ISSN 2350-1022

International Journal of Recent Research in Mathematics Computer Science and Information Technology
Vol. 2, Issue 1, pp: (310-313), Month: April 2015 – September 2015, Available at: www.paperpublications.org

A Task Scheduling Algorithm in Cloud Computing

Ali Bagherinia

Department of Engineering, Dehdasht branch, Islamic Azad University, Dehdasht, Iran

Abstract: **Efficient task scheduling method can meet users' requirements, and improve the resource utilization, then increase the overall performance of the cloud computing environment. Cloud computing has new features, such as flexibility, Virtualization and etc., in this paper we propose a two levels task scheduling method based on load balancing in cloud computing. This task scheduling method meet user's requirements and get high resource utilization that simulation results in Cloud Sim simulator prove this.**

Keywords: **cloud computing; task scheduling; virtualization.**

I. INTRODUCTION

Cloud computing dates back to the 1960's when John McCarthy opined that "computation may someday be organized as a public utility". Amazon played a key role in cloud computing development by launching Amazon web service on utility basis in 2006. Many users share cloud resources by submitting their computing task to the cloud system. Scheduling these tasks is a challenge to cloud computing environment. Task scheduling is an important part of cloud computing, which is a mechanism that maps users' tasks to appropriate resources to execute, its efficiency will directly affect the performance of the whole cloud computing environment. Compared with grid computing, it has some properties such as virtualization and flexibility for cloud computing that grid computing doesn't have [1]. By using virtualization technology, all users have their own virtual machine and don't affect each other. Furthermore utilization of resources is improved effectively, and the users' application is running independency and security of system and service availability is improved. Flexibility is the dynamically increasing or decreasing of resource provided by cloud computing environment according to users' demand. Cause these features, the task scheduling approaches for grid computing cannot work effectively in the cloud computing environment. Different scheduling methods are proposed in [2], [3], [4], [5], [6], [7], [8], [9] and [10].

In this paper, a task scheduling method based on the two levels of load balance is proposed, which consider the flexibility and virtualization in cloud computing to meet improve the utilization of resources and the dynamic task requirements of users.

II. SCHEDULING MODEL

Cloud Computing Architecture includes three layers, infrastructure layer, platform layer and application layer [11]. The infrastructure layer is a set of virtual hardware resources and related management function. The platform layer is a set of software resources, which can provide an environment for developing, running, managing and monitoring cloud application to. The application layer is user oriented; it implements the interaction mechanism between user and service provider with the support of platform Layer. Users can submit tasks through the application layer. According to the above architecture, we propose two levels scheduling model [12] in this paper as shown in Fig. 1.

ISSN 2350-1022

International Journal of Recent Research in Mathematics Computer Science and Information Technology
Vol. 2, Issue 1, pp: (310-313), Month: April 2015 – September 2015, Available at: www.paperpublications.org

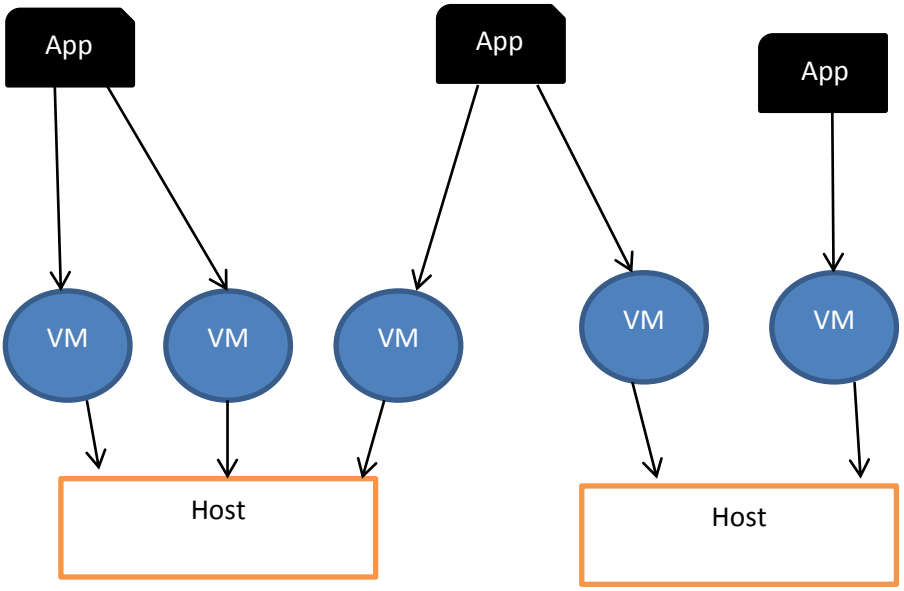

Fig. 1 Two levels scheduling Model

As shown in Fig. 1, the first scheduler create the task description of a virtual machine, including the task of computing resources, network resources, storage resources, and other configuration information, according to resource demand of tasks. Then the second scheduler finds appropriate resources for the virtual machine in the host resources under certain rules, based on each task description of virtual machine. Via the two levels scheduling, the task can obtain the required resources, and it would not lead to the resource allocated to some tasks is less than requirement and increase the implemental time while others are more than their requirements and lead to the waste of resources.

Here all tasks are considered computational, and the tasks are independent each other, and the execution of the task replication is also not considered.

III. SCHEDULING ALGORITHM

The load of virtual machine discussed in this paper is expressed by the predicted executing time of tasks running in the virtual machine, named as VL_i[13].And the load of host is expressed by the average load of the virtual machine that run on it, named as HL_i, the average load value $avgl$ and load balancing evaluating value B of cloud computing environment can be defined as follows:

HLi $= \sum_{k=1}^{n} VL_k / n$,

$avgl = \sum_{k=1}^{n} HL_k / m$ and $B = \sqrt{\sum_{k=1}^{n} (L_k - avgl)^2} / m$,where the n is the number of virtual machines that run on the host and m is the number of hosts ism, the smaller value B the better load balancing and the bigger value B the worse load balancing.

In order to meet users' requirements and increase the utilization of resources, a scheduling algorithm based on load balancing is proposed in this paper. The algorithm is based on the former scheduling model discussed, considering the flexibility and virtualization features of cloud computing, it is divided into two levels scheduling, one is the mapping from task to a virtual machine, another is mapping from the virtual machine to host resources. Generally, for the requirement of the task, users want to get the best response time. Therefore, only task response time and the demand for resources are considered in this paper. At the same time, because tasks are dynamic, they may arrive randomly. If the tasks arrive at same time, they will be sorted ascending according to the resource applied by users. And if they arrive at different time, they will be sorted according to the time sequence arrived. The steps of this algorithm are described as follows:

Step1: sort host resources ascending order of their processing power.

ISSN 2350-1022

International Journal of Recent Research in Mathematics Computer Science and Information Technology
Vol. 2, Issue 1, pp: (310-313), Month: April 2015 – September 2015, Available at: www.paperpublications.org

Step2: According to the task model the first level scheduler establish the virtual machine description according to the properties of task, providing configuration information for allocation of resources and creation of the virtual machine.

Step3: According to the virtual machine description of Task t_i, select a host resource h_j that can meet the required resources and the load is lightest. If the host exists, create the virtual machine and allocate the required resource for it, then update the available resources h with total service ability Fcap of Host h_j , otherwise take the Task t to the tail of the task queue and waiting for the next scheduling.

Step4: If the resource requirements of the Task t_i increase, find whether the host whose virtual machine of Task t_i run on can meet the additional required resources, if it exists, allocate the additional required resources for it, reconfigure the virtual machine, and then update the host's available resources. Otherwise, the virtual machine is migrated to the host with lightest load and the additional required resources to execute continuously.

Step5: If the resource requirements of the Task t_i reduce, release the excess resources that the virtual machine occupied, and update the available resources hold by the host.

Step6: If Task I t has been completed, then destroy the virtual machine of Task t_i and release the occupied resources for the other unfinished tasks.

Step7: Calculate the load balancing evaluating value Bin current environment, if B is greater than the threshold value B_{thr}, that indicates the load balancing state is worse, select a virtual machine with lightest load and migrate it to the host which can meet the resource requirement with the lightest load.

Step8: Repeat step3 to 7 until all tasks are completed.

In the above algorithm, the virtual machine is scheduled to the host with lightest load each time. The advantage is to avoid overloading for the host hold more resources. If the current virtual machine is scheduled to a host, as the computational amount increase, leading to the virtual machine's load is heavy, resulting in load imbalance, then take the dynamic migration operation, keeping load balance in current environment.

IV. SIMULATION

We simulate proposed algorithm and Job Scheduling Using Fuzzy Neural Network Algorithm [14] in the CloudSim toolkit [15]. This simulation mainly shows the advantage of the proposed algorithm compared to Fuzzy Neural Network Algorithm in make span term based on the various number of tasks.

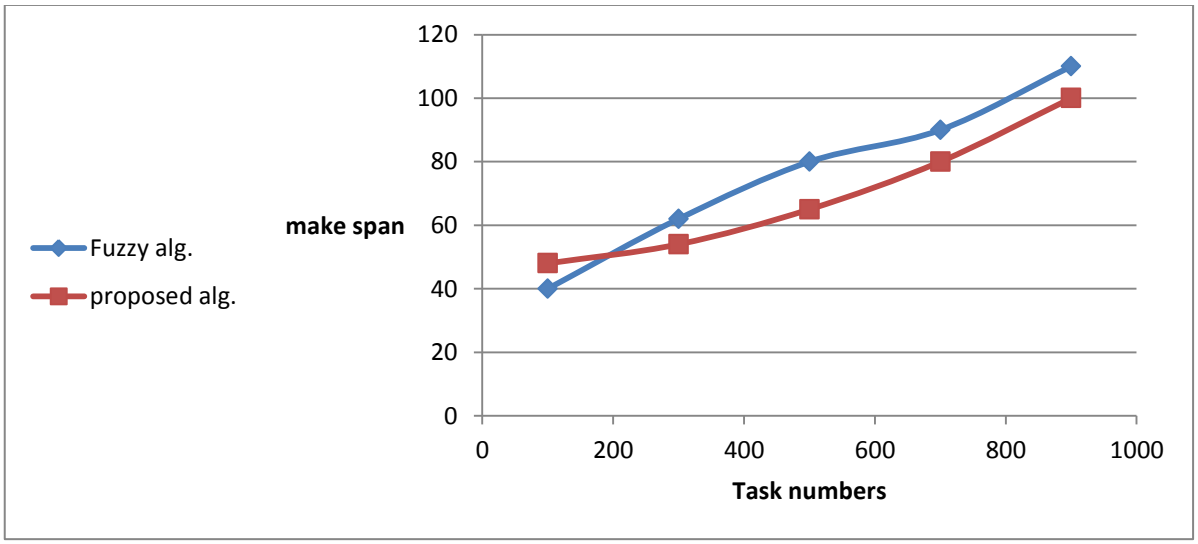

Fig. 2.Makespan Comparing

As shown in Fig. 2, with the task number increasing, the makespan in our algorithm less than Fuzzy Neural networks algorithm, which is due to the flexibility of our algorithm.

REFERENCES

[1] Foster, I., Zhao, Y., Raicu, I., Lu, S.: Cloud Computing and Grid Computing 360-Degree Compared. In: Grid Computing Enviroments Workshop, pp. 1–10 (2008) .

[2] Paul, M., Sanyal, G., "Survey and analysis of optimal scheduling strategies in cloud environment", IEEE, 2012.

[3] Jeyarani, R., Ram, R. Vasanth, Nagaveni, N., "Design and Implementation of an Efficient Two-Level Scheduler for Cloud Computing Environment", IEEE, 2010.

[4] Huang Qi-yi, Huang Ting-lei, "An optimistic job scheduling strategy based on QoS for Cloud Computing", IEEE, 2010.

[5] Meng Xu, Lizhen Cui, Haiyang Wang, Yanbing Bi, "A Multiple QoS Constrained Scheduling Strategy of Multiple Workflows for Cloud Computing", IEEE, 2009.

[6] Hao Li, Huixi Li, "A Research of Resource Scheduling Strategy for Cloud Computing Based on Pareto Optimality M×N Production Model",, IEEE, 2011

[7] Kuan-Rong Lee, Meng-Hsuan Fu, Yau-Hwang Kuo, "A hierarchical scheduling strategy for the composition services architecture based on cloud computing", IEEE, 2011.

[8] Gunho Leey, Byung-Gon Chunz, Randy H. Katzy, "Heterogeneity-Aware Resource Allocation and Scheduling in the Cloud", University of California.

[9] Shu-Ching Wang, Kuo-Qin Yan, Shun-Sheng Wang, Ching-Wei Chen, "A Three-Phases Scheduling in a Hierarchical Cloud Computing Network", IEEE, 2011.

[10] Peixoto, M.L.M., Santana, M.J., Estrella, J.C., Tavares, T.C., Kuehne, B.T., Santana, R.H.C., "A Metascheduler architecture to provide QoS on the cloud computing", IEEE, 2010.

[11] Armbrust, M.: Above the Clouds: A Berkeley View of Cloud Computing. In: EECS Department, University of California, Berkeley (2009).

[12] Sadhasivam Dr., S., Jayarani, R.: Design and Implementation of an efficient Two-level Scheduler for Cloud Computing Environment. In: International Conference on Advances in Recent Technologies in Communication and Computing, vol. 148, pp. 884–886 (2009).

[13] Xiang-hui, P., Er-hu, Z., Xue-yi, W., Guang-feng, L.: Load balancing algorithm of multicluster grid. Computer Enineering and Applications 45(35), 107–110 (2009).

[14] Venkatesa Kumar and K. Dinesh, Job Scheduling Using Fuzzy Neural Network Algorithm in Cloud Environment, Bonfring International Journal of Man Machine Interface, Vol. 2, No. 1, March 2012.

[15] The CLOUDS Lab.CloudSim: A Novel Framework for Modeling and Simulation of Cloud Computing Infrastructures and Services [EB/OL],http://www.gridbus.org/cloudsim.