ISSN:0975-9646

Scheduling Resources in Cloud using Threshold Values at Host and Data Center level

Yatendra Sahu[#1], Neha Agrawal[*2]

[#]*Dept. of CSE, PSIT*
Kanpur, India

[*]*Dept of CSE, MANIT*
Bhopal, India

Abstract— **Cloud computing provides online resources of computing services as per demand of their client. One of the challenging areas in cloud computing is scheduling the computing resources among the client's applications and services. If resource requirement of the cloud applications increase and host machine become overloaded then using migration techniques, cloud system need to transfer some of the virtual machines to another host machine, whose load is low. When cloud applications change their demand very frequently then the frequency of virtual machine migration will be high and our throughput will be low. This paper is proposing a threshold based approach to make efficient resource scheduling with minimizing number of virtual machine migration situation using compare and balance algorithm. The new, simple, flexible and dynamic threshold based resource scheduling algorithm get higher resource utilization with improved throughput for dynamic application as shown by the simulation results using the cloudsim simulator**

Keywords— **Cloud Computing, Dynamic Resource Scheduling, Efficient Resource Utilization, Virtual Machine Migration.**

I. INTRODUCTION

Cloud computing is a recent technology that distributes the computing resource online and allow their client to execute their application services using these distributed resources [1]. Cloud computing works on pay per use model so that it becomes very famous in a few times [2]. The cloud vender who provides the online resources, have to manage the entire configuration of cloud system without interrupting the efficiency and availability of the services. All the work related to software and hardware installation and maintenance are managed by cloud system in back-end and cloud users just use the services on demand with pay per use scheme [2]. The National Institute of Standards and Technology (NIST) define the cloud computing by this definition says that: "Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, data storage, software applications and other computing services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [3]. In the cloud systems, there are many types of scalable computing resources like as software resources, hardware resources and data storage devices etc. Using the attributes and properties of resources, cloud system manages resource scheduling and provides these resources to their client as per demand. In a data center, there may be any number of host machines as shown in Fig. 1 and each host machine distributes their resource to user application with different-different work load according to online demand of resource requirement.

To maximize the recourse utilization of the cloud server, we have to schedule the cloud resources according to the resource requirement of client. Load balancing techniques also based on scheduling algorithms to balance and improve utilization factor of resources. In this paper, we discuss a dynamic compare and balanced algorithm that work on dynamic threshold values. The goal of the paper is scheduling of computing resource of the cloud data center to balance the cloud host machines and improve the resource utilization

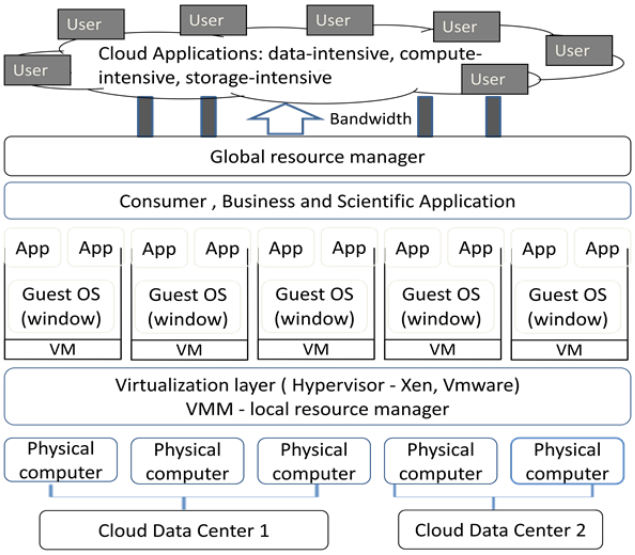

Fig1. Cloud System with VM, Host, Data Center and User Application

. We propose simple, new and effective algorithm for cloud resource scheduling using threshold values of host level and data center level. This paper is organized as follows. In section 2, we discuss the background knowledge of the resource allocation scheme with process scheduling. Section 3 gives brief descriptions about proposed work with compare and balance theory. In section 4, we present the proposed algorithms using flowchart. Section 5 discuss about evaluation of our scheduling algorithm. In section 6, we discuss our experiment and simulation result. Section 7 concludes the paper and section 8 focus on the future work.

II. RELATED WORK

Cloud computing refers an online delivery of computing and storage services to an end-recipients. Cloud system is a collection of distributed data center located throughout the world and these data centers are connected using communication network like as internet. The computing resources are provided to their client using virtualization. Virtualization plays a key role in cloud computing by sharing of single host machine to number of Clint's application. Virtualization technology allows all physical resources to be virtualized and be transparent to client. Client doesn't need to have any information about the hardware type, physical location, level information of computing resources. Client just make a demand for the computing resources and according to the client requirement, the cloud system creates virtual machines inside any one of the host machine of a data center and using that virtual machine clients perform or execute their task [4].

Process scheduling is one of the important parts of cloud system, in which appropriate resources are assigned to users' task. Efficiency of process scheduling algorithm directly affects the performance of the whole cloud system performance [5]. There are several studies [5, 6, 7, 8 and 9] related to the problem of management cloud resources of host machine in various levels like as virtual machine level, host machine level and data center level. As shown in Fig. 2, when server receive the user request for computing resource on demand then it is stored in job pool. In the job pool, many types of the scheduling algorithms may be applied. Using these scheduling techniques, resources of host machine are assigned to user jobs using cloud broker.

Fig2. Job Allocation Policy in Cloud Server

First come first serve, shortest job first, priority bases scheduling and round robin scheduling are some of techniques used by computers. When we assign the computing recourses of the cloud server to the client application then many times the application change their requirement dynamically with respect to time. If system is fully utilized at a particular time and the resource requirement of an application varies then the application requirement is unable to be fulfilled by that particular host machine. At this situation, we need to apply some migration technique to transfer the some of the application of this host machine to the other host machine to remove the situation of over loading. If the frequency of the migration of the application from one host to another host machine due to variation of resource requirement then this overhead affects the efficiency of the cloud system. So there is always an efficient resource scheduling algorithms to allocate the computing resources with minimum upcoming overheads. We should use some of the dynamic and advanced scheduling algorithm then we can decrease frequency of the situation of over loading in the host machine.

III. PROPOSED WORK

We proposed a threshold based compare and balance technique to efficiently allocate the cloud computing resources where user application dynamically changes their resource requirement with respect to time. Generally the online resource distribution system use the combination of resource scheduling algorithm like as first come first serve or priority based approach. But in the cloud application, the requirement of resources are variable so here we should add some of the extra features need some as request of computing resource arrived. Resource allocation policies also concern with the capacity of the cloud server to efficient utilization of the resources. As the cloud system is a collection of data center located in different location of the world. And each data center is also a collection of host machine. So during the allocation of resources to a cloud service or cloud application, scheduling algorithms should be applied on the both level respectively data center level and also host machine level. In our work, we concern with data center level to choose appropriate data center and appropriate host machine having enough resources to fulfil the dynamic requirement of application with respect to time. Here we apply two concepts of resource scheduling algorithms. First concept is selection of a data center that is having some of the host machine with free resources for assign the new cloud application or cloud service. And second concept is to apply the threshold based comparison of data center limit and host machine limit.

In the cloud system suppose there is m total no. of host in particular data center and each host have nj total no. of virtual machine (Vm).

Total no. of virtual machine in the whole data center is n.

Where, $n = n1+n2+n3+\ldots\ldots\ldots +nj+\ldots\ldots\ldots+ nm-1+nm$.

Load of a host machine (H_Load) = α*(CPU Usage) + β*(Ram Usage) + γ*(BW Usage)

Where α, β and γ are respectively weight coefficients.

And, $\alpha + \beta + \gamma = 1$, $(\alpha, \beta, \gamma) \epsilon [0 \tilde{} 1]$.

Here we have used some threshold value to compare the load and these all threshold value are calculated using total capacity of the server multiplied by a respectively weight coefficients.

(1)Host_Limit: Total capacity limits of load value of host machine which may be assigned without effect its efficiency.

(2)DC_Limit: Total capacity limits of load value of data center which may be assigned without effect its efficiency.

(3)Threshold_Value_Of_Host (H_TD) : Host_Limit * δ (Where δ is a constant for host machine set by cloud provider on the basis of dynamic behaviour of applications and services).

(4)Threshold_Value_Of_Datacenter (DC_TD) : sum of Host_Limit of all host machine * σ (Where σ is a constant for data-center set by cloud provider on the basis of dynamic behaviour of applications and services).

Data center threshold (DC_TD) is the upper capacity of the data center at which we can assign the resources of cloud to application services without affecting its efficiency. And the host machine threshold (H_TD) is the upper limit of the host machine. There is also two other values namely data center limit (DC_Limit) and host machine limit (Host_Limit), which tell us the available resource capacity of the cloud server and host machine.

IV. ALGORITHM

We are proposing process scheduling algorithm with global and local process scheduler work on compare and balance technique scheme based on threshold value as shown in Fig. 3. We are implementing this algorithm using two threshold values as threshold value of data center and threshold value of host machine. The entire process of proposed scheduling algorithm includes following stages:

STAGE 1. Measurement of Threshold Values – In the starting stage of scheduling, the threshold value of host and data center are calculated using predefined formulas.

STAGE 2. Global Process Scheduler – When any cloud service is requested then it arrives in job pool. Global process scheduler selects the appropriate data center by comparing their load and limit value to threshold value.

STAGE 3. Local Process Scheduler – After selection of appropriate data center, the appropriate host system is found inside the selected data center. If host system is available with given constraints then control is transferred to next stage to create virtual machines and allocate the computing resources. Otherwise go back to stage 2.

STAGE 4. Resource Allocation – In this stage the requested resource are allocated to the cloud application with the help of virtual machine.

STAGE 5. Finish – In this stage, cloud application are allowed to start execution and update the resource availability database of cloud system.

The flowchart of the algorithm is given like as:

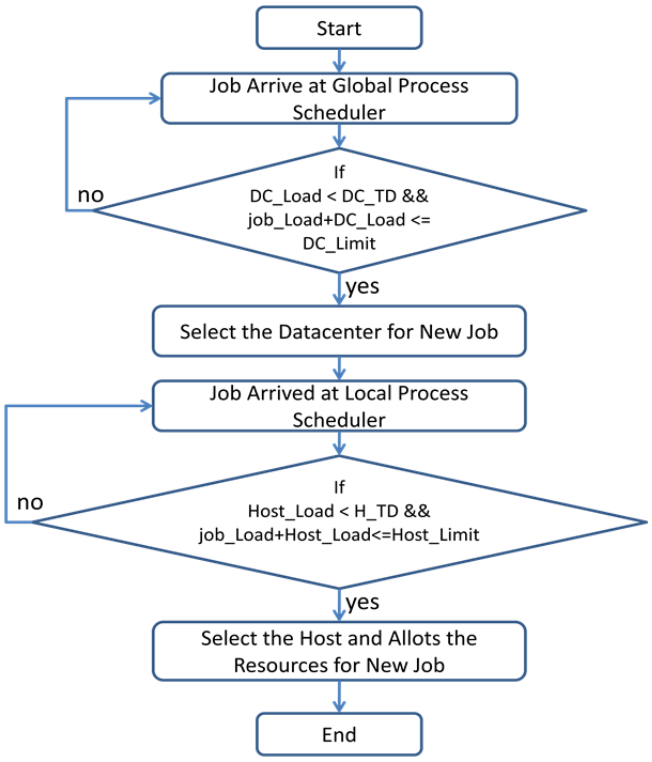

Fig3. Job Allocation Algorithms in Cloud Server

If Load of any host system (Host_Load) is not less than upper threshold value then this host machine is not feasible to create new virtual machine and assign the computing resources of cloud application. If condition is true than only the resources of this host machine are allocated to the cloud application. Otherwise we continuously find the host machine until we get success.

V. DISCUSSION AND EVALUATION

Cloud application are dynamic in nature, means they increase their resource requirement at time to time. If server has enough resources then it can fulfill the dynamic resource requirement otherwise it needs to apply appropriate migration technique to transfer the virtual machine of the particular process into another host computer having enough resource to execute this process. The migration of virtual machine causes come of overhead as service downtime, service privacy and security management etc. If the frequency of process migration from one host machine to another host is high then this overhead affect the efficiency of the cloud system. To reduce the frequency of process migration as shown in Fig. 4, threshold based process scheduling approach are used to improve the resource utilization with decreasing the number of migration of process.

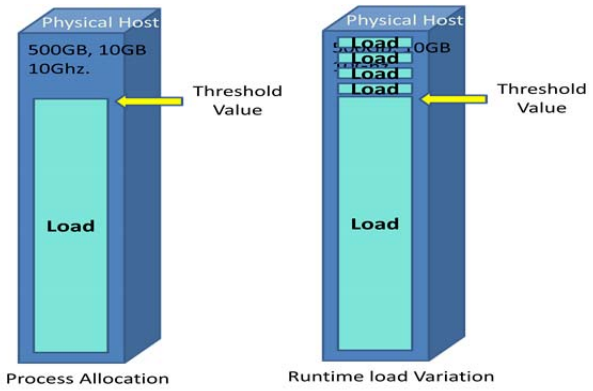

Fig4. Threshold in Job Scheduling Policy in Cloud Server

Here the threshold values are initialised on the basis of cloud application behaviour. If the application resource requirement changes very frequently then the threshold value should be low. If resource requirement increase then host should be able to handle the situation without applying migration of virtual machine.

VI. Simulation and Result

In order to verify the accuracy and feasibility of our approach, we used the cloudSim simulator of cloud computing platform to realize the process scheduling algorithm. CloudSim is a new, generalized, and extensible simulation framework that allows seamless modelling, simulation, and experimentation of emerging cloud computing infrastructures and application services. We simulate our work using cloudsim toolkit that is used to develop a cloud environment. Cloudsim simulation tool of cloud computing is java based tool to create cloud environment with data center and host machine. There is all the functionality of cloudsim toolkit is written in java classes to perform cloud operation. Java class of datacenter, datacenterBroker, datacenterCharacterstics, host, vm, vmAllocationPolicy, vmScheduler, cloudlet and cloudletScheduler are the main classes of the cloudsim simulator. In the first step of establishing cloud environment, datacenters are created using the host machine having its characteristics and physical host machine and its resources description.

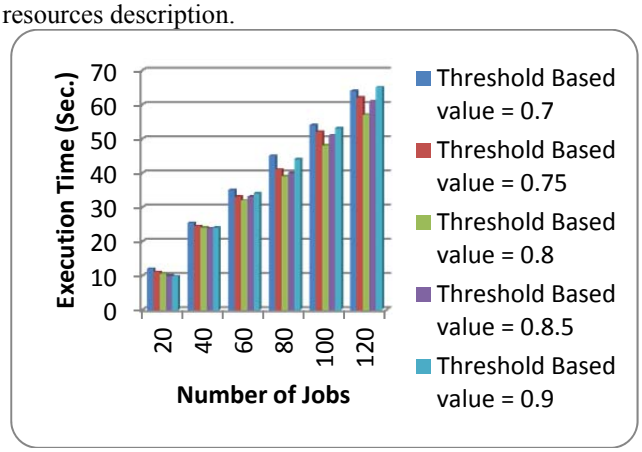

Fig5. Job Execution in Threshold based Scheduling Algorithm.

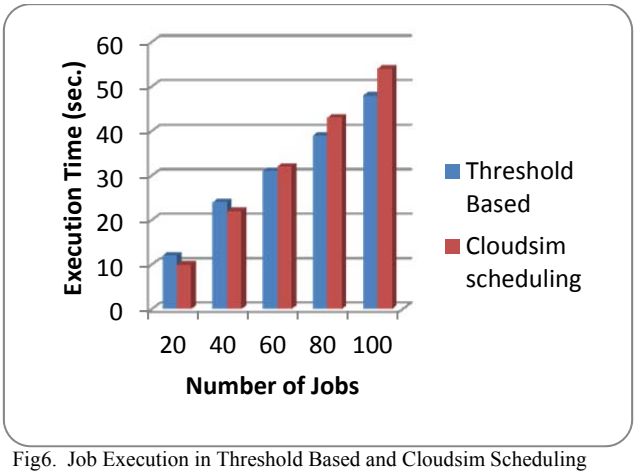

Fig6. Job Execution in Threshold Based and Cloudsim Scheduling Algorithm.

In the threshold based process scheduling technique the job execution time is improved if the cloud application services are dynamic in nature, means the increase their resource requirement at time to time. Our simulation results show that threshold based scheduling algorithm for cloud server presents improved utilization of resources.

VII. Conclusions

This paper has shown the applicability of threshold based process scheduling techniques to obtain measurable improvements in resource utilization, server workload management and efficient process scheduling techniques to the cloud server. The proposed approach could bring cost advantages to cloud vendors who are concerned with utility costs and who are searching for efficiencies that can be relatively easily achieved. As situation to migrate the process due to dynamic requirement of computing resource decreased by this techniques. This is due to the pricing of transfer between regions, availability zones and cloud vendors, which all constitute different pricing strategies. As resource scheduling algorithm internally affects the cloud server load balancing.

VIII. Future Work

As this algorithm are evaluated using simulation environment so there may be some of the new challenges in real implementation of cloud computing. We are implementing it in real cloud environment. In the situation of dynamic changes of resource requirement, the resource of any host is not fully assigned so there may be poor resource utilization. So we are working to improve the resource utilization.

References

[1] Marios D. Dikaiakos, George Pall, Dimitrios Katsaros, Pankaj Mehra, Athena Vakali, "Cloud computing : Distributed Internet Computing for IT and Scientific Research," in *IEEE Internet Computing*, Published by the IEEE Computer Society, 2009.

[2] Panagiotis Kalagiakos, Panagiotis Karampelas, "Cloud Computing Learning," IEEE, in *Hellenic American University Manchester*, N.H. - U.S.A, 2011, pp: 7/11.

[3] Peter Mell,Timothy Grance, The NIST Definition of Cloud Computing, in *National Institute of Standards and Technology* - Computer Security Resource Center-www.csrc.nist.gov.

[4] Yashpalsinh Jadeja, Kirit Modi, "Cloud Computing- Concepts, Architecture and Challenges," in *International Conference on Computing, Electronics and Electrical Technologies*, IEEE, 2012, pp: 4/12.

[5] Jinhua Hu, Jianhua Gu, Guofei Sun, Tianhai Zhao "A Scheduling Strategy on Load Balancing of Virtual Machine Resources in Cloud Computing Environment," in *3rd International Symposium on Parallel Architectures, Algorithms and Programming*, IEEE, 2010, pp: 3/10.

[6] Yiqiu Fang, Fei Wang, Junwei Ge, "A Task Scheduling Algorithm Based on Load Balancing in Cloud Computing," *Springer,* 2010, pp: 71/77.

[7] Qin Zheng and Bharadwaj Veeravalli "On the Design of Mutually Aware Optimal Pricing and Load Balancing Strategies for Grid Computing Systems," *Published by the IEEE Computer Society*, 2013.

[8] Wang Han, Tang Xiao-Qi, Song Bao, Tang Yu-Zhi, "Dynamic Task-Scheduling Algorithm in CNC System Based on Cloud Computing" *Second International Conference on Instrumentation & Measurement, Computer, Communication and Control*, IEEE, 2012, pp: 4/12.

[9] Liang Luo, Wenjun Wu, Dichen Di,Fei Zhang,Yizhou Yan,Yaokuan Mao, "A Resource Scheduling Algorithm of Cloud Computing based on Energy Efficient Optimization Methods" *IEEE Computer Society*, 2012, pp: 9/12.

[10] R. G. Babukarthik, R. Raju, P. Dhavachelvan, "Energy-aware scheduling using Hybrid Algorithm for cloud computing," in *ICCCNT*, Coimbatore, IEEE, July 2012.

[11] V.Vinothina, R.Sridaran, PadmavathiGanapathi, "A Survey on Resource Allocation Strategies in Cloud Computing," in *International Journal of Advanced Computer Science and Applications,* 2012, Vol. 3, No.6.

[12] Yi Zhao, Wenlong Huang, "Adaptive Distributed Load Balancing Algorithm based on Live Migration of Virtual Machines in Cloud," in *Fifth International Joint Conference on INC, IMS and IDC*, IEEE, 2009, pp: 6/09.