International Journal of Latest Trends in Engineering and Technology
Special Issue SACAIM 2017, pp. 344-348
e-ISSN:2278-621X

A STUDY ON LOAD BALANCING AND CONSOLIDATION TECHNIQUES FOR GREEN CLOUD COMPUTING

Sreenivasa B.L[1], Ransom Faria[2], Anzinto Fernandes[3] & Dr. S Sathyanarayana[4]

Abstract- Cloud computing is a computing based technology that uses the Internet and central remote services to manage data and application. Wide usage of Cloud concepts has increased number of data centers over the world resulting into significant increase in power consumption by datacenters which affects environment. With the help of virtualization, multiple virtual machines (VM) can be deployed and combined into a physical machine (PM). These VMs hold and implements the Cloud workload. Efficient use or allocation of resources can save energy consumption. Virtualization is an important technology embedded in Cloud Computing. Mapping the virtual machines to the appropriate physical machines is called VM placement. The effectiveness and elasticity of virtual machine placement has become the main issue in cloud computing environments. In this paper we do study on load balancing and consolidation techniques in order to understand how these algorithm can manage the load and power-on physical machine.
Keywords– Cloud computing, Data centers, virtual machines.

1. INTRODUCTION

As cloud computing provides various services to users across the world [1], applications from customers which are hosted by very large-scale data centers. In such platforms, virtualization technology is engaged to multiply the underlying physical resources. It provides away to deliver the platform, infrastructure, and software as services accessible to consumers in a pay-as-you-go manner. Some of the commercial service providers include Amazon, Google, and Microsoft
cloud computing environments, large-scale data centers are usually the crucial computing infrastructure, which it consist of different number of physical nodes with different virtual machines running upon them. Cloud computing is a new example for the dynamic provisioning of computing services supported by modern data centers. With the development of cloud computing, the data centers are growing at a faster rate. However, average data centers require as much energy as 25,000 household and it was predictable that the energy consumption from data centers would have accounted for two percent of the worldwide energy consumption by 2020.

Heuristics for dynamic reallocation of VMs according to recent resources requirements, while ensuring reliable Quality of Service. The aim of the reallocation is to minimize the number of physical nodes serving existing workload, whereas nodes which are not used are switched off to lower the power consumption. Many researches have been done in power efficient resource managing in data centers. Virtualization is the important technology in cloud computing. Virtualization is the powerful tool used in today's computing to get best results. Over virtualization multiple operating system can made running on one physical machine. With the help of virtualization number of VM can be formed on the server with multiple OS running on it and on single OS one application can run well. OS running on other virtual machines running on same host will not be affected and work continuously, even if a virtual machine not working. Virtual machine is created for each user, with the help of hypervisor. Many numbers of virtual machine are hosted in a single host and One or many VM are created to one user.

The un-required servers are switched off in order to lower the number of running servers, thus reducing the energy use at the points of consumption to offer benefits to all other levels. Also, several additional servers are added to help guarantee service-level agreement. To reduce the total cost of ownership while ensuring quality of service and the lowermost line is to protect the environment. The rest of the paper is organized as follows. Section II a study on load balancing and consolidation algorithm. Section III conclusion and future work.

2. LOAD BALANCING AND CONSOLIDATION TECHNIQUES

2.1 ETC

In this paper Ching-Hsien Hsu et al., an energy-aware task consolidation (ETC) technique is presented aims to optimize energy consumption of virtual clusters in cloud data center. Conforming most cloud systems, a 70% principle of CPU utilization is proposed to manage task consolidation among virtual clusters. The ETC algorithm significantly reduce power consumption in managing task consolidation for cloud systems [2].

[1] Assistant Professor, Department of Information Technology, Aloysius Institute of Management and Information Technology Mangalore, Karnataka, India
[2] Department of Master of Computer Application, St. Aloysius College, Mangalore, Karnataka, India
[3] Department of Master of Computer Application, St. Aloysius College, Mangalore, Karnataka, India
[4] Asst Professor, First Grade Women's College, Mysore, Karnataka

The main idea of our Energy aware Task Consolidation method (ETC) is to consolidate tasks and keeps the total CPU utilization of each Vi being under 70% if applicable. Given a cloud system that is composed with multiple virtual clusters (VC), for example, three VCs A, B and C (denoted by VCA, VCB and VCC), the task consolidation strategy within a VC can be described as follows. - Scheduler of VCA dispatches task tj to its virtual machines if any of them conforms with the 70% principle. If more than one virtual machines can be chose, the best fit strategy will be applied. - If there is no such Vi conforms with the 70% principle, VCA asks resource support from other VCs, e.g. VCB or VCC. - If both VCB and VCC can provide VMs that conforms the 70% principle, then select the lower one of energy consumption in transmitting and executing the task. - If no resource (virtual machine) in other VCs conforms with the 70% principle, assigning tj to Vi with less energy consumption in VCA. Let's use few examples to clarify the above idea. Figure 1 outlines the basic information of a set of tasks. To simplify the presentation, we assume there are three virtual machines in VCA. The first scenario illustrates that a task is assigned to virtual machines locally in VCA that conforms the 70% principle. As shown in Figure 4, after five tasks being assigned to all virtual machines, t5 could be assigned to both V0 and V1. Applying the best fit strategy, task t5 is assigned to V1 because the total CPU utilization is closest to 70% [2].

Task t_j	Arival Time ($T_{a,j}$)	Processing Time ($T_{p,j}$)	CPU Utilization	Data Size
t_0	0s	50s	30%	150Mb
t_1	10s	20s	30%	75Mb
t_2	12s	35s	40%	20Mb
t_3	15s	15s	30%	150Mb
t_4	20s	30s	60%	250Mb
t_5	30s	25s	30%	110Mb
t_6	35s	10s	50%	210Mb

Figure 1

To dispatch and execute task t6, because t6 will consumes 50% CPU utilization, assigning it to any VM in VCA will break the 70% principle. To conform the 70% principle, cluster A asks for resource support from other clusters, for example, VCB and VCC. As shown in Figure 3, both VCB and VCC have available resources conforms the 70% principle, thus VCA needs to decide the destination cluster that reflects better energy usage. To estimate the energy consumption in consolidating a task to other resources that located in different clusters, both CPU computation and network transmission should be considered [2].

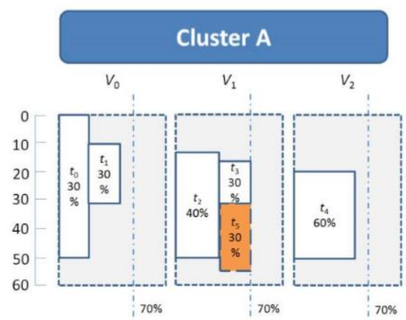

Figure 2

It's easy to see that VCC presents better energy consumption. Thus task t6 will be consolidated to virtual machine in VCC. Figure 2 shows the above scenario [2].

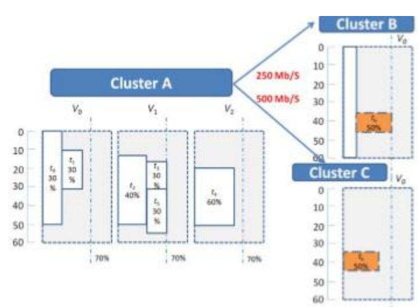

Figure 3

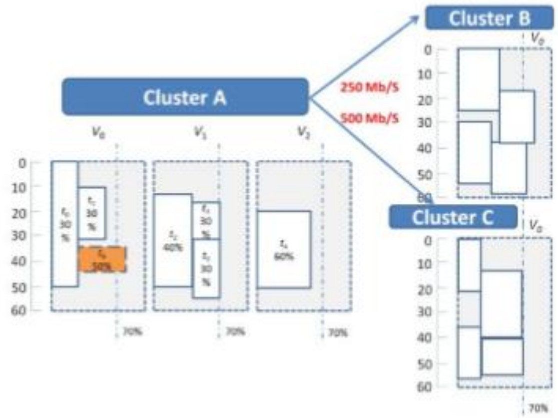

Figure 4

The last example shows what if VCB and VCC do not have enough resources while VCA asks support for t6. In Figure 4, CPU utilization of VMs in VCB and VCC are higher than or equal to 70%. Thus VCA will not seek outside resources due to extra overheads might be incurred. As a result, it assigns t6 to local resources even the VM could not conform the 70% principle. In this example, V0 is chose [2].

An energy-aware task consolidation (ETC) technique minimizes the energy consumption. Considering the architecture most cloud systems conforms, a 70% principle is proposed to manage task consolidation among virtual clusters. Idle state of virtual machines and network transmission are assumed to be 7 times and 2 times of basic energy consumption unit in this study. Note that these values could be adjusted in the ETC method to be adaptive in different cloud systems. ETC can significantly reduce power consumption in managing task consolidation for cloud systems [2].

2.2 NBST

Cloud Computing is one of the fastest growing technologies in present day. Cloud is the metaphor used for internet and cloud computing is sharing pool of resources over internet. Count of users switching to cloud services is increasing day by day, which in turn requires better performance of cloud. One of the major challenges faced in cloud is efficient distribution of resources for serving millions of user requests in less time. Distribution of workload across multiple computing resources is known as Load balancing, its main purpose is to achieve high resource utilization and user satisfaction. Therefore, different load balancing algorithms are used in cloud in order to serve the requests from user in efficient manner. In this paper Shubham Sidana et al., an NBST algorithm is used to balance load on cloud based on arrangement of resources, according to processing speed for virtual machines and then allocating cloudlets to the resources according to their processing requirement. The main aim of NBST is to find the least loaded node and assign application to it, thereby equally distributing the node in the network [3].

Since load balancing is a major issue in Cloud Computing and utilization of resources is one of the important aspects of Cloud, an algorithm is proposed where the load balancing is done among all the virtual machines and load is distributed by sorting these virtual machines on the basis of their processing speed and requests on the basis of their length i.e., number of instructions [3].

In NBST algorithm, jobs are present in the queue and length is known i.e., number of instructions in request. The load balancing algorithm aims at reducing the load over resources. For achieving this, we have arranged all the virtual machines in order according to their execution speed that is in MIPS (Million instructions per second). After arrangement of machines, sorting of cloudlets is performed on the basis of their length (million instructions). Mid-point is taken of those sorted cloudlets list and sorted virtual machines list and then the divided cloudlet lists are mapped to the corresponding lists of virtual machines [3].

In NBST algorithm, tries to balance the load by arranging the virtual machines on the basis of their processing power and arranging the cloudlets according to their length i.e. number of instructions in the cloudlet. The algorithm allocates the resource in such a manner that job requiring less processing are not allocated to the machines with high processing power [3].

2.3 Enhanced Active Monitoring Load Balancing (EAMLB)

Cloud computing provides different information technology services as a commodity to its users. When it was introduced, business industries were using large-scale main frames and they are increasing in numbers and sizes. All are aims to increase its number of users to get more revenue. Hence, they need different types of services having infrastructure, platform, software and many more in less response time. Two types of policies exist in cloud computing systems namely static load balancing and dynamic load balancing. Static load balancing policies do not consider the current state of system. Dynamic load balancing depends on the present behavior of the system. There are various dynamic load balancing policies for virtual machines already existing in this context including Throttled and Active Monitoring in cloud computing systems. These types of algorithms are more flexible in nature. Enhanced Active Monitoring Load Balancing (EAMLB) algorithm is designed to minimize the response time in cloud systems. EAMLB policy provides better response time than Active Monitoring policy that improves the performance of cloud systems [4].

Response time is one of the important factors in case of performance of cloud computing. Minimum response time increases the quality of service to the users and they can better utilize the services provided by the cloud. Enhanced Active Monitoring Load Balancing (EAMLB) algorithm is designed by keeping this parameter in mind. It enhances the concept of already existing Active Monitoring Load Balancing (AMLB) algorithm [4].

An algorithm is developed by enhancing the concept of Active Monitoring load balancing (AMLB) algorithm, which finds the least loaded virtual machine among all virtual machine. In Enhanced Active Monitoring Load Balancing (EAMLB) algorithm, Enhanced Active VmLoad Balancer also considers the virtual machine which is allocated recently along with the least loaded. It gives the benefit that one VM will not be allocated in continuous manner if it is least loaded. One VM needs not to process the task again and again. This minimize the response time and increases the performance of cloud systems [4].

Inputs are given as different jobs jl, j2, ... , jn submitted by the user and the virtual machines vl,v2, ... ,vn created on data centers. In this, two hash maps are created. First is vrnStateList that manage the total number of available and non allocated VMs and second is current Allocation Counts that holds the VMs with their allocation counts. When Data Center Controller (DCC) receives a new request then it queries to the load balancer for new allocation [4].

The load balancer first checks in vrnStateList either any VM is available or not. If VM is available then allocate the job to it, otherwise check the least loaded VM which is not recently loaded. In this case load balancer moves to the next least loaded VM and allocate the request on it. Different variables are used in this algorithm as vmId that refers the Vm which has to be allocated, temp VmId that stores the recent allocated vrnId, minCount that refers the minimum value of allocation that any VM can have, currCount that stores the current number of allocation for virtual machine which has to be checked for leas load [4].

Enhanced Active Monitoring Load Balancing (EAMLB) algorithm is developed by enhancing the working of Active Monitoring algorithm. It improves the response time better than Round Robin and Active Monitoring [4].

2.4 Reinforcement Learning---based Dynamic Consolidation (RL-DC)

Dynamic consolidation techniques optimize resource utilization and reduce energy consumption in Cloud data centers. They should consider the variability of the workload to decide when idle or underutilized hosts switch to sleep mode in order to minimize energy consumption. In this paper Fahimeh Farahnakian et al., Reinforcement Learning-based Dynamic Consolidation method (RL-DC) is to minimize the number of active hosts according to the current resources requirement. The RL-DC utilizes an agent to learn the optimal policy for determining the host power mode by using a popular reinforcement learning method. The author says that the host should be switched to the sleep or active mode and improves itself as the workload changes. Therefore, RL-DC does not require any prior information about workload and it dynamically adapts to the environment to achieve online energy and performance management.

Reinforcement learning (RL) is a machine learning paradigm that has been applied for energy management in large-scale systems. In RL, a decision-maker or agent percepts the environment and chooses an action at each state. After each action execution, the agent receives a feedback indicating the quality of the applied action. The final goal of the agent is to learn a policy for selecting the best action among all possible actions.

In this paper, a dynamic consolidation method is proposed in order to reduce energy cost and SLA violation of data center and named Reinforcement Learning---based Dynamic Consolidation (RL-DC) algorithm. RL-DC can dynamically adapt a number of active hosts to the variable workload. An important part of the consolidation algorithm, is to decide whether (1) additional host are required to provide efficient resource utilization with an increasing workload, or (2) redundant hosts can be put sleep to save energy or (3) the current amount of hosts is sufficient. To make this decision, a learning agent is assumed as an essential part of RL-DC.

The agent first percepts the information about the current power consumption, total CPU utilization and power mode of hosts at beginning a time slot. The time between two iteration of the consolidation algorithm is called the time slot. Then, the host power mode (active or sleep) in the next time slot based on this information and its experience of previous host state is determined by the agent. The RL-DC algorithm optimizes the resource allocation according to the specified power mode of hosts. While the learning agent decides a host should be switched to the sleep, RL-DC migrate all VMs from the host to other hosts. The VM allocation algorithm selects a host to allocate VM from the host that must be switched to the sleep mode. Therefore, the energy cost and CO_2 emissions can be reduced in a data center by switching the under-loaded hosts to the sleep mode. Moreover, when the host power mode is decided to be active and the current mode is not the active mode; the host will be switched to the active mode. RL-DC employs a prediction method, LiRCUP, to avoid the SLA violation. Based on the past CPU utilization values in a host, LiRCUP approximates a function based on the linear regression. The function can forecast the short-term utilization of host by considering on the historical data of usage. If the predicted usage exceeds of available host usage, the host becomes over-loaded. So some VMs on the host must migrate to other hosts before a SLA violation happen (while loop). The VM selection algorithm selects which VM should be migrating to other hosts. The selected VM reallocate to the host that is chosen with the VM allocation algorithm.

A dynamic consolidation method reduces power consumption and SLA violation in the cloud data centers. It employs the reinforcement learning approach to learn the host power mode detection policy without prior knowledge of the environment and workload. Therefore, the method can adapt the number of active hosts to the current resources requirements. reinforcement learning-based dynamic consolidation method is able to minimize energy cost and SLA violation rate efficiently.

3. CONCLUSION

Virtual machine migration acts as an important tool for dynamic resource management in modern day data centers. Load Balancing and Consolidation is a strong technique, which speed up the concurrent execution of number of tasks in virtual machines and reduce the energy usage. Consolidation in cloud computing provides a significant proposal for energy usage. In this paper we studied on load balancing and consolidation techniques in order to understand how these algorithm can manage the load and power-on physical machine. In future we practically compare the algorithms to check the efficiency in power and load in different environment setup.

4. REFERENCES

[1] B. Corona, M. Nakano, H. Pérez, "Adaptive Watermarking Algorithm for Binary Image Watermarks", Lecture Notes in Computer Science, Springer, pp. 207-215, 2004.

[2] Ching-Hsien Hsu, Shih-Chang Chen, Chih-Chun Lee, Hsi-Ya Chang, Kuan-Chou Lai, Kuan-Ching Li and Chunming Rong, "Energy-Aware Task Consolidation Technique for Cloud Computing.".

[3] Mr. Shubham Sidana, Ms. Neha Tiwari, Mr. Anurag Gupta, Mr. Inall Singh Kushwaha,"NBST Algorithm: A load balancing algorithm in cloud computing".

[4] Shikha Garg, Dr. D.V. Gupta and Dr. Rakesh Kumar Dwivedi,"Enhanced Active Monitoring Load Balancing Algorithm for Virtual Machines in Cloud Computing"

[5] Fahimeh Farahnakian, Pasi Liljeberg, and Juha Plosila, "Energy-Efficient Virtual Machines Consolidation in Cloud Data Centers using Reinforcement Learning"

[6] International Journal of Computer Applications (0975 – 8887) Volume 138 – No.11, March 2016 25 A Review on Secure and Energy Efficient Approaches for Green Computing Shivangi Sharma Department of Computer Science and Engineering JMIT, RadaurYamunanagar, India.

[7] International Journal of Computer Applications (0975 – 8887) Volume 138 – No.11, March 2016 25 A Review on Secure and Energy Efficient Approaches for Green Computing Gaurav Sharma, PhD Department of Computer Science and Engineering JMIT, Radaur Yamunanagar, India.

[8] International Journal of Scientific & Engineering Research, Volume 6, Issue 12, December-2015 313 ISSN 2229-5518 IJSER © 2015 https://www.ijser.org Resource Allocation and Server Consolidation Algorithms for Green Computing Mustaffa-Sami, M. Haggag, Dina Saleem.

[9] Arshjot Kaur et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 4886-4890A Survey of Resource Scheduling Algorithms inGreen ComputingArshjot Kaur, Supriya KingerDepartment of Computer Science and Engineering, SGGSWU,Fatehgarh Sahib, Punjab,India (140406).

[10] An Energy Efficient Ant Colony System forVirtual Machine Placement in Cloud Computing by Xiao-Fang Liu, Student Member, IEEE, Zhi-Hui Zhan, Member, IEEE, Jeremiah D. Deng, Member,IEEE, Yun Li, Member, IEEE, Tianlong Gu, and Jun Zhang, Senior Member, IEEE.

[11] Thanu-shree, Dr.Neelender Badal, "Energy Efficient Load balancing Algorithm for Green Cloud", IJERT Vol.5 Innue 03, March-2016

[12] T.Jenifer Nirbah, Rose Rani John "Energy efficient Task Scheduling Algorithms For Cloud Data Centers" IJERT eISSN:2319-1163

[13] Truong Vinh Troung Duy, Yukinori Sato and Yasushi Inoguch "Performance Evaluation of a green scheduling algorithm for Energy savings in cloud" 1-1 Asahidai, Nomi, Ishikawa, 923-1292 Japan [yukinori, inoguchi]@jaist.ac.jp, 978-1-4244-6534-7/10/$26.00 ©2010 IEEE.

[14] Andrew J.Younge, Gregor von Laszewski, Lizhe-Wang ,Sonia Lopez-Alarcon, Warren-Carithers "Efficient Resource Management for Cloud Computing Environments".

[15] Jayant Adhikari, Sulabha Patil, "Comparison of Energy Aware Load Balancing Algorithms in Cloud Computing", International Journal of Scientific & Engineering Research, Volume 4, Issue 12, December-2013 364,ISSN 2229-5518