IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 20, Issue 4, Ver. III (Jul - Aug 2018), PP 23-32
www.iosrjournals.org

Virtualization and Scheduling In Cloud Computing Environment – A Study

Manoj Mathew
ITIL V3 Certified, PRINCE2 Practitioner, IT Consultant, Bangalore, Karnataka, India
Corresponding Author: Manoj Mathew

***Abstract*:** *Cloud computing is the latest technology that is ruling the IT world with faster internet connectivity to the rage of 5G's. The intention behind the cloud computing is to provide effective and instant access to the widely diverged resources, irrespective to their geographical location based on the arrived request. The availability of faster internet connectivity makes the cloud offerings more adaptable. The Cloud solutions are categorized into three major groups - Public, Private and Hybrid. The Cloud provides a various set of applications, infrastructures, solutions, and services that comprises of both computational as well as storage facilities. Such cloud grows every day and scheduling the resources among the requesting processes (client) seems to have numerous issues due to lack of proper dynamic scheduling plan and no certain policies in order to regulate the order by which the computing environment must execute the jobs. This paper gives a wide idea on the virtual machine scheduling algorithms that are in existence. It also discusses the advantages and disadvantages.*

Keywords: *Cloud Computing, IaaS, SaaS, PaaS, Virtual Machine, Scheduling Algorithms, Private Cloud, Public Cloud, Hybrid Cloud*

Date of Submission: 13-08-2018 Date of acceptance: 30-08-2018

I. Introduction

Cloud computing is a distributed parallel computing environment over datacenter resources. The major objective behind cloud computing was to move the desktops to a service-oriented platform. This is achieved by using clusters of servers for providing services and a host of datacenters across geographies containing the huge databases. Cloud computing model provides shared access to the resources at anytime from anywhere. The services are provided dynamically, at a low cost based on the demand by its customers. The elasticity and simplicity characteristics of the cloud attract billions and billions of users towards it on a daily basis. In the market, there are so many public cloud service providers' available providing different resources as per the demand. Amazon , Microsoft Azure, Orange, Google, etc are leading cloud providers. Figure 1 shows the eagle eye view of cloud computing.

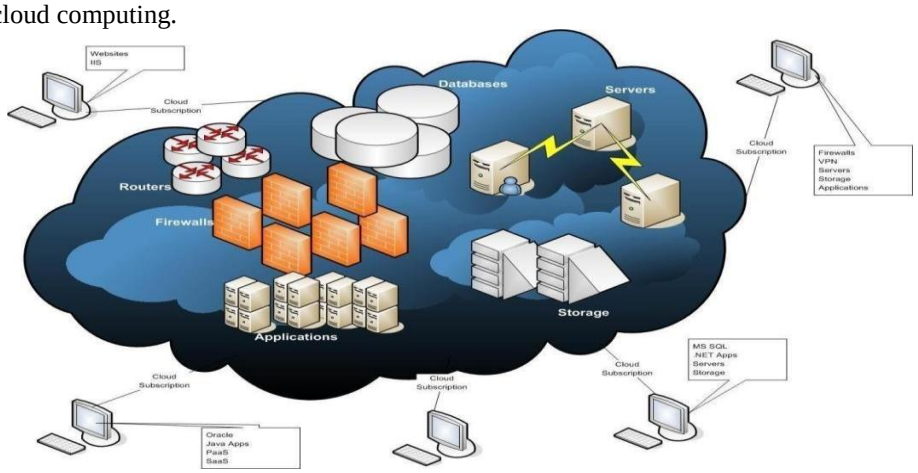

Figure 1 – Cloud – Eagle Eye View

a) Characteristics of Cloud

Resource Pooling – Shared pool of resources are provided by cloud for providing rapid services with minimal efforts. The users can draw a resource from any remote data center according to the requirements.

On Demand Self Service – Refers to the services that are requested by the customers for managing their own computing resources. The customers can manage their own computing resources and these services are provided by the cloud service provider via Internet.

Rapid Elasticity – Since rapid services are provided by the cloud service provider via internet, the services can be easily managed as per the growing or shrinking demand peeks.

Measured Service – The basic objective of cloud is providing service on demand for defined services and pay as per usage. So, customers can request for the services as required and pay for the services as billed by the cloud service provider.

b) Types of Clouds

Private Cloud

A solution in which compute, storage, memory and all related infrastructure are hosted in one or a few private datacenters and accessed seamlessly via dedicated Internet with controls established at datacenter level. Such services are generally meant for corporate employees or to their selected clients/customers via dedicated Public IP's or registered Uniform Resource Locators'. These are expensive but provide flexibility to users and security to the information stored at thesedatacenters.

Public Cloud

A most common cloud solution in which compute, storage, memory are hosted across the globe and can be accessed seamlessly via Public Internet. This is available for general public with a subscription to these service providers. These are widely accepted and used due to the flexibility but security is a concern to it due to the uncertainty of data location. The datacenters are spread across the globe and provides seamless to access to services any point in time.

Hybrid Cloud

Hybrid cloud is a mix of Public and Private Cloud, in which selected compute, storage, memory are hosted across the globe and accessed seamlessly via Public Internet. Access to these are restricted to a very extend but not fully contained like Private Cloud. These are required when a few part of information is shared to Public while the other part is restricted for general public. The datacenters may spread across the globe and provides faultless access to services any point in time with controls to access certain confidential data.

c) Cloud Service Models

The Cloud Services through virtualization is available on various forms in the markets and it's named as following (a) Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS).

Infrastructure as a Service (IaaS)

Infrastructure as a Service forms the base layer for cloud computing. It provides massive computing power and unlimited storage area for large scale customers without requiring any physical hardware on site. It basically deals with servers, storage, virtual machines, networks, load balances. The IaaS service providers supply these resources to the customers' on-demand basis. The minimal requirement for building IaaS cloud includes: Hypervisor - VMM (Virtual Machine Monitor) and Networking Topology, could be either private or public. IaaS maintains the hardware at the local level and mitigates the need for a data center and sometimes it is known as Hardware as a Service (HaaS). Examples include Amazon Web Service (AWS) [15], Windows Azure [16], Rackspace, etc. Amazon Web Services is a cloud computing platform which consists of a collection of remote computing services. Popular customers like Netflix, SmugMug, Pinterest, NASDAQ are benefited by AWS.

Software as a Service (SaaS)

SaaS aka on-demand software options are predominantly for the software's published on cloud and end users pay it as per usage. SaaS is a boon for the business fraternity as it reduces the IT operational costs by outsourcing their hardware and software support as well as maintenance to the cloud service provider. These services could be accessed via internet by using a web browser or an application or a light-weight desktop/VDI (Virtual Desktops Infrastructure) or through mobile applications. The end users data is stored over cloud and easily accessible irrespective of locations Microsoft's Office 365 is a perfect example of SaaS service over Cloud, which is making the Emails and Office Packages/Software's like Excel, Word, and Power Point etc... are

delivered over cloud on a pay-per-use model.

Platform as a Service (PaaS)

PaaS provides computing platforms that instantly scales in order to meet the application demands. PaaS includes resources like database, operating system, programming language, web server that is provided as service to the customer on demand. In this model, developers with proprietary APIs of an application will run on specific environment. It further controls the configuration settings and software deployment. PaaS reduces deployment complexity and cost by casting off the need to buy the required hardware and software.

Virtualization in Cloud Computing

Virtualization is a software layer that provides a consistent interface for the cloud computing architecture. It helps in providing portable and ease management for decoupling the software system from the existing hardware components. Such virtualization technique reduces the inventory, upkeep and maintenance of physical hardware's thus reducing the capital expenses. It is achieved through consolidating the compute, memory and storages, by reducing the physical space that is required for the individual data centers.

The concept of virtualization has been a major point of discussions in IT industry in last few decades it become more relevant for business operations. The virtualization technology makes the cloud computing environment infrastructure to be elastic as well as scalable. Due to elasticity and scalable in nature, scheduling became inevitable. VMware ESX, Oracle Virtual Box, Microsoft HyperV are a few virtualization products widely used in industry. Figure 2 shows the hypervisor model.

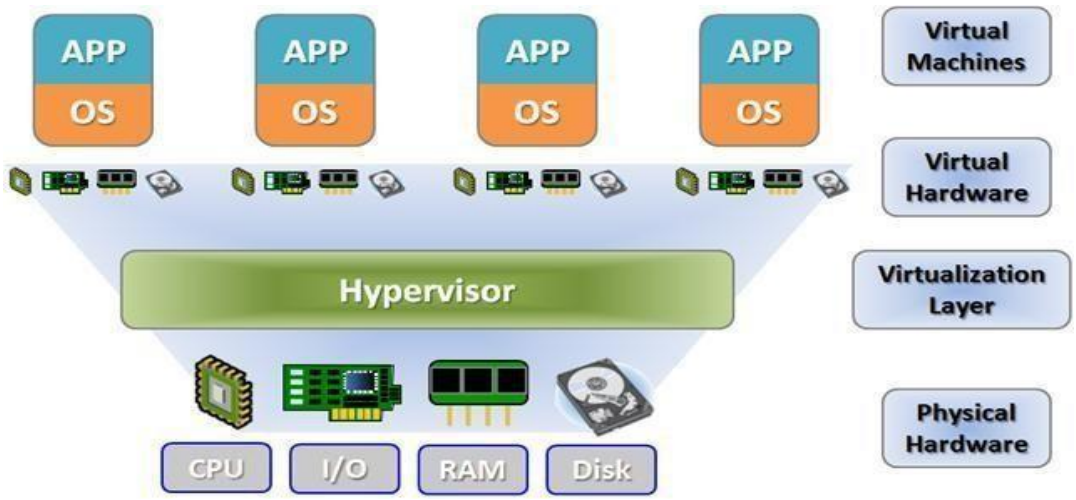

Figure 2 Hypervisor Model

d) Benefits of Virtualization

There are numerous benefits provided by virtualization that includes:

- Reduction in cost of IT infrastructure
- Quicker and Flexible allocation of resources
- Efficient utilization of resources
- Better accessibility
- Minimization of risk
- Improvement in productivity
- Remote access
- Rapid scalability
- Enables running multiple operating system
- Reduce the risk of data loss as the data is backed up across multiple storage locations
- Lower the hardware, manpower requirements and energy consumption
- Easy migration
- Maintainability
- High availability
- Fault tolerance

Types of Virtualization in Cloud Computing

Depending upon the type of application that is used and the hardware's need to be utilized for completing the computation, virtualization takes many. The main types are as follows:

Hardware Virtualization

Hardware virtualization aka server virtualization or hardware-assisted virtualization is a common type of virtualization that runs on individual independent segment hardware or on a physical server. The underlying idea is to combine as many as small physical servers into a large physical server in such a way that the processors can utilize it more effectively. The hypervisor controls and co-ordinates the memory, processor and other components attached to the machine by allowing the different operating system to run on the same machine without any need for source code. Figure 3 shows the hardware virtualization. Hardware virtualization is further subdivided into three types as:

i. Full Virtualization - The entire simulation of the actual hardware takes placein order to allow the software to run as an unmodified guest operatingsystem.
ii. Para Virtualization - The software unmodified runs in modified operating system as a separate system.
iii. Partial Virtualization - The software requires being modified for running.

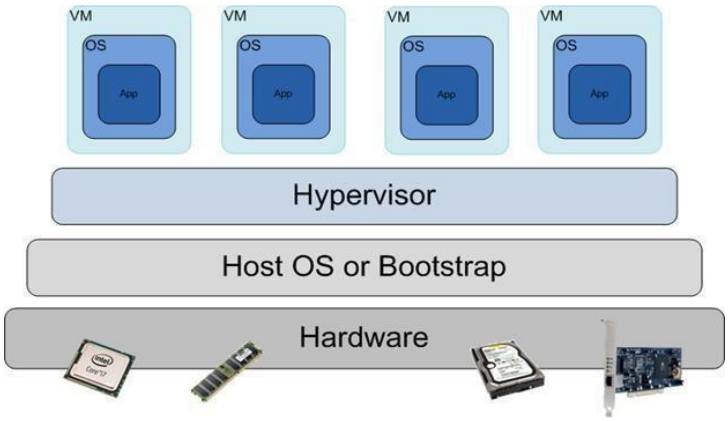

Figure 3 Hardware Virtualization [1]

e) Network Virtualization

Network virtualization refers to the management and monitoring of a computer network as a single managerial entity from a single software-based administrator's console. It provides many networking administrative tasks via automation techniques, reduces the limitations, investments & complexities of IT Network Change Management [17]. Network virtualization provides the network optimization for transferring data, scalability, flexibility, reliability and security. It is specifically useful for huge networks that experiences rapid and unpredictable network traffic rate when it increases. The objective of network virtualization is to provide an enhancement in network productivity as well as to improve the efficiency. Figure 4 shows the network virtualization. It is further categorized as:

i. Internal network enable a single system to function like a huge network.
ii. External network combines more than one network or parts of networks into a single virtual unit or segregate a sing network into multiple unit, based on the requirement.

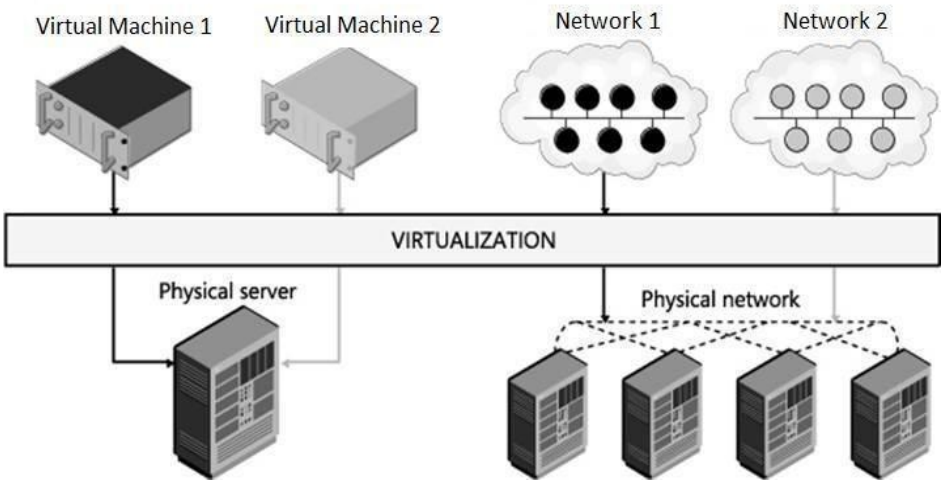

Figure 4 Network Virtualization

a) Storage Virtualization

a. In storage virtualization, multiple network storage resources are presented as a single device for storage. It provides an efficient and easier management of the required resources. It provides several advantages as better availability, improved storage management in heterogeneous environment, better storage utilization, automated management, reduced downtime, etc. Figure 5 shows the storage virtualization. In general, the storage virtualization is of two subtypes:

Block Virtualization – It was in existence, before the file system where multiple storage devices are consolidated into one for providing storage.

File Virtualization – Grants access to the files that are stored over multiple hosts. For enabling file-level usage, the server that uses this kind of storage must have the software installed unit.

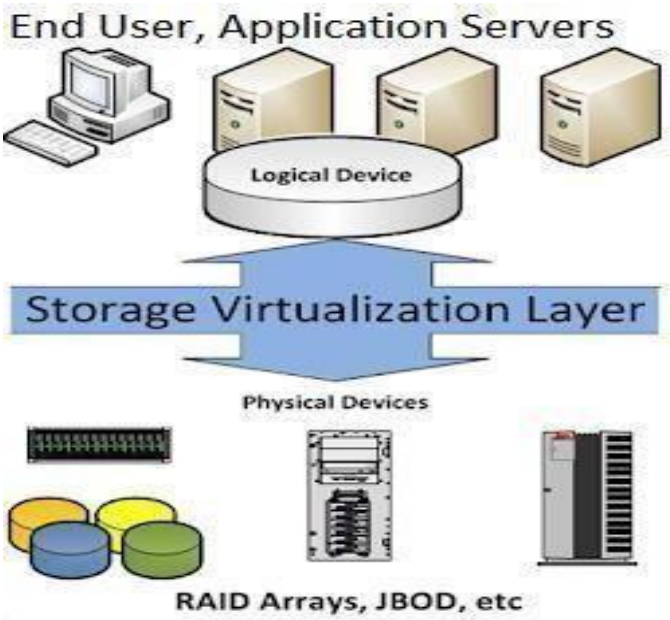

Figure 5 Storage Virtualization [2]

a) Memory Virtualization

Memory virtualization provides a single virtualized memory pool by aggregating the physical memory that exists across different servers. The provided memory pool can be distributed, shared or networked in function. It enhances the performance by providing an enlarged greater contiguous memory space without providing any additional components to the main memory. That is a familiar virtualization that can be found in the Microsoft Windows OS, where major portion of the storage disk drive serves as an extension of main memory. There are two subtypes of this virtualization as:

i) Application-level Integration – The applications that are running on the connected computers can access the memory pool directly through the file system or the provided Application Program Interface. Figure 6 shows the memory virtualization of application level integration.

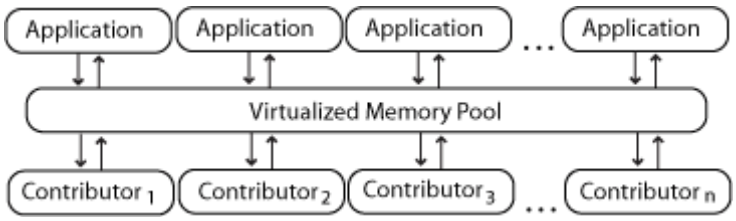

Figure 6 Memory Virtualization - Application Level Integration [3]

ii) Operating System Level Integration – Provides access to the memory pool through the operating system, by first connecting the operating system to the memory pool. Figure 7 shows the memory virtualization of addressable memory integration.

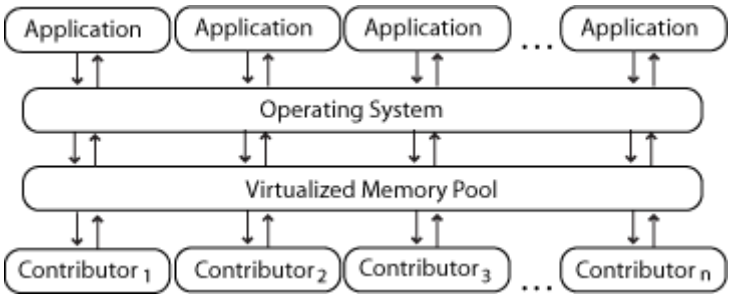

Figure 7 Memory Virtualization – Addressable Memory Integration [3]

b) Software Virtualization

Software Virtualization provides the host machine to creation one or more virtual environments and run on it. It enables the guest operating system with a complete hardware and to run on the host machine. For example, the Android operating system or Linux hosted on a natively running Microsoft Windows Operating system can utilize the hardware as the host machine does without any constraints (or vice versa, running Microsoft Windows on Linux as a guest machine). The software virtualization is categorized into three sub typesas:

i) Operating System Virtualization – Hosting more than one Operating System on the native Operating System.

ii) Application Virtualization – Hosting individual applications separated fromthe native Operating System in a virtual environment.

iii) Service Virtualization – Hosting specific services and processes that arerelated to a particular application.

c) Data Virtualization

Data virtualization is presented as an abstract layer that is completely independent of any implementation details such as data structure or database system. It ease the manipulation of the data without knowing the technical details of it like how its formatted, where it is located physically etc. It helps in decreasing the workload and data errors. Figure 8 shows the data virtualization.

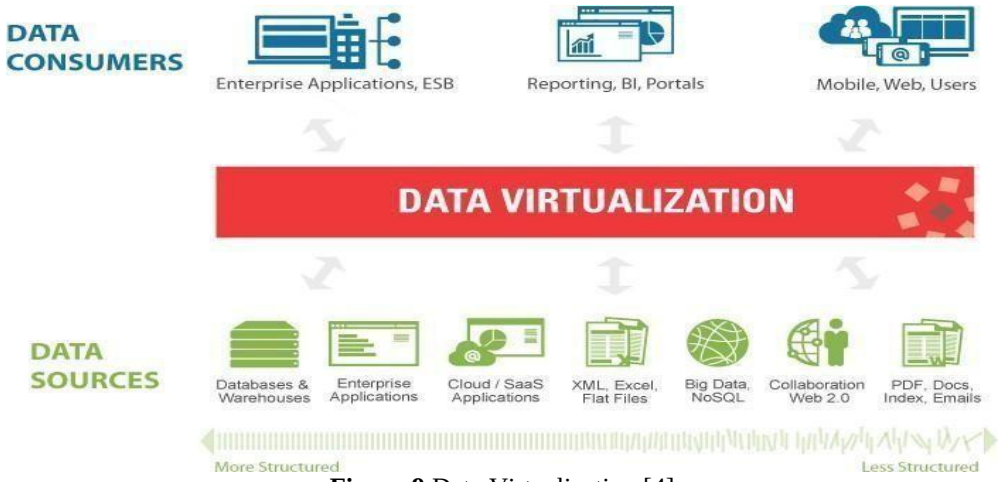

Figure 9 Data Virtualization [4]

d) Desktop Virtualization

Desktop Virtualization is one of the most common forms of virtualization that is in existence for any regular IT employee where the user's desktop is stored on a remote server, which allows the user to access his/her desktop from any remote location through any device. It thus provides a convenient and secured work as if the employee might feel he/she is working from their home. It provides data confidentiality by keeping the data stored securely on the central server and minimizes the data theft risk rate. Figure 10 shows the desktop virtualization.

Figure 10 Desktop Virtualization [5]

II. Virtual Machine Scheduling

The major objective of scheduling in distributed systems is to maximize the processor utilization and minimize the task execution time by distributing the processor load when a need arises in the dynamic environment. Job scheduling, an optimized problem, plays a vital role for improving the reliability and flexibility of the systems [17]. The role of scheduling algorithms is to find out a proper sequence in which the jobs can be executed under the constraints within a reasonable time [6]. There are two major categories of scheduling algorithms as: static scheduling algorithm and dynamic scheduling algorithm.

Static Scheduling Algorithm	**Dynamic Scheduling Algorithm**
Suitable for homogeneous and stable environment	Suitable for heterogeneous and dynamic environment
Gives low performance and might have lots of overhead compared with dynamic scheduling algorithm	Gives higher performance than static scheduling algorithm
Attributes are mostly not taken into consideration during execution	Considers different types of attributes in the system both prior and during the execution time
As it is more suitable for static environment, this scheduling might provide good results in such environment	Some of the considered attributes might make the system complicated and inefficient as well as over head during execution and might degrade the services provided

When virtualization is achieved, for effective management of the resources, a scheduler is put into place for utilizing the resources. The important issue that needs to be addressed in computing environment is scheduling [7]. In cloud environment, scheduler's responsibility is to order the arriving jobs in such a manner that it must maintain the fairness among the requesting job as well as must schedule those in an efficient way without compromising the Quality of Service (QoS).

In the cloud computing environment, virtual machine scheduling algorithms are used for scheduling the Virtual Machine request to the Physical Machines (PM) that belongs to a particular Data Center (DC). The scheduling is done as per the requirements received and based on the resources/compute availability i.e, RAM, Bandwidth, Storage, Processors etc.

The scheduling criterions to be considered while performing VM scheduling are:

- Resource Utilization – It is used for testing the resource utilization. It is expected maximum for an efficient scheduling algorithm.
- Throughput – It is defined as the number of tasks executed at a fixed interval of time.
- Response Time – It can be defined as the amount of time taken by the system to produce the first output from the time of submission. It is expected to be reduced for a better performance.
- Performance – It is defined as the system's efficiency and the scheduling algorithm must be able to improve it.
- Scalability – It is used for measuring the quality of service as the number of nodes increases in the computing network. It is expected to be the same even if maximum numbers of nodes are getting added. The nodes must be added without affecting the services.
- Association of overhead – Overhead occurs due to Interprocess communication and task movement. The overhead must be reduced and the loads must be balanced properly. The algorithm must work well to satisfy the criteria.
- Fault Tolerance – It is defined as the ability of the system to perform uniform balancing of the load across the available resources in spite of failure of the node(s).

III. Virtual Machine Scheduling Algorithms

Cloud computing is rapidly growing and clients demands more and more service. As a result of this scheduling the jobs and balancing the load during real time has become a challenging task. Many algorithms are there in existence for providing effective and efficient algorithms for assigning the client's request to the available cloud cluster nodes. The major objective of scheduling algorithms in a distributed environment is to distribute the works in a balanced way across the existing processors, thus maximizing the utilization and minimizing the time of execution. The important parameters that are related to virtual machines are time that is required to create a new virtual machine, number of applications using the virtual machine, time taken for allocating the additional resources for scheduling the virtual machine basic processing unit in the computing environment. The goal of scheduling in virtual machine is to share the resource for better utilization and fasten the execution of jobs with minimum consumption of energy.

a) Round Robin Algorithm

The round robin algorithm's major objective is to distribute the jobs equally to all of the existing nodes. The scheduler allocates one virtual machine to one node in a cyclic manner. This algorithm ensures fairness in allocation of jobs to the available virtual machine and utilizes the resources in a balanced manner.

b) Weighted Round Robin Algorithm

Supreeth and Shobha [8] proposed an algorithm for scheduling virtual machine's processing in the cloud computing environment known as weighted round robin algorithm. This algorithm works on the basis of weight's allocated. The incoming requests are allocated to the available virtual machine in a round robin manner that takes the weight into consideration. This algorithm does not consider the current load that is allocated to the virtual machine at the given instance. This might make a machine overloaded where as another machine under loaded.

c) Priority Scheduling Algorithm

The general idea behind priority based scheduling algorithm is, each VM is assigned with an internally defined priority based on the characteristics such as the amount of work it can do, time taken for executing the job, and it is allowed to run on each machine. Instances with equal priority are scheduled in FCFS order. Such assigned priorities can be changed dynamically using the aging technique, where the priority of the VM is increased based on the total amount of time VM remains in the ready queue for execution. By implementing such technique, the priority of the VM keeps increasing and at one point of time it reaches a higher priority. This helps in preventing starvation of an instance. Such priority algorithm can be designed either as pre-emptive one

or non preemptive one based on the application area. Priority based scheduling can also be improvised in terms of execution time by implementing SJF policy, where the priority is the inverse of the next CPU burst [9].

d) Gang Scheduling Algorithm

Gang scheduling algorithm is an efficient job scheduling algorithm for parallel system that schedules related VM to run simultaneously on different machines. It is well suited for time sharing system. Gang scheduling algorithm has already been applied in distributed and parallel systems as well. The major advantage of this algorithm in cloud environment is that it selects the best suited resource for executing a task at a given instance by considering the static and dynamic parameters and the necessary restrictions need to be considered [10]. It also allows the parallel jobs to communicate with each other when required. The communication overheads and context switching overheads sometime introduces unwanted chaos and leads to lower the execution time that results in lowering the response time. Such unnecessary delays must be identified and eliminated. Even then the research on gang scheduling has show that there is an improved throughput in the time sharing system [11] which is a good sign for implementing this algorithm for such type of systems. Gang scheduling can be applied in cloud environment which could improve the cost as well as performance.

e) Genetic Algorithm

The objective behind implementing genetic algorithm in cloud computing VM machine scheduling algorithm is to minimize the make span. It is basically used for scheduling sets of independent VMs with a goal to reduce the scheduling time. Initially, many individual solutions are randomly generated that aids in forming the initial population. Size of the population depends on the number of VMs to be run on the system and the type. During each successive generation, a new breed of generation is created by selecting a proportion of the existing population. The implementation of genetic algorithm involves two steps:

Step 1: Select individual solution through a fitness-based process

Step 2: Generate a second generation population of solution from the selected one through genetic operators like crossover and mutation.

Repeat step 1 and step 2 until an optimal solution is achieved by satisfying the criteria of minimum response time. Genetic algorithm increases the time and space cost. It provides improved response time and throughput by parallel execution. It also improves the quality of service. The disadvantage of this algorithm is the migration cost becomes higher.

f) Adaptive Algorithm

Adaptive algorithm uses dynamic priority for notes based on which the VM's are scheduled and assigned for execution. Priority is assigned based on the load factor and capacity of the VMs [12]. Using the priority, the VM's are scheduled, that could vary dynamically based on the load factor. This algorithm is quiet efficient and it provides a right balance between the performance and power efficiency. The dynamic priority concept leads to better resource utilization. It also helps in improving the throughput, and achieves high bandwidth utilization.

g) Efficient Resource Utilization Algorithm

Efficient Resource Utilization Algorithm (ERUA) is a 3-tier cloud architecture based algorithm [13] that considers Consumers, Service Provider and Resource Provider. It aims at providing a Quality of Service (QoS) to the user and cost benefit to service provider. The objective is achieved by providing an effective resource reallocation based on the utilization ratio and enhancing the rate. This is done by the service providers by hiring the resources that are available with the resource provider and the instances of VMs are created dynamically as per the request in order to provide customer service. Customers will based on pay-as-you use policy and the service providers use the massive pool of existing resources and schedule it dynamically on demand.

h) Renewable Energy Source Provisioned Algorithm

The objective of this algorithm is to schedule the VM in a network of cloud that are deployed in different geographical areas in such a way that the cost of total power consumption is reduced for the operator [14]. The cloud providers create a VM with resource requirements for the request and execute the VM before the deadline approaches. This approach helps in preventing the bandwidth from getting wasted. Maximum throughput can be achieved by this scheduling method. The disadvantage of this algorithm is if the required resources are not allocated immediately to the requested job, there is a policy that the dead line of the job to be completed may not be met and couldn't satisfy the objective.

IV. Conclusion

IT industry is going through a challenging phase where the cost of operation is getting much in to the radar of CIO/CTO's, adding real-estate cost, electricity, power and other components to keep a datacenter operated for 24/7 is a huge overload for even major companies. So, the adoption of cloud and cloud services are recommended to manage the CAPEX (Capital Expenditure/Investment) and run business operations on a "Pay as You Go" model. Barring a few data security concerns, cloud can be a major breakthrough due to its easy adaptability, scalability and flexibility, with much required redundancy without huge expense or initial investments. The virtualization and automation scripts in cloud help to streamline the resource utilization. The companies may follow Public/Private/Hybrid Cloud models according to their data security standards. So, over all adoption of SaaS, PaaS, IaaS models and schedule those services via algorithms are beneficial for the organization's financial stability. In simple terms, scheduling is an automated job that happens in the background of cloud computing environment. The scheduling algorithms are meant to improve the resource utilization and must increase the performance. The different scheduling algorithms are discussed in this paper that effectively schedules the jobs in the computing background.

References

[1]. https://blog.cloudpassage.com/2011/07/22/the-difference- between-virtualization-and-cloud- computing/
[2]. http://www.cloudcomputingat.com/2016/05/storage- virtualization.html
[3]. https://en.wikipedia.org/wiki/Memory_virtualization
[4]. https://www.denodo.com/en/data- virtualization/overview?utm_source=BDA-Blog
[5]. https://www.znetlive.com/blog/virtualization-in-cloud- computing/
[6]. Pinal Salot , "A Survey Of Various Scheduling Algorithm In Cloud Computing Environment" in M.E, Computer Engineering, Alpha College of Engineering, Gujarat, India , Volume: 2Issue:2
[7]. Hadi Salimi, "Advantages, Challenges and Optimizations of Virtual Machine Scheduling in Cloud Computing Environments" in International Journal of Computer Theory and Engineering Vol. 4, No. 2, April 2012.
[8]. Supreeth S , Shobha Biradar , "Scheduling Virtual Machines for Load balancing in Cloud Computing Platform" International Journal of Science and Research (IJSR), India, Vol-2 Issue- 6,
[9]. June -2013
[10]. Vignesh V, Sendhil Kumar KS, Jaisankar N , " Resource management and scheduling in cloud environment" in International Journal of Scientific and Research Publications, Volume 3, Issue 6, June 2013.
[11]. Dr. Chenna Reddy, "An Efficient Profit-based Job Scheduling Strategy for Service Providers in Cloud Computing Systems" in International Journal of Application or Innovation in Engineering and Management (IJAIEM), Volume 2, Issue 1, January 2013.
[12]. Junliang Chen, Bing Bing Zhou, "Throughput Enhancement through Selective Time Sharing and Dynamic Grouping" in 2013 IEEE 27th International Symposium on Parallel and Distributed Processing.
[13]. Kiran Kumar et. al., "An Adaptive Algorithm For Dynamic Priority Based Virtual Machine Scheduling In Cloud" in IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 6, No 2, November 2012
[14]. Ramkumar N, Nivethitha , "Efficient Resource Utilization Algorithm (ERUA) for Service Request Scheduling in Cloud" in International Journal of Engineering and Technology (IJET) , Vol 5 No 2 Apr-May 2013.
[15]. Dimitris Hatzopoulos, "Dynamic Virtual Machine Allocation in Cloud Server Facility Systems with Renewable Energy Sources" at IEEE International Conference on Communications (ICC) 2013, Budapest, Hungary.
[16]. Kumar, Rakesh, and Shilpi Charu. "Comparison between Cloud Computing, Grid Computing, Cluster Computing and Virtualization." International Journal of Modern Computer Science and Applications **3**(1): 42-47.
[17]. "What is cloud computing?" http://searchcloudcomputing.techtarget.com/sDefinition/ 0sid201gci 1287881, 00.html.
[18]. Manoj Mathew, "Information Technology Change Management in Networking", International Journal of Latest Trends in Engineering and Technology (IJLTET), Vol 1, Issue 1, pp. 176-181, May 2016.
[19]. H. Yuan, C. Li, M. Du, "Optimal Virtual Machine Resources Scheduling Based on Improved Particle Swarm Optimization in Cloud Computing", JOURNAL OF SOFTWARE, 2014.

Manoj Mathew." Virtualization and Scheduling In Cloud Computing Environment – A Study ." IOSR Journal of Computer Engineering (IOSR-JCE) 20.4 (2018): 23-32.