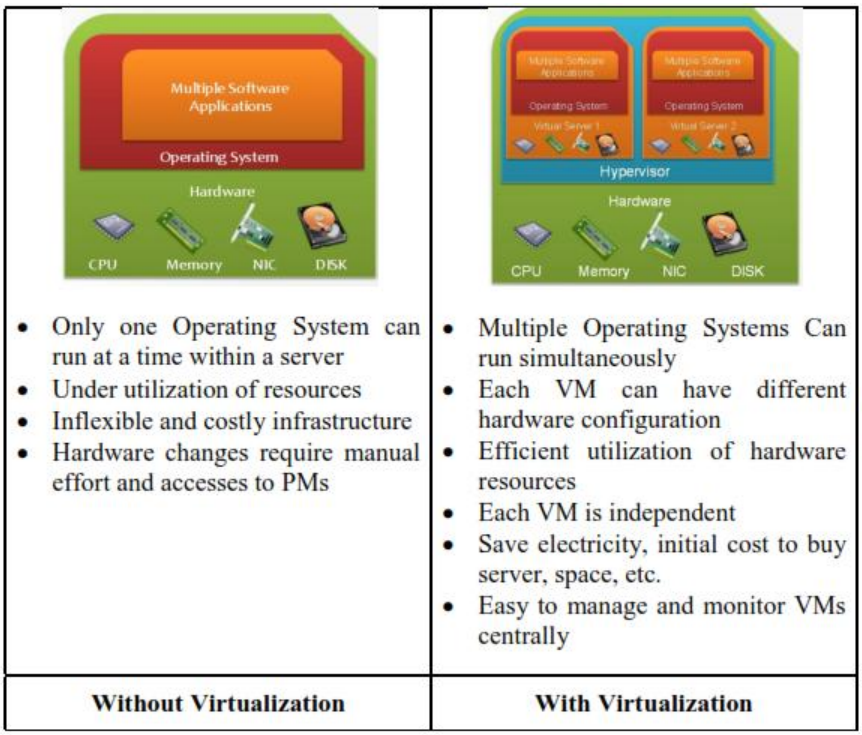

Figure 5: With and Without Virtualization

Virtualization technology has two main elements, VM and VMM. VM is an isolated runtime environment (guest OS and applications). Multiple virtual systems (VMs) can run on a single PM. The VMM is a program that allows multiple operating systems to share a single hardware host. Each guest operating system appears to have the host's processor, memory, and other resources all to itself. However, the VMM is actually controlling the host processor and resources allocating what is needed to each operating system in turn and making sure that the guest operating systems (or VMs) cannot disrupt each other.

There are three types of virtualization techniques, namely, server virtualization, client (or desktop) virtualization and storage virtualization (Figure 6).

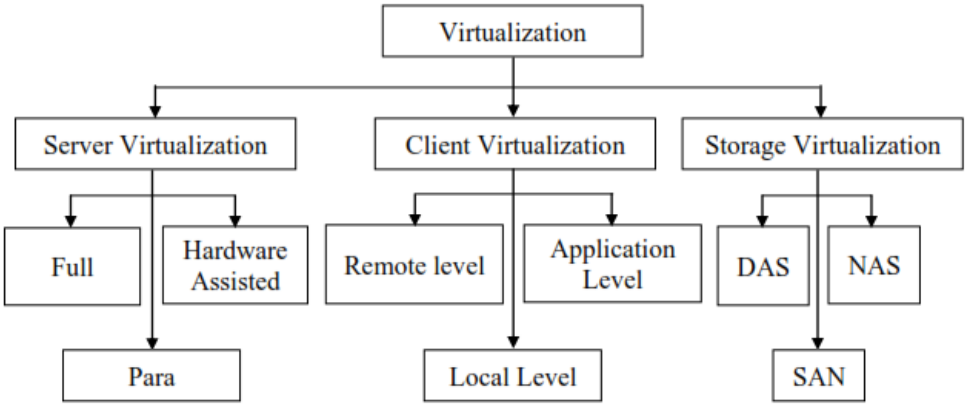

Figure 6: Types of Virtualization

In server virtualization, a single server performs the task of multiple servers by portioning out the resources of an individual server across multienvironment. The hypervisor layer allows for hosting multiple applications and operating systems locally or remotely. The advantages of virtualization include cost savings, lower capital expenses, high availability and efficient use of resources. With cloud computing, there are three important types of virtualization, namely, full virtualization, paravirtualization and hardware assisted virtualization. In full virtualization, a complete installation of one machine is carried out on another machine and results in a VM with all the software that are present in the actual server. Here, the remote datacenter delivers the services in a fully virtualized manner.

The execution environment of virtualization techniques can be classified into two forms, namely, process-level and system level. Process level virtualization is implemented on top of an existing system, while system level virtualization is implemented directly on hardware with minimum requirement of existing operating system.

Virtualization in cloud computing involves three broad stages.

- **Stage 1 - Application Profiling:** In this stage, the applications are profiled in its physical environment in order to obtain its resource utilization.
- **Stage 2 - Generation of VM Configurations:** In this stage, the above obtained knowledge is used to generate configurations for VMs.
- **Stage 3 - VM Placement:** This stage uses the generated VM configurations to identify the optimal manner of mapping them onto PMs.

In general terms, virtualization is defined as a technology such that there is a software abstraction layer between the hardware and operating system and applications running on top of it. This software abstraction layer, as described earlier, is called VMM. The VMM hides the physical resources from the operating system because hardware resources are controlled by the VMM. This is the reason that user can have two or more operating systems running on the same machine in parallel. Therefore hardware can be partitioned into two or more logical units called virtual machine (VM).

5. VM 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. Virtualization technology provides strong isolation amongst the VMs. For example, security isolation prevents one virtual machine from accessing data of another and fault isolation prevents failure in one machine to impact the other. The process involves categorizing the VMs hardware and resources requirements and the anticipated usage of resources and the placement goal. The placement goal can either be maximizing the usage of available resources or it can be saving the power by being able to shut down some servers. The autonomic virtual machine placement algorithms are designed keeping in mind the above goals. The VM requests follow a two-tier distribution approach. 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.

6. Conclusion

The important to have a VM placement algorithm along with a loading monitoring algorithm that can perform VM provisioning in an time-efficient manner and that can also take care under- and over-utilization of resources efficiently so as to increase the efficiency of the cloud computing system. When the number of VMs and PMs was small, mapping of VMs to appropriate PMs would be possible manually. However, the current scenario faces a tremendous increase in the number of VMs and PMs, which makes automation of placement task mandatory. Existing automated solutions have to evaluate

several numbers of possible mappings for a given set of VMs and PMs and thus, require the improved intelligent placement heuristics to narrow down the search for a solution to obtain near-optimal placement plans. Moreover, the following issues have been identified in the existing solutions.Thus, developing an accurate VM placement algorithm has become crucial and also considered as a challenging research area where the goal is to address these two extremes. Furthermore, VM boot up time has been reported to span various time durations before it is ready to operate specifically from between 5 and 10 minutes; Amazon Elastic Compute Cloud, and between 5 and 15 minutes. It is believed that during this time of system and resource unavailability, requests cannot be serviced which can lead to penalty on the part of the cloud providers. Multiplying this lag time over several server instantiations in a data center can result in heavy cumulative penalties. These penalties or compensations to the client cannot redeem the poor QoE which the customers must have perceived. Thus, it is important to have a VM placement algorithm along with a loading monitoring algorithm that can perform VM provisioning in a time-efficient manner and that can also take care under- and over-utilization of resources efficiently so as to increase the efficiency of the cloud computing system.

References

1. Tawalhen, L., Darwazeh, N.S., Al-Qassas, R. and Aldosari, F. (2015) A Secure Cloud Computing Model based on Data Classification, The 6th International Conference on Ambient Systems, Networks and Technologies (ANT-2015), the 5th International Conference on Sustainable Energy Information Technology (SEIT-2015), Procedia Computer Science, Vol. 52, Pp. 1153-1158.
2. Shayan, J., Azarnik, A., Chuprat, S., Karamizadeh, S. and Alizadeh, M. (2013) Identifying benefits and risks associated with utilizing cloud computing, International Journal of Soft Computing and Software Engineering, Vol. 3, No. 3, Pp. 416-421.
3. Obasuyi, G.C. and Sari, A. (2015) Security challenges of virtualization hypervisors in virtualized hardware environment, International Journal of Communications, Network and System Sciences, Vol. 8, Pp. 260-273.
4. Tamane, S. (2015) A review on virtualization : A cloud technology, International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 3, Issue 7, Pp. 4582-4585.
5. Janani, N., Shiva, R.D., and Prakash, P.(2015) Optimization of virtual machine placement in cloud environment using genetic algorithm, Research Journal of Applied Sciences, Engineering and Technology, Vol. 10, No. 3, PP. 274-287.
6. Gahlawat, M. and Sharma, P. (2014) Survey of virtual machine placement in federated clouds, 2014 IEEE International Conference on Advance Computing (IACC), Pp. 735–738.
7. Medhat, A.T., Ashraf, B.E., Arabi, E.K. and Fawzy, A.T. (2013) Cloud Task Scheduling Based on Ant Colony Optimization, 8th International Conference on Computer Engineering & Systems ICCES, Pp. 64-68.
8. Banerjee, S., Mukherjee, I. and Mahanti, P.K. (2009) Cloud Computing Initiative using Modified Ant Colony Framework, World Academy of Science and Technology, Vol. 56, Pp. 221-224.
9. Mishra, M. and Sahoo, A. (2011) On theory of VM placement: Anomalies in existing methodologies and their mitigation using a novel vector based approach, IEEE International Conference on Cloud Computing, Pp. 275-282.
10. Gao, R. and Wu, J. (2015) Dynamic Load Balancing Strategy for Cloud Computing with Ant Colony Optimization, Future Internet, Vol. 7, Pp. 465- 483.
11. Suseela, B.B.J. and Jeyakrishnan, V. (2014) A multi-objective hybrid ACOPSO optimization algorithm for virtual machine placement in cloud computing, International Journal of Research in Engineering and Technology, Vol. 03, Issue 04, Pp. 474-476.
12. Gupta, E. and Deshpande, V. (2014) A Load Balancing Technique for Servers of Datacenter of Cloud using Ant Colony Optimization, International Journal of Recent Advances in Engineering & Technology, Vol. 2, Issue 6 & 7, Pp.44-47.
13. Khan, S. and Sharma, N. (2014) Effective Scheduling Algorithm for Load balancing using Ant Colony Optimization in Cloud Computing, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, Issue 2, Pp. 1-8.
14. Shilpa D. and Chaudhari, S. (2014) Reviews of Load Balancing Based on Partitioning in Cloud Computing, International Journal of Com puter Science and Information Technologies, Vol. 5, No. 3, Pp. 3965-3967.
15. Sarma, V.A.K., Rajendra, R., Dheepan, P. and Kumar, K.S.S. (2015) An Optimal Ant Colony Algorithm for Efficient VM Placement, Indian Journal of Science and Technology, Vol 8, No. S2, Pp. 156–159.
16. Patel, M. and Kadian, R. (2016) A Review on ACO based Scheduling Algorithm in Cloud Computing, International Journal of Computer Science and Mobile Computing, Vol. 5, Issue 5, Pp. 489-493

