A Review on Virtual Machine Scheduling Algorithms and Techniques in Cloud Computing

Subhash B. Malewar, Prof.Deepak Kapgate

Abstract— **Cloud Computing provide the various services to the user on pay as you use mode as per requirement. The role of Virtual Machine's (VMs) has emerged as an important issue because, through virtualization technology, it makes cloud computing infrastructures to be scalable. Load Balancing is essential for effective operations in distributed environments hence in cloud it is very important research. There are various algorithms present for scheduling the clients request to appropriate available cloud node. In this paper we explain various virtual machine management techniques and the scheduling algorithms also compare them based on some performance matrices to provide an overview of latest technique and algorithms in the cloud computing field for virtual machine scheduling.**

Index Terms— **Cloud Computing, scheduling techniques, VM scheduling algorithms**

I. INTRODUCTION

Cloud computing is the delivery of computing services over the Internet. Cloud services allow individuals and businesses to use software and hardware that are managed by third parties at remote locations. In IT infrastructure cloud computing appears to be very important for the users to organize their applications in a distributed environment. This cloud computing enhances scalability, fault tolerance and availability. The Cloud computing is the delivery of computer resources through a Web service interface [1]. The term "cloud" refers to the organization of the underlying physical infrastructure remaining opaque (not visible) to the end user. In other words, cloud computing gives a user access to computer resources (i.e. machines, storage, operating systems, application development environments**,** application programs) Over a network through Web services, while the actual physical location and organization of the equipment hosting these resources be it in the next room or spread across the globe is not necessarily known to the user. Cloud computing delivers infrastructure, platform, and software (application) as services, which are made available as subscription-based services in a pay-as-you-go model to consumer. The services provided by cloud computing are:

A) Infrastructure as a Service (IaaS): Cloud Infrastructure as a Service (IaaS) consists of shared data-centers, virtual hardware and appearing as a single point of access for consumer's computing needs. IaaS can deliver software, data centre space, virtualization platforms and network instruments with advantages like flexibility, scalability and cost effectiveness [3].

B) Platform as a Service (PaaS): Cloud Platform as a Service (PaaS) facilitates deployment of applications without the cost and complexity of buying and managing the underlying hardware and software layers [2]. PaaS provider provides core cloud competences those are required to develop applications onto the Cloud like operating system, programming language execution environment, web server, database etc**.**

C) Software as a Service (SaaS): It provides the use of applications running on the Cloud Provider's infrastructure. These services can be accessible from any heterogeneous systems or any interfaces [3]. These services may be defined with exception of limited user specific usage.

To gain the maximum benefit from cloud computing, developers must design mechanisms that optimize the use of architectural and deployment paradigms. The role of Virtual Machine's (VMs) has emerged as an important issue because, through virtualization technology, it makes cloud computing infrastructures to be scalable.

Cloud computing enjoys the many attractive attributes of virtualization technology, such as consolidation, isolation, migration and suspend/resume support [4]. A virtual machine (VM) is a software implementation of a computing environment in which an operating system (OS) or program can be installed and run. Important parameters related to virtual machines are Number of virtual machines used by applications, Time taken to create a new VM, Time taken to move an application from one VM to another, Time taken to allocate additional resources to VM Scheduling the basic processing units on a computing environment has always been an important issue. Like any other processing unit, VMs need to be scheduled on the cloud in order to minimize response time, Do the job faster, Consume less energy.

II. VM MANAGEMENT TECHNIQUES

This section describes different virtual machine management technique in cloud environment. Various VM scheduling technique like as:

(A)Migration Techniques

To improve the utilization of cloud resources we use virtual machines. Virtual machine is a software implementation of a computing environment in which operating system or program can be installed and run. In live virtual machine migration, virtual machine moves from one physical host to other host without disturbing others. All migration techniques are trying to reduce total migration time and down time [5].

Manuscript received Nov 06, 2014

Subhash B. Malewar, Dept. of CSE, GHRAET,Nagpur, RTM Nagpur University, India

Prof. Deepak Kapgate, Dept. of CSE, GHRAET,Nagpur, RTM Nagpur University, India

In pre-copy migration first transfer the memory contents to the target machine. After completing the Memory transfer processor states are transferred to destination. In post copy memory data are transferred after the processor states transfer. In order to reduce the total number of pages transfer during the migration.

(B) Real-Time Scheduling

T. Cucinotta[7] proposes a mechanism for providing temporal isolation based on a CPU real time scheduling strategy. This allows not only to have control over the individual virtual machine throughput, but also on the activation latency and response-time by which virtualized software components react to external events. A real system validating the approach by recurring to soft real-time scheduling strategies at the virtualization layer, it is possible to provide a good level of isolation between the concurrently running VMs. Furthermore, it is possible to achieve both a good throughput of the VMs and to keep the individual guarantees at the latency level, something that is not possible with the standard Linux scheduling strategies.

(C) Fault Tolerant Technique

While concerning on large scale system, fault tolerance is a very critical issue, since the cloud resources are placed various locations. This leads to a higher probability of failures while solving huge problems, thus the cloud service reliability could be relatively low. Therefore, providing an effective fault tolerance technique for a cloud system is mandatory.

(D) Memory-aware cloud scheduling techniques

J. Ahn[6] proposes and evaluates two cluster-level virtual machine scheduling techniques for non-uniform memory accesses (NUMA) Affinity and cache sharing which do not require any past or prior knowledge on the behaviors of VMs. For memory-aware scheduling, the cloud scheduler collects the cache behavior of each VM from computing nodes, and migrates VMs if such migration can potentially reduce the overall cache misses and the average memory access latencies by NUMA affinity in the cloud system.

III. VM SCHEDULING ALGORITHMS

In this section describe various virtual machine scheduling algorithms:

A) GREEDY ALGORITHM

The Greedy algorithm is the default algorithm used for scheduling of Virtual Machines. The Greedy algorithm is very simple and straight forward. As a matter of fact, it was the only scheduling policy which was in use for a long time. Only after the cloud started evolving, more complex scheduling policies came into effect [11].The greedy algorithm uses the first node that it finds with suitable resources for running the VM that is to be allocated. The first node that is identified is allocated the VM. This means that the greedy algorithm exhausts a node before it goes on to the next node. As an example, if there are 3 nodes and the first node's usage is 40% while the other two are under loaded and if there are two VMs to be allocated, then both are allocated to the first node which might result in the increase of its usage to 90% while the other two nodes will still remain under loaded. As obviously seen, the main advantage of the Greedy algorithm is its simplicity. It is both simple to implement and also the allocation of VMs do not require any complex processing. The major drawback would be the low utilization of the available resources. As illustrated in the example above, even if there are under loaded nodes, an overloading of a node might result.

B) ROUND ROBIN ALGORITHM

The Round Robin algorithm mainly focuses on distributing the load equally to all the nodes. Using this algorithm, the scheduler allocates one VM to a node in a cyclic manner. The round robin scheduling in the cloud is very similar to the round robin scheduling used in the process scheduling [10]. The scheduler starts with a node and moves on to the next node, after a VM is assigned to that node. This is repeated until all the nodes have been allocated at least one VM and then the scheduler returns to the first node again. Hence, in this case, the scheduler does not wait for the exhaustion of the resources of a node before moving on to the next. The main advantage of this algorithm is that it utilizes all the resources in a balanced order. An equal number of VMs are allocated to all the nodes which ensure fairness. However, the major drawback of using this algorithm is that the power consumption will be high as many nodes will be kept turned on for a long time. If three resources can be run on a single node, all the three nodes will be turned on when Round Robin is used which will consume a significant amount of power [10].

C) WEGHTED ROUND ROBIN ALGORITHM

Shobha Biradar,[8] proposed algorithms weighted round robin, In order to improve the virtual machine's processing in cloud computing infrastructure. The bottom of cloud computing is composed with the virtual machines. When the user catches the large distributed data in cloud computing, the operation will affect the performance of virtual machines. In this a weighted round robin algorithm, which allocates all incoming requests to the available virtual machines in round robin fashion based on the weight's without considering the current load on each virtual machine.

Weighted round robin method improves the performance by consuming less time for scheduling virtual machines.

D) PRIORITY SCHEDULING ALGORITHM

In priority algorithms Virtual Machine is assigned a priority which is allowed to run. Same or Equal-Priority instances are scheduled in First come fist out manners [8]. Priorities are assigned based on characteristics of VM's such as amount of workload, execution time, priority assign by user. Priority once assigned to VM can be changed dynamically by using concept of aging i.e. here the priority of VM keeps increasing based on the total amount of time VM remains in ready queue waiting for execution[8]. If priority of VM increased greater than the VM executing on physical hardware the executing VM preempts with VM having higher priority. It has High processor utilization, High throughput, Minimize turnaround time.

E) DYNAMIC PRIORITY BASED SCHEDULING ALGORITHM

Subramanian S, [10] proposed algorithm uses dynamic priority for the nodes based on which the virtual machines are scheduled. It schedules the VMs to the nodes depending upon

their priority value, which varies dynamically based on their load factor. This dynamic priority concept leads to better utilization of the resources. Priority of a node is assigned depending upon its capacity and the load factor. This algorithm strikes the right balance between performance and power efficiency. It prevents a particular node from being overloaded by considering the load factor. The idle nodes are turned off. Hence it is power efficient.

F) GENETIC ALGORITHM

Genetic algorithm has main objective is to minimize the make span and improve the performance [2]. Initially in Genetic Algorithms many individual solutions are randomly generated to form an initial population. The population size depends on the nature of the problem i.e. no of VM's and type to be run effectively on system. During each successive generation, a proportion of the existing population is selected to breed a new generation. Individual solutions are selected through a fitness-based process, where fitter solutions (VM's schedule likely to give effective response time) are typically more likely to be selected. In the next step is to generation of second population of solutions from those selected through genetic operators: crossover and mutation. This generational process is repeated until a termination condition has been reached i.e. a solution is found that satisfies minimum response time criteria [9].

Genetic Algorithm will increase the cost of time, space, throughput and improve the quality of service of the entire .The goal of GA is to reduce the scheduled time of VM. Genetic algorithm provides both improved response time to VM via parallel execution.

G) CONTENT-BASED SCHEDULING ALGORITHM

Sobir Bazarbayev,[12] proposed content based scheduling algorithm in which VM disk images are stored in specialized storage racks and then transferred to compute nodes in other racks when a VM based on the image is deployed. Different VM disk images share many common pages, especially if they use the same operating system and operating system version. The algorithm returns the selected node and the VM on that node with the highest similar content. When deploying a VM, we search for potential hosts that have VMs that are similar in content to the VM being scheduled. Then, we select the host that has the VM with the highest number of disk blocks that are identical to ones in the VM being scheduled.

Once we have chosen that host node, we calculate the difference between the new VM and the VMs residing at the host; then, we transfer only the difference to the destination host. Finally, at the destination host, we can reconstruct the new VM from the difference that was transferred and the contents of local VMs. Content based VM scheduling algorithm that can significantly reduced the network traffic associated with transfer of VM's from storage racks to host racks in cloud data center.

H) DEADLINE-AWARE ALGORITHM

K. Parrott,[6] proposed deadline-aware scheduler that responds to job execution to optimize job deadlines when run in virtual machines by delays in real time and dynamically optimizes jobs to meet their deadline obligations. The algorithm schedules the jobs and learns over time about the missed deadlines under various conditions and tries to predict whether job would be meeting its deadline and if not then take appropriate measures to improve it chances in meeting deadline .The main goal of the Deadline aware scheduler is to guarantee a start service time for a request. It does that by imposing a deadline on all I/O operations to reducing job deadline miss rate and increasing job throughput rate. Deadline-aware algorithm has flexible use and high utilization of the datacenter resources [13].

IV. COMPARISON OF SCHEDULING ALGORITHMS

Scheduling Algorithms	Response time	Throughput	Fault tolerant	Migration time	Performance
Greedy algorithm	Yes	Yes	-	-	-
Round Robin Algorithms	Yes	Yes	-	-	-
Weighted round robin Algorithm	Yes	Yes	-	-	Yes
Priority Algorithm	Yes	-	-	Yes	Yes
Dynamic Priority Based Scheduling Algorithm	Yes	Yes	-	-	Yes
Genetic Algorithm	Yes	Yes	-	-	Yes
Content Based algorithm	-	-	Yes	-	-
Deadline-Aware Algorithm	-	Yes	-	Yes	-

CONCLUSIONAND FUTURE WORK

In this paper, we study existing management techniques and the algorithms for the virtual machine management in cloud computing .We studied and discuss the advantages and disadvantages of various virtual machines scheduling algorithm. Then we compare studied algorithms based on the matrices. As our future work, we are planning to improve existing virtual machine scheduling algorithms for the better performance, consume less energy and also suitable for the multimedia services and applications.

REFERENCES

[1] Neeraj,MS.Alankrita,Aggarwal, "Load Balance Based on Scheduling and Virtual Machine" IJARCSSE Vol-3, Issue- 7, July -2013.

[2] Karan D.Prajapati,"Comparison of Virtual Machine Scheduling Algorithms in Cloud Computing" IJCA 2013.

[3] MR.Nishant.S.Sanghan,"Pre-Emptable Shortest Job Next Scheduling in Private Cloud Computing JIKRCE Vol -2, Issue -2, Oct -13.

[4] 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.

[5] Getzi Jeba Leelipushpam.P, "Live Virtual Machine Migration Techniques – A Survey" in International Journal of Engineering Research & Technology (IJERT), Vol-1, Sep- 2012.

[6] Jeongseob Ahn," Dynamic Virtual Machine Scheduling in Clouds for Architectural Shared Resources".

[7] Tommaso Cucinotta," Providing Performance Guarantees to Virtual Machines using Real-Time Scheduling" in Tommaso Cucinotta, Dhaval Giani, Dario Faggioli, and Fabio Checconi Scuola Superiore Sant'Anna, Pisa, Italy.

[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, June -2013.

[9] Tarun Goyal, "Host Scheduling Algorithm Using Genetic Algorithm In Cloud Computing Environment", International Journal of Research in Engineering & Technology (IJRET) Vol-1, Issue-1, June-2013.

[10] 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, Nov-2012.

[11] T.Thiruvenkadam, V.Karthikeyani"A Comparative study of VM Placement Algorithms in Cloud Computing Environment"Proceedings of 7th SARC-IRF International Conference, August-2014, New Delhi, India.

[12] Sobir Bazarbayev,"Content-Based Scheduling of Virtual Machines (VMs) in the Cloud" in University of Illinois at Urbana-Champaign, AT&T Labs Research.

[13] K. Parrott," Deadline Aware Virtual Machine Scheduler for Grid and Cloud Computing" in 2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops.