International Journal of Latest Trends in Engineering and Technology
Special Issue SACAIM 2016, pp. 369-374
e-ISSN:2278-621X

DESIGN AND IMPLEMENTATION OF NEW VIRTUAL MACHINE SCHEDULING ALGORITHM

Neenu Joseph[1], Melody M Fernandes[2], Priyanka R Kamble[3] and Mr. Santhosh B[4]

Abstract- For efficient operations in a distributed environment load balancing is an important factor. Load Balancing strategies are of huge significance in improving the reliability and performance of resources in data centres. One of the challenging problems in scheduling resources in a cloud data centre is the allocating and migrating of virtual machines, these virtual machines are reconfigurable. Therefore load balancing due to its challenges and importance for the cloud has become a major research area. Various algorithms were proposed to supply productive mechanisms for tending to the client's requests to available Cloud nodes. In this paper, we investigate the different multidimensional resource scheduling algorithms used to achieve scheduling of various task and balancing load in Cloud Computing.

Keywords –scheduling, integrated load balancing, imbalance value, cloud computing

I. INTRODUCTION

Cloud computing is one of the latest upcoming technologies, that is internet based with various advancement in virtualization, grid, web computing, utility computing and other related technologies. Many organizations use cloud computing for various purposes. Cloud computing allows us perform tasks with high quality, low cost, and uses pay-per-use model. It provides platform and applications on demand though the internet. The benefits of could computing are hiding and abstraction of complexity, virtualization of resources, and efficient use of distributed resources. Examples of could computing platforms are: Google App Engine,IBM blue cloud,Amazon EC2,Microsoft Azure.

A cloud datacenter is a distributed network structure that has many nodes. Resources like CPU, memory, and network bandwidth form a node. These different resources are the multi-dimensional resources. The traditional data centers contained physical machines. This configuration was expensive, with wasted energy and floor space, low resource utilization, and management overhead. The concept of virtual cloud data centers have the ability to be allocated form one set of resources to another in a proper manner. This allows us to use cloud more efficiently and usefully. We can handle large data centers with reduced maintenance cost. To use the cloud resources more efficiently scheduling of the resources plays an important role.

In this paper we have presented a modified version of the current LIF scheduling algorithm. This increases the efficiency of operations in a distributed environment.

A. Challenges in Load Balancing–

Modified resource allocation techniques and better improved strategies through efficient job scheduling are the main provisions of the load balancer. The load can be network load, CPU load,

[1] *Aloysius Institute of Management and Information Technology St.Aloysius College(AUTONOMOUS) BeeriKotekar, Mangalore, Karnataka, India.*
[2] *Aloysius Institute of Management and Information Technology St.Aloysius College(AUTONOMOUS) BeeriKotekar, Mangalore, Karnataka, India.*
[3] *Aloysius Institute of Management and Information Technology St.Aloysius College(AUTONOMOUS) BeeriKotekar, Mangalore, Karnataka, India.*
[4] *Aloysius Institute of Management and Information Technology St.Aloysius College(AUTONOMOUS) BeeriKotekar, Mangalore, Karnataka, India.*

memory capacity or delay. Load balancer performs the distribution of load among various nodes in a distributed system while also being able to avoid a situation where some of the nodes are idle or less loaded while others are overloaded and to improve both utilization of resources and response time of jobs. At any instant of time the load balancer ensures that equal amount of work is done by every node in the network or all the processors in the system. This is the important factor to be considered during the resource allocation but this has become more difficult especially in elastic cloud computing where the user can dynamically request for the resource.

The performance prediction also plays an essential role in load balancing. But cloud environment is highly variable and unpredictable. Providers try to oversubscribe as many users to a shared infrastructure to increase resource utilization. This results in resource contention and interference. Heterogeneity within the same instance type and administrative action (e.g., eviction) to maintain the service level are some other factors that cause unpredictability of the environment.

II.PROBLEM STATEMENT

At a certain time period in a cloud data center there are *M* Physical Machines (PMs), also called hosts, which configuration may be heterogeneous. Allocating and migrating virtual machines (VMs) which are reconfigurable and taking into consideration integrated features of hosting physical machine (PMs) are one of the challenging problems in scheduling resource in cloud data centres. This problem can also be defined as given a set of*n* requests (VMs) and a set of *m* identical machines (PMs) *PM*1, *PM*2, ...,*PMm*, each request has a processing time, the objective of load-balance is to balance load on every machine while they are being assigned requests.

III. EXISTING ALGORITHMS

A. LIF Algorithm–

WenhongTian et al [1]. proposedan algorithm based on demands characteristics (for example, CPU intensive, high memory, high bandwidth requirements etc.), always selects lowest integrated imbalance value in different physical machines(as stated in equation (2-3)) and available resource to allocate virtual machines.

LIF algorithm considers imbalance values integrated CPU, memory and network bandwidth utilization, and the following parameters are considered: Average CPU utilization $CPU_i^U$of a single server *i*: is averaged CPU utilization during observed period. For example, if the observed period is one minute and utilization of CPU is recorded every 10 seconds, then $CPU_i^U$is the average of six recorded values of server *i*.

Average utilization of all CPUs in a Cloud data center. Let CPU_i^nbe the total number of CPUs of server *i*,

$$CPU_u^A = \frac{\sum_{i=1}^{N} CPU_i^U CPU_i^n}{\sum_{i=1}^{N} CPU_i^U} \quad (1)$$

Where*N* is the total number of physical servers in a Cloud data center and CPU_i^nrepresents the number of CPUs in physical server *i*. Similarly, average utilization of memory, network bandwidth of server *i*, all memories and all network bandwidth in a Cloud data center can be defined as MEM_i^u,NET_i^u, MEM_u^A, NET_u^Arespectively.

Datacenter-wide integrated imbalance value *ILBi*, of server *i*. In statistics variance is used as a measurement of how far a set of numbers are spread out from each other, therefore it is widely used. Using variance, an integrated load imbalance value (*ILBi*) of server *i*is defined

$$\frac{(Avg_i - CPU_u^A)^2 + (Avg_i - MEM_u^A)^2 + (Avg_i - NET_u^A)^2}{3} \quad (2)$$

Where

$$Avg_i = \frac{(CPU_i^u + MEM_i^u + NET_i^u)}{3} \quad (3)$$

is average utilization of multi-dimensional resource in a physical machine, also called integrated load in LIF.

B. ZHCJ Algorithm–

Wood et al. [2], introduced a few virtual machine migration techniques. One integrated load balance measurement is applied as follows:

$$V = \frac{1}{(1-\text{CPU}_u)(1-\text{MEM}_u)(1-\text{NET}_u)} \quad (4)$$

Wherecpu, net and mem are the corresponding utilizations of that resource for the virtual or physical server. The higher the utilization of a resource, the greater the volume; if multiple resources are heavily utilized, theabove product results in a correspondingly higher volume. The volume captures the degree of (over)loading along multiple dimensions in a unified fashion and can be used by the mitigation algorithms to handle all resource hotspots in an identical manner. The algorithm always chooses physical machines with lowest referred V value and available resource to allocate virtual machines.

C. ZHJZ Algorithm –

Zheng et al [3].proposed integrated load-balancing measurement as following:

$$B = \frac{aN1_iC_i}{N1_mC_m} + \frac{bN2_iM_i}{N2_mM_m} + \frac{cN3_iD_i}{N3_mD_m} + \frac{dN4_iNet_i}{Net_m} \quad (5)$$

The referred physical server *m* is selected firstly. Then other physical servers *i*is compared to server *m*. *N*1*i* is the CPU capability, *N*2*i* is for memory capability, *N*3*i* is for hard disk. *Ci*, *Mi*is for average utilization of CPU and memory respectively, *Di* is for transferring rate of hard disk, *Neti*is for network throughput. *a, b, c, d* is for weighting factor of memory, network bandwidth, CPU and harddisk respectively.
The algorithm selects a physical machine, and calculates the value and chooses lowest referred B value in different physical machines and available resource to allocate virtual machines.

D. D. Rand Algorithm –

Randomly assigns requests (virtual machines) to physical machines which have available resource.

E. E. Round Robin (RR) –

The round-robin is one of the most used algorithm for scheduling (for example by Amazon EC2 and Eucalyptus [4]), in which PM’s are allocated VM’s in turns. Simplicity in implementation is the advantage of this algorithm.

IV. PROPOSED ALGORITHM:Extended_LIF

LIF algorithm considers imbalance values integrated CPU, memory and network bandwidthutilization. It always selects lowest integrated imbalance value in different physical machines and available resource to allocate virtual machines.

Extended_LIF algorithm considers hard disk utilization along with CPU, memory andnetwork bandwidthutilization. The following parameters are considered:Average CPU utilization CPU_i^U of a single server *i*: is averaged CPU utilization during observed period. For example, if the observed period is one minute and utilization of CPU is recorded every 10 seconds, then CPU_i^U is the average of six recorded values of server *i*.Average utilization of all CPUs in a Cloud data center. Let CPU_i^n be the total number of CPUs of server *i*,

$$CPU_u^A = \frac{\sum_{i=1}^{N} CPU_i^U CPU_i^n}{\sum_{i=1}^{N} CPU_i^U} \quad (6)$$

Where$N$ is the total number of physical servers in a Cloud data center and CPU_i^n represents the number of CPUs in physical server *i*. Similarly, average utilization of memory, network bandwidth and hard disk of server *i*, all memories, all network bandwidth and all hard disks in a Cloud data center can be defined as MEM_i^u,NET_i^u, MEM_u^A, NET_u^A, HD_u^A, HD_u^A respectively.

Datacenter-wide integrated imbalance value *ILBi*, of server *i*. In statistics variance is used as a measurement of how far a set of numbers are spread out from each other, therefore it is widely used. Using variance, an integrated load imbalance value (*ILBi*) of server *I* is defined

$$\frac{(Avg_i - CPU_u^A)^2 + (Avg_i - MEM_u^A)^2 + (Avg_i - NET_u^A)^2 + (Avg_i - HD_u^A)^2}{4} \quad (7)$$

Where $$Avg_i = \frac{(CPU_i^u + MEM_i^u + NET_i^u + HD_i^u)}{4} \quad (8)$$

is average utilization of multi-dimensional resource in a physical machine, also called integrated load in LIF

V.EXPERIMENTAL RESULTS

Algorithms are compared using the cloud simulator –cloudsched. Similar to Amazon eight EC2 instances with high CPU, high memory and standard configuration but not exactly the same, eight types of virtual machines with equal probability of requests are generated.Cloudsched simulator generates different requests as follows: the total numbers of arrivals (requests) can be randomly set, all requests follow Poisson arrival process and have exponential length distribution. Therefore to test the algorithm, it is executed six times and its average has been taken. Experiment is conducted using the following data center characteristics.

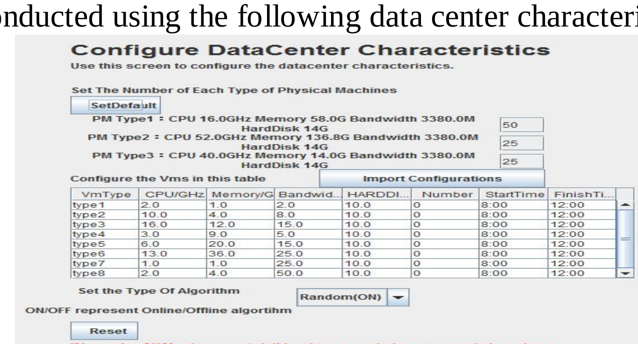

Figure 1. DataCenter characteristics_1

PM Type 1: CPU 16.0 GHz Memory 58.0 G Bandwidth 3380.0 M Hard disk 14 G= 50
PM Type 2: CPU 52.0 GHz Memory 136.8 G Bandwidth 3380.0 M Hard disk 14 G= 25
PM Type 3: CPU 40.0 GHz Memory 14.0 G Bandwidth 3380.0 M Hard disk 14 G= 25

Table - 1: Research utilization and unbalanced degree of data center

	LIF	Extended_LIF	Random	Round Robin	ZHCJ	ZHJZ
Unbalanced degree of data center (Variance)	0.17%	0.14%	0.24%	0.20%	0.21%	0.19%

Table 1 shows average imbalance value of a cloud data center, it can be seen that Extended_LIF algorithm has lowest imbalance value of a cloud data center.

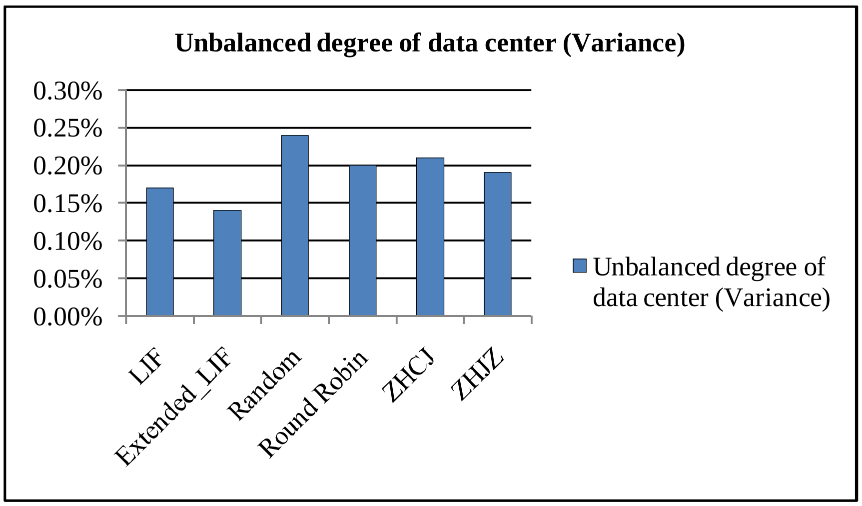

Figure - 2: Unbalanced degree of data center

VI.CONCLUSION

Essential requirements of a dynamic resource scheduler is to reduce imbalance level of the datacenter and to make all the resources are equally loaded. Simulation results shows that proposed Extended_LIF algorithm has low unbalance degree of data centre. It makes load on resources are equally distributed.

VII. ACKNOWLEDGEMENT

The authors would like to thank the dedicated SIG-research group in the area of Cloud & Grid Computing, wireless networks at the Dept. of Computer Science, AIMIT, St Aloysius College,, Mangalore, India, for many simulating discussions for further improvement and future aspects of the Paper. Lastly but not least the author would like to thank everyone, including the anonymous reviewers

REFERENCES

[1] Wenhong Tianet.al LIF: A Dynamic Scheduling Algorithm for Cloud DataCenters Considering Multi-dimensional Resources ,Journal of Information & Computational Science 10:12 (2013) 3925–3937.

[2] T. Wood et. al., Black-box and gray-box strategies for virtual machine migration, Proceedings of Symp. on Networked Systems Design and Implementation (NSDI), 2007.

[3] H. Zheng, L. Zhou, J. Wu, Design and implementation of load balancing in web server cluster system, Journal of Nanjing University of Aeronautics & Astronautics, Vol. 38, No. 3, Jun. 2006.

[4] eucalyptus, www.eucalyptus.com.

[5] Wenhong Tianet.al LIF: A Dynamic Scheduling Algorithm for Cloud DataCenters Considering Multi-dimensional Resources ,Journal of Information & Computational Science 10:12 (2013) 3925–3937.

[6] A. Gulati, G. Shanmuganathan, A. Holler, Cloud-scale resource management: Challenges and techniques, Proceedings of HotCloud 2011, Portland, OR, USA, June 14-17, 2011.

[7] A. Singh, M. Korupolu, D. Mohapatra, Server-storage virtualization: Integration and load balancing in data centers, Proceedings of the 2008 ACM/IEEE Conference on Supercomputing, 2008,1-12.

[8] B. Wickremasinghe et al., CloudAnalyst: A CloudSim-based tool for modelling and analysis of large scale cloud computing environments, Proceedings of the 24th IEEE International Conference on Advanced Information Networking and Applications (AINA 2010), Perth, Australia, April 20-23, 2010.

[9] E. Arzuaga, D. R. Kaeli, Quantifying load imbalance on virtualized enterprise servers, Proceedings of WOSP/SIPEW'10, San Jose, California, USA, January 28-30, 2010.

[10] W. Tian, Adaptive dimensioning of cloud data centers, Proceeding of the 8th IEEE International Conference on Dependable, Automatic and Secure Computing, DACS 2009, 2009.