A COMPARATIVE STUDY OF VM PLACEMENT ALGORITHMS IN CLOUD COMPUTING ENVIRONMENT

[1]T.THIRUVENKADAM, [2]V.KARTHIKEYANI

[1]Asst.Professor, Department of Computer Science, K.S.Rangasamy College of Arts and Science, Tiruchengode
[2]Asst.Professor, Department of Computer Science, Thiruvalluvar Govt., Arts College, Rasipuram.
E-mail: mailone.thiru@gmail.com, drvkarthikeyani@gmail.com

Abstract- Cloud computing delivers infrastructure, platform, and software (application) as services. In Infrastructure-as-a-Service (IaaS) service providers generally combine several physical machines and different hardware components to form a single infrastructure. Virtual machine monitors allows a single physical machine to hold many virtual machines hence virtual machines can be employed as computing resources for IaaS. Thus, efficient virtual machine scheduling is necessary in cloud computing environment for increasing resource utilization and efficient use of applications in virtual machine. VM scheduling helps in efficient sharing of virtual machines to available datacenters and these scheduling policies help to increase the cloud performance. Different allocation policies are available and they have their own advantages and limitations. In this paper we propose a comparative study for different types of VM scheduling algorithms and are provisionally discussed and analyzed. Then we concluded that the development of an efficient and enhanced algorithm is needed for scheduling the VMs with the perspective of QOS constraints and enhance user performance of VMs.

Keywords- Cloud Computing, VM allocation, Scheduling, Virtualization , Virtual machine

I. INTRODUCTION

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 consumers. These services in industry are respectively referred to as Infrastructure as a Service (Iaas), Platform as a Service (PaaS), and Software as a Service (SaaS). “Cloud computing, the long-held dream of computing as a utility, has the potential to transform a large part of the IT industry, making software even more attractive as a service”. Clouds [10] aim to power the next generation data centers by architecting them as a network of virtual services (hardware, database, user-interface, application logic) so that users are able to access and deploy applications from anywhere in the world on demand at competitive costs depending on users QoS (Quality of Service) requirements [1].

The following are the major services provided by cloud computing

- Infrastructure as a Service
- Platform as a Service
- Software as a Service

Infrastructure as a Service (IaaS), [2] is a service-oriented model. The users of IaaS can access the virtual hardware resources including virtual machines, virtual networks and virtualized storage through cloud. The most prominent service provider is Amazon EC2.

Platform as a Service (PaaS),[3] is a service platform that developers can use to deploy their own applications. The platform contains a pattern of compute power for hosting some features of the application as well as some of the software such as mail server, web servers, databases, etc. The famous PaaS service providers include Amazon Web Services and Google App Engine. The basic cloud model is Software as a Service (SaaS),[4] which has applications, offered to end users through a web browser or thin client, which is completely stored managed and updated in the cloud. Some of the SaaS service providers are Microsoft’s online update service, Trend Micro Internet Security and so on.

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.

VM allocation is a keyword used in cloud computing for virtual sharing of physical machine among the datacenters. It provides the knowledge of allocated VM to a particular datacenterId. This allocation is based on different policies that make it efficient and easy to understand. These allocation policies can be implemented at virtualization level. The virtualization of cloud elements takes place at the infrastructure layer. Depending on these allocation policies, cloud infrastructure is highly structured and scalable.

II. SERVICE WORKFLOW OF IAAS MODEL

We will use two operation flows to analyze the scheduling flow of cloud platform.[5] The flow of user request to resources is as follow:

- The registered users access the portal server and request the virtual machines with the parameters including quantity of core, frequency, memory, storage space, OS and etc.
- Portal Server sends the request to the scheduling server.
- The scheduling server searches the physical machines to find the host to create the virtual machine according to the metadata of physical machine, which records the operation and configuration details.
- The scheduling server chooses one optimal server and then sends the creating command of virtual machine to its agent server.
- The scheduling sever chooses the virtual machine template form the stored templates within cloud storage administration center, and sends one request for the template to the agent server.
- The requested virtual machine image will be sent (or mapped) to the physical sever based on the template, the scheduling sever will start the virtual machine if the image is loaded successfully.
- If the virtual machine starts successfully, the user can access the virtual machine through RDP, VNC, ICA or SSH.
- The monitor server will send the resource-information renew request, the scheduling server will send the request to the agent servers. If the agent severs acquire the information then send them to the monitoring server.
- The monitoring server will renew the information within the metadata database, to guarantee its correctness, and to improve the efficiency of scheduling operations.

Therefore developing an optimal scheduling algorithm for virtual machines is an important issue. In this paper an analysis of existing Virtual Machine's (VM's) scheduling algorithms have been done.

III. EXISTING SCHEDULING ALGORITHM IN CLOUD COMPUTING

In general scheduling algorithm works for the set of VMs find the appropriate Physical Machine. Scheduling in Eucalyptus determines the method by which Virtual Machines are allocated to the nodes. This is done to balance the load on all the physical machines effectively and to attain a specified quality of service. The demand for a good scheduling algorithm begins from the requirement for it to perform multitasking and multiplexing [10].The scheduling algorithm in cloud environment is anxious mainly with Throughput - number of VMs that are successfully scheduled in the given time unit. Response time - Sum of time taken by the algorithm to place the first VM after a request was submitted. Load Balancing among servers - All the requests for an allocation of a physical machine should be considered in the same manner without any partiality. Resource Utilization –the amount of resource utilized by the VMs scheduled in a physical machine Energy consumption – The total power consumption of a datacenter. Number of active PMs – the total number of physical machines which are currently used for allocating set of VM requests

A. Round Robin Algorithm

The algorithm assigns one VM to a physical machine in a recurring order. The round robin scheduling algorithm used in the IaaS cloud is basically follows the characteristics of the round robin process scheduling in an operating system scheduling. The scheduling algorithm 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. The Round Robin algorithm mainly focuses on dispensing the load equally to all the physical machines. Thus, in this case, the algorithm does not wait for the exhaustion of the resources of a node before moving on to the next.

Round Robin algorithm helps in the Fast Execution and also results into Lower Cost as the VM's are prioritized according to its Cost only. The major benefit of this algorithm is that it consumes all the resources in a balanced order. All the VMs are equally placed in physical machines to guarantee equality.

This algorithm does not consider the overload or under load factor of the host machines. Since it uses more number of physical machines the datacenter utilizes more power.

B. Greedy Algorithm

The open source IaaS Eucalyptus uses Greedy algorithm is its default algorithm for placement of Virtual Machines in physical hosts. The Greedy algorithm is very simple and uncomplicated. This scheduling strategy was the only policy used for a long time. Only after the cloud started growing, more multifaceted scheduling strategies came into effect. The greedy algorithm uses the first physical machine that it come across with appropriate resources for running the VM that is to be allocated and the first VM request is allocated for the identified physical machine. The further VM requests are allocated to the same node till it meets the requirement. This means that the greedy algorithm exhausts a node before it goes on to the next node.

Straightforwardness is the major benefit of the Greedy algorithm. It is easy to employ and also the allocation of VMs do not involved any composite processing. The major drawback would be no fairness among the available resources and not meeting quality of service requirements some times because of consolidating all types of VMs in same physical machine. Minimal consumption of the available resources is the important shortcoming of this algorithm.

C. Backward Speculative Placement

Backward Speculative Placement (BSP) is a placement technique which is based on a simple and fast heuristic producing high quality placements with respect to the optimization goal described in the previous Section. Notice that, due to the unpredictable nature of the deploy stream, the problem cannot be solved using traditional scheduling techniques.

In BSP has two phases for cloud placement optimization: Continuous deployment: Cloud infrastructures continuously receive an unpredictable stream of deploy requests over time. This kind of behavior demands for a (i) quick placement of newly arrived deploy requests, (ii) limited number of time expensive operations (i.e., computation of new placement, relocations), and (iii) smart placement decision so that future state of the infrastructure remains stable and efficient. Due to the high responsivity requirement, the continuous deployment phase does not allow relocations of VMs (i.e., all existing VMs are locked in their host, except those with expired lifetime). The placement decisions deal only with the mapping of newly arrived VMs to hosts.

Ongoing optimization: Long periods using a continuous deployment setting might bring the infrastructure to a suboptimal state which is far from the optimization goal for this reason, BSP periodically reoptimize the placement also considering migration of VMs.

By monitoring historical demand traces of deployed VMs, BSP projects the past demand behavior of a VM to a candidate target host, and capture the VMs correlation aspect in an efficient way Besides, authors also restrict the maximum number of migration. The optimization component is defined such that the deference between the most and the least loaded hosts is minimized. However, the resource that BSP is considering is only focused on CPU

D. Dynamic Priority Based Scheduling Algorithm

It prevents a particular node from being overloaded by considering the load factor. The idle nodes are turned off. Hence it is power efficient. It prevents fluctuations around the load factor of 80% in most cases. Fluctuation occurs only under extreme cases, when all the nodes have load factor which are approximately 80%.

In Dynamic Priority algorithm is discussed. Mainly this scheduling a virtual machine in Eucalyptus platform and it will work under various circumstances. But this algorithm does not handle certain cases because of failure of nodes. Also the uptime and downtime of nodes have not been measured.

E. Genetic Algorithm

In the genetic algorithm, the resources are deployed and are arranged to every physical node. By this way it solves the problems. The genetic algorithm introduces an average load distance in order to measure the overall load balancing effect of the algorithm. Virtual machine (VM) migration is used to avoid the conflicts on traditional systems like CPU and memory, micro-architectural resources such as shared caches, memory controllers, and non uniform memory access (NUMA).

In Genetic Algorithm is discussed, in genetic algorithm the problem is the load balancing. So, the strategy for scheduling the VM resources on load balancing is based on the genetic algorithm. According to historical data and the current state of the system through the genetic algorithm, this scheduling strategy computes the needed VM resources after the deployment and chooses the least-affective solution through which it achieves the best load balancing and avoids or reduces the dynamic migration.

The goal of this algorithm is to prioritize resources those are most suitable for the VM. Those resources with a higher rank are used first to allocate VMs.

In Architectural Shared Resources it shows the live VM migration which is used to mitigate the contentions on micro-architecture resources. This reduces conflicts. It shows the evaluation of two-cluster level virtual machine scheduling techniques for cache sharing and it does not require any prior knowledge on the behaviors of VMs.

In Broker Virtual Machine Communication Framework they have proposed an efficient algorithm to provide an effective and fast execution of the task assigned by the user. So there is an effective communication framework between broker and virtual machine for assigning the task and fetching the results in optimum time and cost using Broker Virtual Machine Communication Framework (BVCF).

F. Power Save Algorithm

The Power Save algorithm optimizes the power utilization by switching off the physical machines which are not presently used by the VMs.[6] Instead of keeping all the physical machines turned on, resulting in a lot of power utilization, this algorithm aims at turning the unallocated physical machines off which will decrease the power utilization to a reasonable extent. The scheduler allocates a VM to the node and then traverses through the list of nodes to check if the node is unused and if found to be so, turns it off.[6] [10] If the resource of a node which has been turned off is required for the allocation of a VM, the scheduler turns it on again and then allocates the VM to that node.

As an example, consider the scenario in which there are three nodes, two of which are unused. When a new VM is to be scheduled, the scheduler may allocate it to the node which is already being used and would turn off the two nodes which are unused. The Power Save algorithm results in the reduction of power consumption but this is at the expense of lower utilization of resources. This algorithm is used only at places where there is an extreme need for reducing the consumption of power.

IV. COMPARISION OF EXISTING ALGORITHMS

Determination of best scheduling algorithm for cloud computing is depends on various factors. For scheduling and provisioning of resources different algorithms are available that are aware of particular factor. We have compared such algorithms with factors like Least Response Time, Load Balancing among servers, Reasonable Resource Utilization, Least energy consumption, Minimum number of active PMs and Higher profit. The below table list the features available in some of the most popular VM scheduling algorithm used in cloud datacenter environment.

Table 1: Comparison of Scheduling Algorithms

Algorithm	Least Response Time	Load Balancing among servers	Reasonable Resource Utilization	Least energy consumption	Minimize the no of active PMs	Higher profit
Round Robin	Yes	Yes	-	-	-	-
Greedy	Yes	-	Yes	-	Yes	Yes
Genetic	-		Yes	Yes		Yes
Dynamic Priority	-	Yes	-	Yes	Yes	-
BSP	-	Yes	Yes	-	-	-
Power Save	-	-	Yes	Yes	Yes	-

CONCLUSION AND FUTURE WORK

various scheduling algorithms with the perspective of Least Response Time, Load Balancing among servers, Reasonable Resource Utilization, Least energy consumption, Minimum number of active PMs and Higher profit are studied. To solve the resource scheduling problem various scheduling algorithms based on various factors have been tried by various researchers. The majority of the previous studies have ignored the dynamic nature of the incoming stream of VM deployment requests to which the cloud infrastructure is subject over time Moreover, different algorithms woks well and satisfies some of the above discussed factors but the cloud providers need a dynamic algorithm which will perform well in all situations and satisfies all the factors so as future work we plan to develop such a dynamic algorithm which will consider various factors for VM placement to avoid SLA violations and manages QOS constraints.

REFERENCES

[1] William Voorsluys, James Broberg, and Rajkumar Buyya. Cloud Computing: Principles and Paradigms, chapter Introduction to Cloud Computing, pages 1–44. Wiley Press, 2011.

[2] K. Chen and W. M. Zheng, “Cloud Computing: System instance and Current State,” Journal of Software, Vol. 20, No. 5, 2009, pp. 1337-1348. doi:10.3724/SP.J.1001.2009

[3] Z. W. Xu, H. M. Liao, et al., “The Classification Reaserch of Network Computing System,” Journal of Computing Machine, Vol. 18, No. 9, 2008, pp. 1509-1515.

[4] G. W. Zhang, R. He and Y. Liu, “The Evolution Based on Cloud Model,” Journal of Computing Machine, Vol. 7, 2008, pp. 1233-1239

[5] Aobing Sun,Tongkai Ji, Qiang Yue, Feiya Xiong “IaaS Public Cloud Computing Platform Scheduling Model and Optimization Analysis” Int. J. Communications, Network and System Sciences, 2011, 4, 803-811

[6] Subramanian S, Nitish Krishna G, Kiran Kumar M,Sreesh P4and G R Karpagam, “An Adaptive Algorithm For Dynamic Priority Based Virtual Machine Scheduling In Cloud” IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 6, No 2,November 2012.

[7] Jianhua Gu, Jinhua Hu, Tianhai Zhao, Guofei Sun, “A New Resource Scheduling Strategy Based on Genetic Algorithm in Cloud Computing Environment”, Journal Of Computers, Vol. 7, No. 1, January 2012.

[8] N. M. Calcavecchia “VM Placement Stratrgies for Cloud Scenarios” in Proc IEEE 5th International Coference on Cloud Computing.

[9] H. Viswanathan, “Energy-Aware Application-Centric VM Allocation for HPC Workloads”

[10] Supreeth S 1, Shobha Biradar “Scheduling Virtual Machines for Load balancing in Cloud Computing Platform”, International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064

★★★